Desktop Game with AI Support: Part II – Nuclear Weapons Research Site
AI as a Collaborator in Physical Protection and Response Planning

Lev Neymotin

1. Purpose and Context
This document completes the nuclear portion of the Desktop Game series by presenting **Part II**, in which a dedicated security-trained AI participates as a collaborator in planning and response. Unlike more advanced variants where AI may act autonomously or where an adversarial AI is introduced, this version keeps AI firmly in an advisory and analytical role. It supports human decision-makers but does not replace them.

The facility modeled in this exercise is the same 2-mile by 2-mile nuclear weapons research site used in Part I. It contains large storage facilities with significant quantities of highly enriched uranium (HEU) and metal plutonium. The scale of the site, the material attractiveness, and the geopolitical implications of compromise make it a high-consequence environment. The tabletop exercise is designed to test whether human-led security planning becomes measurably stronger when supported by a dedicated AI collaborator.

The central question of Part II is simple:
Does a well-trained AI collaborator reduce blind spots, compress decision time, and strengthen coordination without undermining human authority?

The exercise does not assume the answer. It is structured to discover it.

2. Facility Model and Security Environment
The modeled site covers approximately four square miles. It includes outer controlled areas, layered perimeter fencing, access control points, internal roads, research buildings, storage bunkers, vaults, and hardened facilities containing HEU and plutonium. The protected area contains multiple high-value storage structures separated by distance. The physical layout requires realistic travel times for response forces, which becomes an essential part of the game.

Security infrastructure includes perimeter intrusion detection systems, cameras, thermal imaging, lighting, motion sensors, access control systems, and alarm monitoring centers. Response forces are staged at defined positions. Communications operate through secure radio and network systems. Surveillance may include drones and fixed sensor networks, but all assets remain under human command.

The threat environment includes external assault, insider misuse of credentials, coordinated insider–external scenarios, and attempts to exploit confusion or delayed recognition. In this version of Part II,

the adversary does not use AI. The purpose is to isolate and evaluate the value of AI on the defensive side only.

3. Structure of the Table and the Role of the AI
The tabletop is organized around defined roles that mirror real-world authority. Participants typically include a security director or incident commander, a response force leader, a monitoring center supervisor, an operations representative, and a communications or cyber lead. Observers may attend, but decision-making authority remains concentrated in defined roles.

The AI is introduced as a dedicated security collaborator seated at the table. It has access, within the simulation, to the same sensor inputs and reporting streams available to human participants. It does not issue orders and does not control weapons, drones, or robotic systems directly. All actions require human authorization.

The AI's responsibilities include analyzing sensor patterns, identifying inconsistencies, estimating breach probabilities, suggesting response deployments, highlighting timing risks, and continuously calculating likely adversary routes and objectives. It may also flag potential cognitive bias–for example, when humans appear to dismiss an alarm as a false positive without sufficient evaluation.

The rules of engagement are explicit:
The AI may recommend.
Humans decide.
All recommendations and decisions are logged for later review.

This structure ensures that AI remains a collaborator, not a commander.

4. Planning Phase: Building the Baseline Security Architecture
The exercise begins with a structured planning session. Participants design or review the site's physical protection architecture, including sensor placement, patrol routes, response staging, communication protocols, and escalation thresholds. The AI participates by modeling detection coverage gaps, estimating response travel times, and identifying areas where simultaneous alarms could overwhelm the response force.

During this phase, the AI may simulate breach attempts and produce probabilistic heat maps showing which facilities are most vulnerable under current deployment patterns. It may also highlight areas where redundancy is assumed but not operationally effective, for example, where two sensors cover the same field but both depend on the same communication node.

Humans retain full authority. The purpose is not to accept AI suggestions automatically but to use them as structured prompts for better planning. The planning phase ends only when participants are satisfied that the security posture is internally consistent and operationally realistic.

5. Live Simulation Phase
Once the baseline plan is established, the facilitator introduces live event injects. These may include:

- A perimeter intrusion alarm at night.
- A vehicle approaching a restricted zone.
- Conflicting camera feeds.
- A suspicious badge access event.

- A communication disruption in one sector.

