# Expediting the Targeting Process Responsibly using Artificial Intelligence

Christopher J. Lowrance, Elliott R. Harris, and C. Anthony Pfaff

United States Army War College, 47 Ashburn Drive, Carlisle, PA 17013 USA

## ABSTRACT

Absent the use of artificial intelligence (AI), each stage of the targeting process requires human input, which consumes time – a precious resource during high-intensity conflict. To counter that problem, the application of AI and related technologies could rapidly automate certain steps within the targeting process. The challenge, however, becomes how to responsibly speed up the process using AI, while also maintaining appropriate levels of human oversight based on level of risk acceptable to the commander. We address this challenge with the introduction of a fuzzy logic controller that is designed to automatically adapt and optimize the level of human-machine interaction based on the current targeting conditions. The logic controller uses the confidence of the AI algorithm at each stage of the targeting process, as well as the commander's risk tolerance at the time, to automatically determine what stages of the targeting process can be trusted for AI to accelerate and which ones require explicit human verification. The final stage of verifying and authorizing a fire mission is reserved for humans only to make.

**Keywords:** AI-enabled targeting process, military artificial intelligence, Department of Defense modernization, fires warfighting function, fuzzy control and decision making

## 1. INTRODUCTION

The U.S. Army, like many others, is placing greater focus on how it integrates rapidly advancing technology to support its core function of fighting and winning, especially against peer and near-peer competitors in large-scale combat operations (LSCO).[1] In fact, the Army has set the year of 2035 as its goal to employ new warfighting doctrine that is underpinned by artificial intelligence (AI) and other technologies, so that it can prevail throughout the spectrum of competition.[2] Unlike counter-terror or counter-insurgency operations, which generally involved the targeting of individuals or small groups, LSCO will require a much higher operations tempo with thousands of targets a day that must be engaged faster than the enemy can target one's own assets. Attaining that speed will not only require the introduction of AI technologies, but it will also require networking these technologies with space-based sensors and networked command and control systems. Done successfully, these technologies will enable militaries to *sense, understand, decide, act, and assess* to achieve *decision dominance* faster than ever before.[3] To achieve decision dominance, the Army and Joint Force must leverage AI's speed in all six warfighting functions: mission command, movement & maneuver, intelligence, fires, sustainment, and protection. However, while each warfighting function is important, "success in LSCO is dependent on the Army's ability to employ fires",[4] which includes the integrating process of targeting as the means to prioritize targets, effects, and resources for the commander. Thus, leveraging AI to expediate the targeting process must be given the utmost attention, which aligns with the Army's modernization strategy that has "Long-Range Precision Fires" as its top priority.[2]

While AI should be leveraged in targeting, it brings with it challenges not only associated with how to integrate it effectively, but how to ensure that commanders and their staffs can provide meaningful human control to mitigate potential collateral harms both to friendly forces and noncombatants. Some challenges arise as a function of AI's robustness, where erroneous behavior may be caused by inputs not representing the training data[5–7] or adversaries attempting to spoof or hack the system.[8] Another drawback is the opaqueness of AI,

Further author information: (Send correspondence to Christopher Lowrance)
E-mail: christopher.lowrance.mil@armywarcollege.edu, Telephone: (717) 245-4138
The views presented in this paper are those of the authors and do not necessarily represent the views of Department of Defense or its components.

where operators do not adequately understand how and why a system recommends or takes an action.[9] Failure to address these concerns can undermine the trust commanders have in the system.[5,10]

Trust and risk, of course, are the central concerns here. Commanders need a reliable way to know when AI can be trusted and when it should be allowed to execute some stages of the targeting process with less supervision for the benefit of speed, but at the cost of more risk. Fortunately, there are precedents that can help provide more transparency regarding the performance of AI-enabled systems[11] and human-centric measures that focus on improved interfaces and teaming aspects.[12,13] What these precedents show is that adaptive human-machine teaming that adjusts the level of human intervention as conditions change can make the targeting process not just more effective, but safer as well. To manage the appropriate level of human-machine interaction, a fuzzy logic controller can be used to automate the decision as to the number of human interventions required for a given target based on a series of conditions. Fuzzy logic provides the ability to impart human reasoning in an automated decision-making process through a series of rules, and for the application of controlling the amount of interaction between humans and AI in the process of targeting, the fuzzy controller could consider multiple criteria before making each teaming decision, such as: 1) the commander's risk tolerance for the given mission, 2) the AI algorithm's confidence in the existence of a target, 3) the choice of best effect (or shooter) as selected by an AI optimization algorithm, and 4) the likelihood of collateral damage and fratricide as assessed by an air and ground deconfliction algorithm enabled by AI like the one being developed by the Defense Advanced Research Projects Agency.[14]

Such an adaptive teaming model is much like other safety-critical systems that are starting to integrate AI. For instance, the automotive industry is using AI in its advanced driver-assist systems (ADAS) to automate much of the driving for a human who supervises and takes control whenever the system loses confidence or when the human observes errant behavior.[15–17] Just like this kind of teaming model, military applications of AI could execute parts of the targeting process with only human supervision, but whenever the AI is not highly confident about its assessment, the process flow would be blocked until a human explicitly verifies the output of the AI. The key is getting the human-machine pairing right because failure to do so results in slowing down the process and undermining the advantages that AI is supposed to provide. Thus, integrating AI into the targeting process presents an optimization problem where the goal is to find the right mix of human-machine teaming that maximizes the speed of AI without overly restrictive control measures, while still ensuring the appropriate amount of human oversight for safety and risk reduction. The proposed fuzzy controller presented in this paper will help solve this optimization problem.

