



Against Strategic Surprise: AI-INT as a Synthesis-and-Simulation Layer for Actor Modelling and Strategic Warning

Citation

Tzukert, Nimrod (2026). Against Strategic Surprise: AI-INT as a Synthesis-and-Simulation Layer for Actor Modelling and Strategic Warning

Link

<https://dash.harvard.edu/handle/1/42756179>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access License Articles (IOAL), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

Against Strategic Surprise

AI-INT as a Synthesis-and-Simulation Layer for Actor Modeling and Strategic Warning

Abstract

Intelligence organizations often fail to anticipate strategic developments even when relevant information is available because signals remain fragmented, ambiguous, overwhelming, or filtered through entrenched assumptions. The September 11 attacks, Russia's 2022 invasion of Ukraine, and the October 7, 2023 Hamas attack illustrate how, even within some of the world's most capable intelligence organizations, available indicators do not reliably translate into effective strategic warning, with outcomes shaped by differences in collection, interpretation, dissemination, and uptake.

This article examines these failures and introduces AI-INT, or Artificial Intelligence Intelligence, as a response. AI-INT is an integrated, actor-specific synthesis-and-simulation layer within all-source analysis. It assembles and fuses relevant multi-source intelligence to build and update actor models, then uses them to run simulations at scale under systematically varied assumptions and conditions. It builds on wargaming's value in deepening understanding, challenging assumptions, exposing blind spots, and exploring alternative trajectories by combining fused intelligence with LLM- and ML-based methods to expand the scenario space.

As a bounded architectural illustration, the article presents SESBot, a Behavioral-AI framework for simulating the decision-making of selected Semi-State Terrorist Organizations (SESTOs). SESBot combines LLM-based agents with machine-learning behavioral anchors, illustrating one implementation of intelligence-grounded actor simulation within AI-INT.

AI-INT's promise lies not in predicting the future with certainty or replacing human judgment, but in helping analysts and decision-makers test assumptions, compare alternative scenarios, identify analytic blind spots, surface overlooked trajectories, connect them to monitorable indicators, and communicate uncertainty rigorously.

Keywords: AI-INT; Artificial Intelligence Intelligence; strategic surprise; intelligence failure; actor modeling; blind spots; scenario simulation; LLM agents; OSINT; SESBot; SESTOs.

Introduction: Strategic Surprise and the Limits of Traditional Synthesis

Are intelligence failures, as Betts (1978) argued, in some sense inevitable? Even if strategic surprise cannot be eliminated, can contemporary analytic capabilities reduce the analytic failures that contribute to it? Strategic surprise cannot be reduced to a failure of collection. In many cases, the relevant information is present, but the understanding needed to recognize what is unfolding and what may follow is not. Modern intelligence organizations operate in environments saturated with data from open and classified sources, including commercial imagery, financial indicators, social-media signals, intercepted communications, and human reporting. Yet information abundance does not guarantee effective warning. Relevant signals may remain fragmented across agencies, classified compartments, technical systems, and analytic frames, or be filtered through entrenched assumptions about what is possible, plausible, or rational.

This article examines a possible approach to narrowing this synthesis gap: strengthening the capacity to assess what dispersed and ambiguous signals may collectively indicate before events make their meaning clear.

Research on intelligence failure helps explain why this gap persists. Whether available information becomes effective warning depends not only on detection, but also on how evidence is interpreted, how prevailing assumptions are challenged, how uncertainty is communicated, and whether the resulting assessment receives serious institutional and political consideration (Betts, 1978, 1982; Handel, 1984; Heuer, 1999; Jervis, 2010; Wirtz, 2024). Strategic surprise may therefore occur even when individual indicators have been collected and recognized, if their combined significance is not understood, conveyed, or acted upon in time.

One established method for addressing precisely these interpretive and imaginative constraints is wargaming. By placing participants in the roles of relevant actors, requiring them to make decisions in an interactive setting, and confronting them with the responses and consequences generated by those decisions, wargames can expose assumptions that remain hidden in conventional analysis and surface trajectories excluded from the prevailing assessment (Perla & McGrady, 2011; Lin-Greenberg et al., 2022). Their principal value lies not in predicting a particular outcome, but in the disciplined exploration of how actors may perceive, decide, interact, and adapt under changing conditions (Schwarz, 2009). In this sense, wargaming helps analysts and

decision-makers ask not only whether a scenario is expected, but what would have to be true for it to occur, how others might respond, and which indicators would make it more or less plausible.

Yet conventional wargaming is difficult to conduct continuously and at scale. It requires time, expert participants, scenario design, facilitation, and adjudication, which necessarily limits the number of actors, assumptions, and scenario branches that can be examined in a single exercise. It may also be difficult to integrate and update large volumes of heterogeneous intelligence systematically across repeated games. These constraints do not diminish the analytic value of wargaming. Rather, they point to an opportunity to extend its underlying logic by grounding actor representations in fused multi-source intelligence, varying assumptions and conditions systematically, and exploring a much larger scenario space than human-led exercises can ordinarily sustain.