The clock starts at the first alarm. All decisions are time-stamped.

The exercise deliberately separates response into measurable segments: detection, recognition, decision, deployment, and interception. Recognition time–the moment between hearing an alarm and mentally committing to action–is treated as real and often decisive. The AI continuously tracks these intervals and may warn when hesitation threatens response effectiveness.

As information arrives, the AI provides structured recommendations. For example, it may suggest dispatching two response units along specific routes while simultaneously repositioning surveillance drones to verify a suspected diversion. It may warn that responding to one alarm leaves another sector exposed. It may calculate that if movement does not begin within a defined time window, interception probability drops below a critical threshold.

Humans may accept, modify, or reject these recommendations. The AI does not protest beyond providing reasoning and updated projections. The tension between human intuition and AI modeling becomes part of the learning process.

The simulation continues until the threat is resolved, neutralized, or reaches a defined consequence point.

6. Integration of Modern Technology
In this version of Part II, modern tools such as drones, robotic reconnaissance platforms, enhanced signaling systems, and integrated communication networks are available but always under human control.

The AI may recommend drone redeployment to close a detection gap or propose using a robotic platform to delay an intruder while human forces converge. It may suggest temporary lockdown of a sector based on risk modeling. However, execution depends entirely on human authorization.

The tabletop examines whether these technologies truly shorten response timelines or whether they introduce additional layers of decision complexity. The AI's role is to reduce that complexity by presenting options clearly and by filtering raw sensor data into operationally relevant insights.

7. Live Simulation Phase and the Reality of Mental Response Lag
Once the baseline security architecture is established, the exercise moves into the live simulation phase. At this point, planning assumptions are replaced by unfolding events. The facilitator introduces the first inject—typically an alarm, anomaly, or ambiguous indicator somewhere within the two-mile by two-mile research site. From that moment forward, time becomes an explicit variable. All actions and discussions are time-stamped.

The simulation deliberately separates the response process into distinct stages: detection, recognition, decision, deployment, and interception. Detection refers to the moment a sensor, camera, badge reader, or observer identifies something abnormal. Recognition is the psychological transition from "this is probably routine" to "this may be a real attack." Decision is the formal commitment to act. Deployment begins when forces or assets move. Interception marks physical engagement or neutralization.

In traditional exercises, response time is often measured from alarm to arrival. However, this model conceals a critical interval—the mental delay between hearing an alarm and accepting that it represents a credible threat. In real facilities, especially those that operate for years without experiencing an actual attack, personnel are conditioned by stability. Alarms are frequently benign. Equipment malfunctions occur. False positives are common. The human brain is wired to normalize irregularities before escalating them. This hesitation is not incompetence; it is cognitive economy. Yet in a high-consequence environment containing highly enriched uranium and metal plutonium, even seconds of hesitation can materially affect interception probability.

During the simulation, recognition time is treated as measurable. When an alarm occurs, the facilitator does not assume immediate mobilization. If participants debate, seek confirmation, reinterpret the signal, or attempt to downgrade the event before committing to action, that time is recorded. The purpose is not to penalize caution but to make visible the operational cost of disbelief.

The AI collaborator plays a structured role at this moment. Unlike humans, it does not experience surprise, denial, or normalization bias. Upon receiving the same alarm data, it immediately generates probabilistic threat assessments and projects likely adversary paths and timelines. If hesitation begins to consume the available response window, the AI may state that interception probability decreases beyond an acceptable threshold if deployment does not begin within a defined number of seconds. It cannot compel action, but it translates psychological delay into quantified operational risk.

This dynamic often produces the most instructive tension of the exercise. Humans may feel that waiting for confirmation is prudent. The AI may demonstrate that delay materially degrades defensive geometry. The discussion that follows is not theoretical; it occurs under a ticking clock. Participants must decide whether to move under uncertainty or risk losing the engagement entirely.
As additional information enters the scenario—secondary alarms, conflicting camera feeds, communications disruptions, or suspicious access events—the system becomes more complex. Redundancy in sensors may either strengthen confidence or produce ambiguity if signals conflict. The AI filters and synthesizes these streams, but humans remain responsible for commitment. The exercise therefore reveals whether the integrated system accelerates clarity or merely increases cognitive load.