## 2. THE TARGETING PROCESS

According to Joint doctrine, the targeting process can be generally grouped into two categories: deliberate and dynamic.[18] Deliberate targeting is a staff process, usually conducted 24-96 hours in advance, to anticipate targets and assign resources against them for future operations.[19] On the other hand, *dynamic targeting* is executed during current operations and involves unanticipated, unplanned, or newly detected targets.[20] Newly discovered targets may have actions and resources already planned for them; however, those that are *unanticipated and unplanned* are referred to as *targets of opportunity*. Such targets must go through the complete targeting cycle under tighter time constraints given the absence of prior planning and the need to perform other duties to support current operations. In either case, deliberate or dynamic targets might be *time-sensitive* in nature, requiring them to be addressed more quickly to avoid the targets either fleeing or harming the force. Compound that fact with the nature of LSCO that will present many targets, potentially on the order of thousands in short order, assuming AI is used to quickly identify targets based on data from space-based sensors that can rapidly scan large geographic areas. Given each target may potentially pose its own threat, the U.S. military will need to execute its targeting process faster than the enemy does in LSCO regardless of whether newly detected targets are classified as either deliberate or dynamic.

During LSCO, the deliberate targeting process will be especially challenged. For instance, sophisticated enemies will attempt to conceal their assets and thwart the means of sensing, making it difficult to completely plan for some targets. Furthermore, an advanced military will possess many mobile vehicles, which have a high probability of changing their positions since the last time they were detected during the deliberate targeting process conducted 24 to 96 hours ago. Although US forces can anticipate certain targets, the locations of them

will likely not be known until an operation is underway. Consequently, more targets will likely be deferred from the deliberate targeting cycle to the dynamic one, ultimately placing increased pressure on teams to provide attention to more targets within a narrower time span. Additionally, LSCO will entail more targets on the battlefield that must be engaged in a shorter amount of time than the US has faced in recent conflicts. During the last two decades of counterinsurgency operations, the U.S. primarily encountered targets that were individuals or small groups that tended to mask themselves among the local population. Consequently, targets presented themselves infrequently and may have taken weeks or even months to develop, and in general, the process demanded meticulous targeting steps and precision strikes to avoid collateral damage in populated areas. On the contrary, targets in LSCO will be numerous, and many will consist of hardware platforms that pose an immediate threat to the force. Therefore, it will be critical to minimize the time between each target detection and effects on target in LSCO. Fortunately, AI can help address these challenges by augmenting the human-led process in ways that leverage the strengths of AI, while preserving human oversight and decision making.

## 3. AUGMENTING HUMAN DECISION MAKING IN TARGETING

Targeting is a serious process, and if combatants make mistakes, it can lead to the unintended loss of innocent or friendly lives. Mistakes generally come in two forms: false positives and false negatives. Sometimes the system misidentifies an object and sometimes it fails to identify an object. Understanding how those mistakes arise is critical to understanding how one can assess confidence in the system. Fortunately, AI-based object classification systems, which can be used in processing satellite images in the search for targets, provide a level of confidence or likelihood that a given detected object resides within a bounded area of an image. This information can be exploited by a decision agent – in this case, the proposed fuzzy controller – to provide a sense of whether a human should explicitly intervene and verify the detection before it proceeds to the next targeting stage.

AI targeting systems rely on deep learning, which is an approach to AI that uses many layers of artificial neural networks to process information in a way that loosely mimics the biological brain.[21] Deep neural networks (DNNs) have proven to be effective at detecting patterns in data, including imagery.[22] For the military, this means that sensors which generate images, such as such as cameras and radars on satellites, can be processed by DNNs to detect targets. This concept is depicted in Fig. 1. The figure shows pixels of an image being processed through the interconnected layers of a DNN that has multiple outputs. Each output pertaining to a target class is coupled with a confidence score that reflects the probability of that target type being present within the image. These confidence scores are usually assigned by a softmax function added at the end of the neural network.[23] The purpose of the softmax function is to normalize the vector of real numbers coming from last layer of neurons in the DNN into a meaningful probability distribution ranging from 0 to 1. Each output from the softmax can be interpreted as the probability of its corresponding target class existing in the image. However, before a DNN can make such predictions with confidence scores, it must be trained to recognize specific targets. This process involves feeding annotated examples (i.e., a training set) through the DNN iteratively many times. During this training cycle, the machine learning algorithm uses the training data to search for the optimal set of tunable parameters to turn the DNN into a generalized prediction model that can be used against new inputs. However, the predictive performance of the model depends upon the similarity of the new inputs with the data previously used during training.

When the inputs to AI classifiers do not resemble the data in which they were trained, then they are likely to make prediction mistakes. This happens, for instance, when a classifier was trained using only images taken during the summer months of openly exposed targets and then presented images of partially concealed targets taken during the winter. To try and counter such mistakes, it is important to continuously search for and collect new informative data examples as they become available and then use them to retrain and update the classifier periodically. Regardless, the main point is that classifiers can make mistakes given the state of the art and the difficulty of collecting comprehensive datasets, but so do humans – especially those under the pressures of high-intensity LSCO. Ultimately, it is important for commanders to understand the limitations of AI and get a sense of the system's performance in similar conditions, thereby enabling them to make wiser decisions regarding how much control they wish to provide to AI in targeting based on their assessed risk calculus for a given mission.

AI is known to be effective at filtering through data at rapid speed and quickly recognizing relevant patterns.[22] Hence, in the case of targeting, AI can be trained to detect certain targets and then be used to process the data

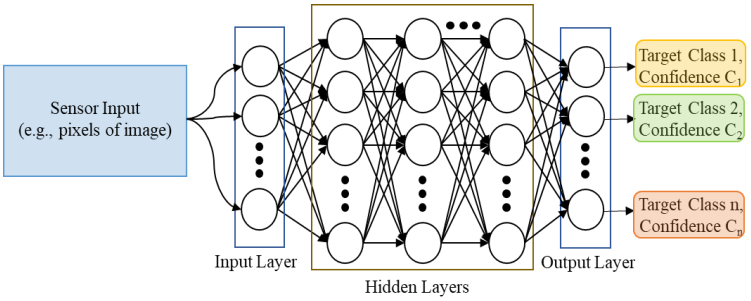


Figure 1. Concept diagram of DNN used target detection. In this case, as new images are captured from a sensor, they would be processed through the DNN which would predict, with a degree of confidence, whether a particular target exists in the image or not.

streams of sensors in near real-time. Fig. 2 depicts one of the potential ways of using AI in the dynamic targeting cycle of "Find, Fix, Track, Target, Engage, Assess (F2T2EA)", where AI can have a significant impact given the inherent time constraints associated with current operations and dynamic targets. As the figure conveys, AI can be used to continuously scan sensor feeds and aid targeting analysts in the steps of "find, fix, and track" by alerting them whenever a new target is detected. It is worth emphasizing that AI can accomplish these steps quicker than humanly possible and can do so continuously without fatiguing or skewing their judgement under the stressors of combat as a team of targeting analysts might.

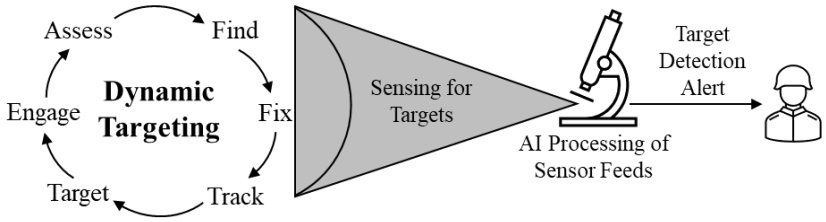


Figure 2. Accelerating the search for dynamic targets using AI to process sensor feeds and send alerts upon making new detections.

One possible drawback to the approach in Fig. 2, where every target detection must be re-inspected by an analyst, is the human bottleneck. It is possible that the number of targets detected by AI in a short period of time greatly outnumbers the number of analysts available to re-inspect them. For instance, it is possible that space-based sensors could quickly scan a wide area near an adversary's base of operations, which in turn the data could be processed by AI in a matter of seconds, producing up to thousands of potential targets for image analysts to verify. In this case, a target queue must be formed, where potential targets wait until an analyst can review them. Any time spent in the queue extends the time between target detection to effects on target. To mitigate such bottlenecks, AI can be paired with humans in a more adaptable way that does not always require the same level of inspection in every scenario. Implementing a simple and static teaming scheme, whereby all targets detected by AI must be individually reinspected by a human may be overly cautious and slow in some circumstances. There may be times when the situation demands the commander to accept more risk for the sake of protecting the force by acting more quickly. In such scenarios, a more agile and adaptive teaming scheme where the commander's risk acceptance and the AI's confidence in its detections could be used to wisely determine which targets can more safely pass the "re-inspection queue" and automatically move to the next targeting stage – especially knowing that a final human verification stage is in place.

In addition to target acquisition, AI can also play an important role in more quickly and effectively designating assets to engage the target and assessing whether the selected effects may cause unintended consequences. Fig. 3 shows three additional, high-level, steps added to the "target" stage of F2T2EA and indicates based on shading which steps are appropriate for AI or other forms of intelligent automation.

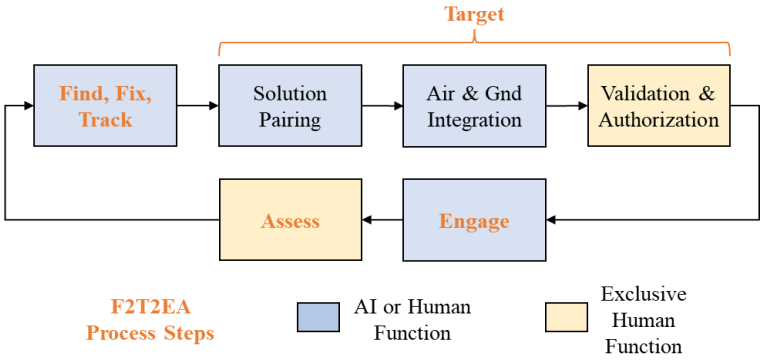


Figure 3. The dynamic targeting steps of F2T2EA expanded to include three more subprocesses within the "Target" step. The figure suggests what steps could be reliably performed by AI and those that must remain exclusively human-based functions, so that the acceleration of the process by AI is done so in a responsible and governable manner.