The simulation continues until the event is resolved or reaches a defined consequence threshold. At the conclusion, the full timeline is reconstructed. Particular attention is paid to the recognition interval. Participants assess whether hesitation was justified, whether AI prompts shortened or lengthened deliberation, and how doctrine might be revised to reduce unnecessary delay without encouraging reckless escalation.

By explicitly modeling mental response lag within the live phase, the Desktop Game moves beyond conventional tabletop assumptions. It acknowledges that the most difficult vulnerability to quantify is not mechanical failure but human disbelief at the moment when disbelief is most costly. In doing so, it transforms cognitive shock from an invisible factor into a structured component of security analysis.

8. After-Action Review
The exercise concludes with a structured reconstruction of the entire event timeline, beginning with the first anomaly and ending with resolution or consequence. The review does not focus only on what decisions were made, but on when they were made and why. The timeline is rebuilt in segments—

detection, recognition, decision, deployment, and interception—so that the often invisible interval of mental response lag becomes visible and measurable.

Particular attention is given to recognition time. Participants examine how long it took to move from alarm awareness to genuine belief that a hostile act might be underway. They assess whether hesitation was operationally prudent or whether it reflected normalization bias, disbelief, or overconfidence in routine stability. This discussion is grounded in recorded timestamps rather than memory or impression. If seconds were lost in debate, reinterpretation, or search for confirmation, those seconds are accounted for explicitly.

The group then evaluates the role of the AI collaborator during those intervals. Did the AI provide clear, actionable recommendations early enough to influence outcomes? Did its probabilistic warnings meaningfully highlight the cost of delay? When humans chose to override or modify AI suggestions, did those decisions strengthen or weaken interception probability? In some cases, the AI may have identified threat patterns before human commitment occurred. In other cases, the AI may have overestimated risk or contributed to information overload. Both outcomes are treated as equally valuable data.

Communication performance is also examined. Participants analyze whether radio traffic, information flow, or hierarchical approval structures created bottlenecks that amplified recognition lag or slowed deployment. They review whether deployment routes were realistic given site geometry and whether forces moved as quickly in the simulation as doctrine assumed. Any gap between assumed mobility and simulated performance becomes a candidate for revision.

The AI's predictive models are compared directly with the actual scenario trajectory. If early AI modeling accurately forecasted adversary movement or highlighted sectors of vulnerability before humans acted, that finding informs future integration policy. If the AI's modeling proved inaccurate or distracting, that too shapes how it should be calibrated and used.

The objective of the after-action review is not to establish AI superiority or to defend human judgment. It is to refine the integrated human-AI system. By grounding discussion in measured intervals—especially the recognition and commitment phase—the review transforms cognitive shock from an abstract concern into a documented operational variable. The result is a clearer doctrine, more precise authority boundaries, and a security architecture that reflects not only technical capability but human psychological reality.

9. Learning Objectives and Strategic Insights

Part II is designed to reveal several structural truths about high-consequence security environments. First, human cognitive delay is real and measurable. Technology does not eliminate it automatically. Second, AI can compress analytical time but cannot replace authority or accountability.

Third, modern surveillance and robotic tools are only as effective as the doctrine that integrates them. Fourth, clarity of command structure is more important in an AI-supported environment than in a purely human one, because ambiguity multiplies when advisory inputs increase.

If Part I examined the limits of human-only security planning, Part II examines whether human-led security meaningfully improves when AI serves as a disciplined collaborator. The answer depends not on the sophistication of the algorithm but on how clearly authority, timing, and integration are defined.

10. Completion of the Nuclear Segment
With Part I and Part II complete, the nuclear portion of the Desktop Game framework now includes ***A human-centered baseline model*** and ***An AI-supported collaborative model*****.**

This structured progression allows organizations to compare performance across both configurations. It also prepares the foundation for future variants, should AI ever move beyond collaboration into greater autonomy or adversarial roles.

In this version, however, the principle remains firm:
AI supports.
Humans command.

Security performance is judged by measurable time, clarity of decision, and protection of high-consequence material.