One factor in asset selection is determining what effect one wants to achieve, such as destroy, disrupt, neutralize, and so on as specified in ATP 3-60.[20] "Solution pairing" refers to matching a target with an appropriate effect. This is a straightforward process in the case of a pre-planned, deliberate target that is already listed in the Attack Guidance Matrix (AGM). However, for some dynamic targets, such plans may not exist. Therefore, when considering the many targets that will be encountered during LSCO and the limited resources available for effects, then this step of selecting the "best shooter" is effectively an optimization problem in which AI is well-suited for solving very rapidly.

A second factor is the risk any asset represents for collateral damage or fratricide. Today, this process is aided by prior planning and the building of "Fire Support Coordination Measures (FSCMs)" and "Airspace Coordinating Measures (ACMs)" and the use of digital systems including the Joint Automated Deep Operations Coordination System (JADOCS) and the Advanced Field Artillery Tactical Data System (AFATDS). FSCMs and ACMs are designed to expedite the attack of targets, while also protecting forces, populations, critical infrastructure, and religious sites from errant fire missions. AFATDS can check to see whether a fire mission violates the FSCMs or ACMs, but nevertheless, leaders typically verify the deconfliction of fires with airspace managers and other team members within the fires cell prior to authorizing a mission, likely due to a lack of confidence that these systems are updated with the latest operating picture. Existing algorithms, to include AI, have the potential to accelerate the clearance of joint fires and make it more reliable, without the heavy reliance on preplanned measures that account for the maneuver of friendly forces, assuming these algorithms are given real-time air and ground common operating pictures (COPs) and access to other useful data like the Army's One World Terrain (OWT) database.[24] With accurate blue-force tracking and 3D digital map data from OWT, an intelligent algorithm can quickly estimate whether an effect is likely to have unintended consequences.

Third, is the subprocess referred to as "Validation and Authorization" in Fig. 3. This step is where a leader within the Fires Cell at any echelon, such as the Deputy Fire Support Coordinator or Targeting Officer uses their best judgement to make a final decision on whether to engage based on guidance, the law of war, rules of engagement, and other considerations outlined in doctrine. As Fig. 3 shows, this stage is reserved for humans only. This guarantees that every targeting mission has at least one human in-the-loop who is responsible for assessing and deciding whether to proceed to engagement using all the information at their disposal. To avoid this stage being a potential bottleneck, the U.S. Army should continue to exercise mission command, whereby multiple targeting officers have the authority to make such engagement decisions based on staying aligned with the commander's intent. Additionally, engagement decisions should have the ability to be made at any echelon who shares an area of responsibility; this would allow for busier fires cells to divert some targets for a faster "validation and authorization".

# 4. AN INTELLIGENT AND AUTOMATED CONTROL SYSTEM FOR ADAPTIVE HUMAN-MACHINE TARGETING

## 4.1 Inputs to the Control System

Ideally, AI can perform the steps indicated in Fig. 3 flawlessly. However, this is not reality. It should be expected that AI will at times make mistakes. Sometimes these errors might be relatively minor like mis-classifying a T-72 tank as a T-90 or not choosing the optimal shooter for a target, while at other times, the consequences can be more severe. For instance, the AI model responsible for the "detection" stage may generate false-positive detections, or even worse, may fail to detect some targets. The latter case can only be addressed with targeting analysts spot-checking sensor data (e.g., imagery) without any cues from the AI for the purpose of quality control and catching system mistakes. If missed targets are identified, those should be annotated and added to the DNN's training set for updating the model. Similarly, any false-positive detections should be added to the training set as negative examples so that the DNN can improve and learn from new data.

In addition to manually spot-checking sensor feeds, the classification confidence scores from the DNN can also be exploited to help mitigate the chances of a mistake by AI going unnoticed. Recall from Fig. 1 that DNNs provide a measure of how confident they are, in a probabilistic sense, about each target detection. A visual representation of this confidence score is depicted in Fig. 4. As indicated in the figure, decision thresholds, as indicated by the horizontal lines separating "Low, Moderate, and High", can be set for triggering certain actions. For instance, whenever the classifier is not highly confident about its classification, then perhaps those data samples are added to the targeting analyst queue for re-inspection. Furthermore, the ones in the moderate confident range should likely be stored and reviewed for potential addition to the DNN's training set given the model's uncertainty between high and low.

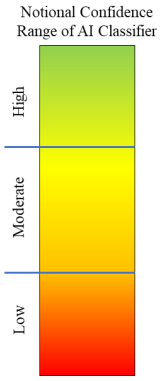


Figure 4. An illustration highlighting the fact that AI classifiers provide a confidence score with each detection, and that score can be scaled to resemble the probability of the classifier getting the classification correct. Thresholds along this range, which scales from 0 to 1, can help to define categories like "low, moderate, and high", which in turn can aid in designing the system to make intelligent decisions.

However, relying solely on static thresholds for deciding whether a data sample should be re-inspected is likely too simplistic and inflexible. For instance, if a single threshold – or dividing point between classifications like "high" and "low" – is used and it is set too high, then it is possible that too many targets will be added to the re-inspection queue, overwhelming targeting analysts. On the other hand, if set too low, then there is more risk of either mis-classified targets or false-positives passing to the next step without correction. The concept in fuzzy logic is to not hardcode single value thresholds as indicated by the blue horizontal lines in Fig. 4. Rather the idea is to program graceful transitions between the input classes of "low, moderate, and high" as indicated by the gradual change in color in Fig. 4. Fuzzy logic accomplishes this by allowing inputs – in this case, the confidence of the AI for a target type – to potentially belong to multiple sets, but with varying degrees of membership. For example, an AI classifier may output that it is 80% confident that it detected a particular target type, and in this case, fuzzy logic may then say that the confidence level belongs 85% to "High" and 15% to "Moderate",

depending on the way the pre-programmed "fuzzy sets" of "High, Moderate, and Low" are configured by the human-expert designer of the system.

Determining the level of acceptable risk for the AI to operate is the commander's decision. Therefore, the commander should be given the flexibility and option to assume more risk at times, if based on their best judgement the conditions merit it. For example, a commander may be risk averse when providing fire support in a counter-insurgency mission or in a dense urban environment with many civilians nearby, whereas a commander may be more risk tolerant if facing a high-intensity battle in mostly open terrain or performing final protective fires when friendly forces are facing the possibility of being overrun by the enemy.

To capture risk tolerance, commanders could be given the option of adjusting the AI classifier's confidence threshold like changing the positions of the vertical blue lines separating "Low, Moderate, and High" as shown in Fig. 4. However, it would not be intuitive for commanders to relate the amount of risk they are willing to accept to an AI's confidence threshold for classifying targets. Furthermore, if the system only has a single input parameter into its decision-making logic, such as only considering the AI's confidence, then the types (or complexity) of responses from the AI-enabled targeting system is limited and simple. Whereas, if the system is given multiple inputs, such as the AI's confidence and the commander's risk tolerance, then a more sophisticated response is possible because multiple factors can be considered in system's decision. Therefore, commanders should be given the ability to convey their risk tolerance directly to the system in a more meaningful way using a variable risk meter like the one shown in Fig. 5.

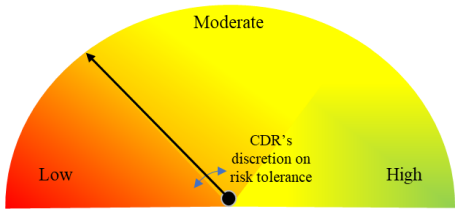


Figure 5. An illustration showing a rheostat-like knob in which commanders can use to relay their risk tolerance to the AI-enabled targeting system. A low risk tolerance translates to the commander being less willing to accept risk, meaning that the targeting system should be more cautious and require more human control and oversight in the process.

### 4.2 Decision-Making Logic within the Control System

Using the inputs above (classifier confidence and commander's risk tolerance), the goal is to produce a fuzzy logic controller that determines, in a rapid and automated fashion, the level of oversight and direct human involvement in the targeting steps shown in Fig. 3. In other words, the controller's purpose is to logically decide the optimal mix of human-machine interaction that should take place in the F2T2EA process based on the conditions of each target. Although a commander's (CDR) risk tolerance will likely remain constant for some time, it should be expected that the AI's confidence will vary with each target it detects. Hence, the recommendation by the controller on the level of oversight in the F2T2EA process could change for each target.

A fuzzy controller is designed to allow humans the ability to impart their decision-making logic into an automated control system via a set of pre-programmed rules. These human-derived rules then drive the action (or output) of the system based on the conditions of the inputs and the rule base logic. A block diagram for the proposed control system is shown in Fig. 6. As indicated in the figure, it is possible to add more than the two criteria into the decision-making process of the controller, but it is easier to illustrate the concept if the rule base is kept to only two dimensions. The rule base would be programmed into the controller's memory using a series of 'if-then' statements and would follow the logic of: *If 'AI Confidence' is LOW and 'CDR Risk Tolerance is LOW, then 'Human Involvement' is MAXIMUM,..., If 'AI Confidence' is HIGH and 'CDR Risk Tolerance' is HIGH, then 'Human Involvement' is MINIMUM.* Assuming two inputs with three categories each (Low, Moderate, and High), the complete set of nine rules can be derived by the two-dimensional rule base shown in Fig. 6.

The inputs and rule base were illustrated using heat maps in the figure for the purpose of implying the use of fuzzy logic in this design. In fuzzy logic, inputs can partially belong to multiple sets, rather than exclusively one

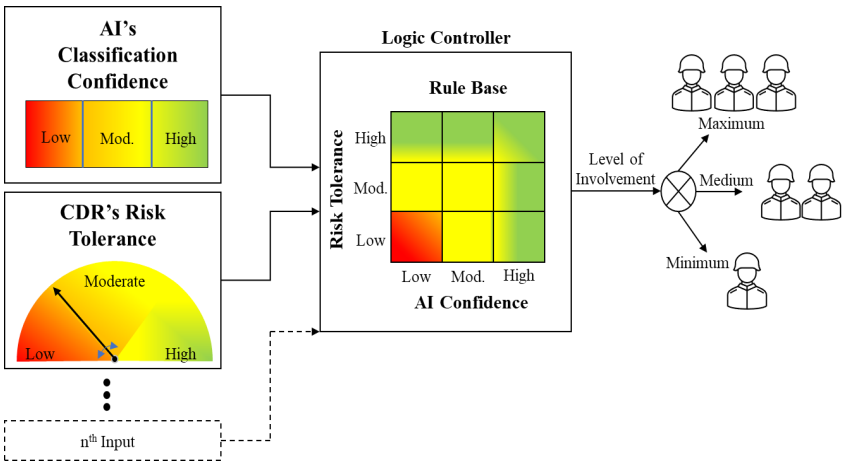


Figure 6. A block diagram for a programmable controller designed to gauge the confidence of AI and the risk tolerance of the commander for every newly detected target, and then logically decide the level of oversight that should take place during the targeting process.

set based on a hard cutoff. For instance, a measured input could be classified as partially belonging to two sets, such as 'moderate' and 'high', but with varying degrees of membership for each. This 'fuzzy' encoding technique avoids 'hard' thresholds – allowing the programmer to use more graceful 'soft' transitions, which in turn provides a smoother response (or output) from the controller as the inputs vary near decision boundaries.

The following example illustrates how fuzzy logic works in this use case; however, a more in-depth explanation of fuzzy logic and control is available from the cited references by Lilly and Lowrance.[25–27] For the purposes of this example, assume that the fuzzy sets shown in Fig. 7 is designed for the input "Commander's Risk Tolerance". Given the other input of "AI Confidence" also ranges from 0 to 100, identical fuzzy sets could be constructed for it. It is important to note the areas where the sets of "low", "moderate", and "high" overlap, as these are the input values which may partially belong to two sets. For instance, assume that the "Commander's Risk Tolerance" is measured to be 33 – which is the midway point along the y-axis. In this case, the input would classified as belonging to both "Low" and "Moderate" at 0.5 or 50% each. This same so-called fuzzification process takes place for the other input of "AI Confidence", and for the purpose of this example, assume that AI confidence is 66 which would then get classified as 50% belonging to both "Moderate" and "High".

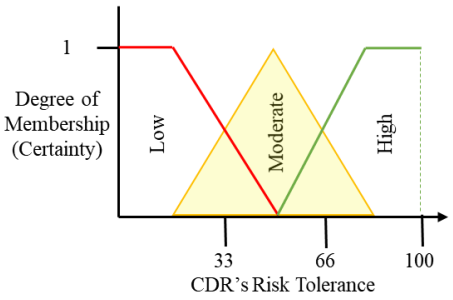


Figure 7. An example of fuzzy sets which could be designed for the inputs of "Commander's Risk Tolerance" and "AI Confidence". Given the overlapping nature of fuzzy sets, it is possible for inputs to be classified as belonging to multiple linguistic sets with varying degrees of membership.

After fuzzifying the inputs into human-understandable linguistic classes of "low, moderate, and high", the next step is to determine which rules in the pre-programmed "rule base" get activated and to what degree. In this case, there are four possible combinations between the two inputs: 1) confidence 'moderate' AND risk 'moderate', 2) confidence 'high' AND risk 'moderate', 3) confidence 'moderate' AND risk 'low', and 4) confidence

'high' AND risk 'low'. But each of these rules gets activated to a varying degree based on the earlier fuzzification step that assigned levels of belongingness to "low", "moderate", and "high" fuzzy sets.

Fuzzy set operations for the AND function of rule premises could consist of taking the minimum or product function of the two membership degrees. For instance, if "AI confidence" is moderate at 50% and "CDR risk tolerate" is moderate at 50%, then the product function would fire the consequent of the rule ("moderate involvement") with 0.5 x 0.5 = 0.25 degree of activation. The same process would take place for each of the other three rules and then taking those degrees of activation and multiplying them by the consequent value assigned by the programmer, such as 0 for minimum, 0.5 for moderate, and 1 to represent maximum levels of oversight. To determine the final output of the fuzzy controller, it is a matter of calculating the average of all four activated rules using the center-average or another method. In this case, if the computed average is below 0.25, then the decision from the fuzzy controller would be "minimum oversight", or "maximum oversight" if higher than 0.75, or otherwise "moderate oversight" if the output is somewhere between those boundaries. As this example illustrates, fuzzy logic provides a more sophisticated way of implemented rule-based logic with subtle transitions between the rules and averaging the effects of multiple rules rather than relying on classical set theory that concretely says a particular number either belongs or does not, which in this case would simply activate a single rule, not four rules for a more averaging effect.

For every newly detected target, the controller is designed to identify one of three different ways to supervise AI given it is now augmenting the targeting cycle. As indicated in Fig. 6, a target may be recommended for either a *maximum, moderate, or minimum* level of human involvement in the process. The corresponding three ways of potentially modifying the targeting process based on those output options are shown in Fig. 8. Each version of the targeting steps is shown to be aided by AI to some degree, but the main difference is the number of steps in which humans must verify the output from an AI-driven stage before proceeding to the next stage. Regardless, the final verification and authorization step is always reserved as a human-only decision.

The controller's decision for *maximum involvement* generally implies a human-driven targeting process, where humans the leading role in each stage, but as Fig. 8 suggests, it does not preclude AI from assisting in those steps for speedup. In other words, AI can augment any step, but a human must explicitly verify the output before the target proceeds. On the opposite extreme, *minimum involvement* translates into AI automating all steps, except for the final *validation and authorization* process, where a leader in the fires cell (e.g., Targeting Officer or Deputy Fire Support Coordinator) would review the targeting information and recommendations from AI before ordering to proceed with a fire mission. The *moderate oversight* process flow is somewhat more nuanced. It is similar to minimum oversight, except that the classification confidence of the AI algorithm and the risk assessment from integration stage must meet stringent thresholds; if a threshold is not met in either case, then a human must inspect the output generated by the AI algorithm.

## 5. CONCLUSION AND FUTURE WORK

When coupled with space-based sensors, AI will enable forces to "sense, understand, decide, and act" faster than ever before. The coordinated use of joint fires, enabled by a targeting process that is accelerated by AI will be instrumental in winning the next war. However, integrating AI into the targeting process is nontrivial given the need to balance the way AI interacts with humans so that it expediates targeting, yet does so in a responsible way that manages the inherent risks to life and the need to ensure appropriate human oversight.

A simplistic way of integrating AI would be to have humans re-verify everything the AI generates, but simply having AI perform these functions as an untrusted teammate whose work at every stage needs to be double-checked, is overly cautious and slow in some circumstances. There may be instances when the commander is willing to accept more risk and trust AI more for the benefit of speed, especially knowing that, as a failsafe measure, a targeting officer will make the final call on whether to proceed with a given fire mission.

The challenge is knowing when AI can be trusted to perform some targeting steps without the time delay of a human manually verifying. This is a function of obtaining a probabilistic sense of the AI's confidence in its work and understanding the commander's tolerance of risk given the current mission at hand. To do this, we developed a logic controller that weighs these inputs and then decides whether a target should proceed through one of three different processes – each with a different level of human oversight (i.e., minimum, moderate, or

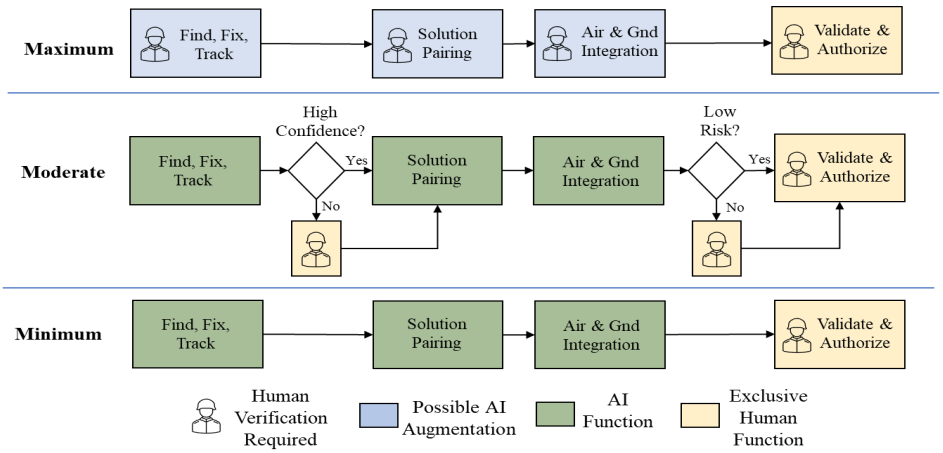


Figure 8. Three options for varying the level of human oversight in an AI-enabled targeting process. Each flow corresponds to either "maximum, moderate, or minimum" levels of oversight. The logic controller shown in Fig. 6 would determine which targeting process flow would be ideal based on the conditions of the AI's confidence and the CDR's risk tolerance.

maximum interactions). The design is based on principles of fuzzy control and is extensible to more inputs that can be added in the future to refine the decision logic.

There are several opportunities to evolve the work proposed in this paper. One option is to take the design discussed in this paper and experiment with it. Through such experimentation, much can be learned about its effectiveness towards accelerating the targeting process in a responsible manner. More specifically, the types and levels of interactions between the AI and the fires team could be studied in detail to further optimize them for speed and safety. It would be interesting to study the effect that the commander's risk tolerance input has on the trust of the system. For instance, does the ability to adjust the acceptable level of risk, which in turn changes the system's behavior, build more trust in the AI – as the system may be seen as being tunable and operating within its expected risk limits? Additionally, there is the possibility of converting the controller's decision logic, which decides the type and number of human interactions for a given target, from fuzzy control over to an AI control system. For example, the subdomain of machine learning known as reinforcement learning could be applied. In this case, AI could learn the optimal "interaction policy" through an iterative process of being rewarded or penalized based on whether its past decisions had to be corrected or not by a human.

## ACKNOWLEDGMENTS

The views presented in this paper are those of the authors and do not necessarily represent the views of Department of Defense or its components.

The authors would like to thank the U.S. Army's storied XVIII Airborne Corps and the Strategic Studies Institute of the U.S. Army War College for supporting this work.



## REFERENCES

[1] Lundy, M., "Large-scale combat operations," *Military Review: The Professional Journal of the U.S. Army*. **98**(5) (2018).

[2] Department of the Arny, "2019 Army Modernization Strategy: Investing in the Future." Army.mil, 2019 https://www.army.mil/e2/downloads/rv7/2019_army_modernization_strategy_final.pdf. (Accessed: 23 January 2022).

[3] Freedberg, S. J., "Army's new aim is 'decision dominance'." Breaking Defense, 17 March 2021 https://breakingdefense.sites.breakingmedia.com/2021/03/armys-new-aim-is-decision-dominance/. (Accessed: 23 January 2022).

[4] U.S. Department of the Army, "Army Doctrine Publication 3-19 Fires." Army Publishing Directorate, July 2019 https://armypubs.army.mil/ProductMaps/PubForm/Details.aspx?PUB_ID=1007599. (Accessed: 23 January 2022).

[5] David, R. A. and Nielsen, P., "Defense science board summer study on autonomy." Defense Technical Information Center, 01 June 2016 https://apps.dtic.mil/sti/citations/AD1017790. (Accessed: 23 January 2022).

[6] Dunnmon, J. and et al., "Responsible AI Guidelines." Defense Innovation Unit, 2021 https://www.diu.mil/responsible-ai-guidelines. (Accessed: 23 January 2022).

[7] Eigner, O. and et al., "Towards resilient artificial intelligence: Survey and research issues," in [*2021 IEEE International Conference on Cyber Security and Resilience (CSR)*], 536–542, IEEE.

[8] Comiter, M., "Attacking artificial intelligence: AI's security vulnerability and what policymakers can do about it." Harvard Kennedy School Belfer Center, August 2019 https://www.belfercenter.org/publication/AttackingAI. (Accessed: 23 January 2022).

[9] Shin, D., "The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI," *International Journal of Human-Computer Studies* **146**, 102551 (2021).

[10] Lewis, L., "Insights for the third offset: Addressing challenges of autonomy and artificial intelligence in military operations." Center for Naval Analyses, September 2017 https://www.cna.org/CNA_files/PDF/DRM-2017-U-016281-Final.pdf. (Accessed: 23 January 2022).

[11] Freeman, L., "Test and evaluation for artificial intelligence," *INSIGHT* **23**(1), 27–30 (2020). https://onlinelibrary.wiley.com/doi/pdf/10.1002/inst.12281.

[12] McDermott, P. and et al., "Human-machine teaming systems engineering guide." Defense Technical Information Center, 11 December 2018 https://apps.dtic.mil/sti/citations/AD1108020. (Accessed: 23 January 2022).

[13] Saenz, M. J., Revilla, E., and Simón, C., "Designing AI systems with human-machine teams," *MIT Sloan Management Review* **61**(3), 1–5 (2020).

[14] Lye, H., "DARPA looks to AI, algorithms to de-conflict airspace." Airforce Technology, 09 April 2020 https://www.airforce-technology.com/features/darpa-looks-to-ai-algorithms-to-de-conflict-airspace/. (Accessed: 23 January 2022).

[15] Arief, M., Glynn, P., and Zhao, D., "An accelerated approach to safely and efficiently test pre-production autonomous vehicles on public streets," in [*2018 21st International Conference on Intelligent Transportation Systems (ITSC)*], 2006–2011, IEEE (2018).

[16] Zhao, X. and et al., "Assessing safety-critical systems from operational testing: A study on autonomous vehicles," *Information and Software Technology* **128**, 106393 (2020).

[17] Bauchwitz, B. and Cummings, M. L., "Evaluating the reliability of tesla model 3 driver assist functions." Duke University, 01 October 2020, http://hal.pratt.duke.edu/sites/hal.pratt.duke.edu/files/u39/dms-min.pdf. (Accessed: 23 January 2022).

[18] Joint Chiefs of Staff, "Joint Publication 3-60 Joint Targeting." Department of Defense, 28 September 2018 https://www.jcs.mil/Doctrine/Joint-Training/Joint-Functional-Schools/JTS/. (Accessed: 23 January 2022).

[19] U.S. Department of the Army, "Army Techniques Publication 3-09.90 Division Artillery Operations and Fire Support for the Division." Army Publishing Directorate, October 2017 https://armypubs.army.mil/ProductMaps/PubForm/Details.aspx?PUB_ID=1003471. (Accessed: 23 January 2022).

[20] U.S. Department of the Army, "Army Techniques Publication 3-60 Targeting." Army Publishing Directorate, May 2015 https://armypubs.army.mil/ProductMaps/PubForm/Details.aspx?PUB_ID=105207. (Accessed: 23 January 2022).

[21] Goodfellow, I., Bengio, Y., and Courville, A., [*Deep learning*], MIT press (2016).

[22] Sarker, I. H., "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science* **2**(3), 160 (2021).

[23] Wood, T., "Softmax function." DeepAI, https://deepai.org/machine-learning-glossary-and-terms/softmax-layer/. (Accessed: 23 January 2022).

[24] McCullough, K., "One World Terrain (OWT) - Institute for Creative Technologies." University of Southern California, 2013, https://ict.usc.edu/research/projects/one-world-terrain-owt/. (Accessed: 23 January 2022).

[25] Lilly, J. H., [*Fuzzy control and identification*], John Wiley & Sons (2011).

[26] Lowrance, C. J. and Lauf, A. P., "An efficient fuzzy-based power control scheme for ad hoc networks," in [*2015 Wireless Telecommunications Symposium (WTS)*], 1–8, IEEE (2015).

[27] Lowrance, C. J., Lauf, A. P., and Kantardzic, M., "A fuzzy-based machine learning model for robot prediction of link quality," in [*2016 IEEE Symposium Series on Computational Intelligence (SSCI)*], 1–8, IEEE (2016).