AI-INT, or Artificial Intelligence Intelligence,¹ is proposed here as such an extension. It is not another collection discipline or a standalone simulation technique, but an integrated, actor-specific synthesis-and-simulation layer within all-source analysis that connects fused multi-source intelligence to persistent actor models, simulation at scale, and iterative analyst interrogation (Section 4). Its purpose is to narrow the gap between information and understanding identified above.

The claim advanced here is deliberately bounded. AI-INT cannot eliminate strategic surprise, predict future decisions with certainty, or substitute for human judgment. Its potential contribution is narrower: to help analysts and decision-makers examine a broader range of scenarios, expose assumptions embedded in prevailing assessments, identify overlooked trajectories, and connect those trajectories to indicators that can be monitored as conditions change. This does not make AI-INT inherently objective; rather, it makes the assumptions and biases embedded in the analysis more visible, contestable, and subject to validation.

Recent advances in artificial intelligence make this conception increasingly feasible by expanding the practical capacity to integrate multiple source streams, build and update actor models, simulate scenario branches at scale, and generate comparative behavioral hypotheses. Research using LLMs

¹ Kendrick (2023) previously used AI-INT, or Artificial Intelligence Intelligence, to describe a proposed sixth intelligence discipline. The present article adopts the term but develops a narrower and distinct conception.

has shown that these models can simulate economic agents (Horton et al., 2023), replicate certain patterns observed in human-subject studies while also revealing systematic distortions (Aher et al., 2023), support interactive behavioral agents (Park et al., 2023), and simulate strategic interactions among state actors in historical conflict scenarios (Hua et al., 2023).

To make this argument concrete, the article presents SESBot as an architectural illustration. SESBot is a Behavioral-AI framework developed by the author to simulate the behavior and decision-making of selected Semi-State Terrorist Organizations (SESTOs). By combining multiple LLM-based agents with machine-learning algorithms that serve as behavioral anchors,² SESBot illustrates one possible implementation of AI-INT within a bounded, actor-specific simulation framework.

The article proceeds in four stages. First, it examines strategic surprise as a failure of synthesis and identifies the cognitive, institutional, and informational mechanisms that prevent available signals from becoming effective warning. Second, it situates AI-INT in relation to established intelligence disciplines and wargaming, defines its core architecture, and locates it within the intelligence cycle. Third, it explains why this function is possible now and presents SESBot as a bounded architectural illustration. Finally, it specifies requirements for validation, uncertainty communication, security, and human oversight, and considers how AI-INT could responsibly support intelligence analysis and decision-making.

² “Behavioral anchors” refers here to structured machine-learning components that constrain or inform the LLM-based simulation by reference to recurring empirical patterns, coded variables, or historically observed behavioral regularities. The term does not imply that the model mechanically predicts behavior.

Section 2: Strategic Surprise as a Failure of Synthesis, Not Only Collection

The synthesis problem is especially consequential for strategic rather than tactical warning. Tactical warning refers here to specific, time-sensitive, and actionable information about a bounded threat, such as an identifiable plot, target, or operational plan. When intelligence reaches this level of precision, it provides a relatively concrete basis for immediate action. Dahl (2013) argues that preventing surprise attacks generally requires specific tactical-level warning together with decision-makers willing to receive and act on it. Strategic warning, by contrast, concerns broader changes in an actor's intentions, priorities, capabilities, or emerging trajectory, often before their timing and form are known. It is therefore less bounded and depends more heavily on synthesizing ambiguous signals, challenging entrenched assumptions, and assessing possible developments under uncertainty. Although a missed tactical warning can be deadly and produce strategic consequences, its immediate object is generally a particular event or operation. A failure of strategic warning can leave decision-makers unprepared for a broader change in the security environment and narrow the range of responses available to them. It is this more synthesis-dependent form of warning that is central to the argument developed here.

Strategic warning does not simply succeed or fail at a single point; it may succeed in detection but fail in interpretation, communication, or decision uptake (Ikani, 2025). The synthesis gap is therefore not produced by a single mechanism. It may arise from how ambiguous evidence is interpreted, how organizations handle competing assessments, how political environments shape receptivity, or how meaningful signals are identified within overwhelming volumes of information. These mechanisms often overlap and reinforce one another, shaping which possibilities receive serious consideration and whether a warning reaches decision-makers with sufficient clarity and credibility. The following discussion separates them analytically while recognizing that strategic surprise usually results from their interaction.

Cognitive entrenchment and failure of imagination constitute a first category of synthesis failure. Analysts and decision-makers do not interpret information in a vacuum; they assess it through established assumptions, professional routines, and expectations about how an actor normally behaves. Experience and expertise are indispensable, but they can also reinforce confidence in familiar patterns, encourage ambiguous evidence to be interpreted in ways that preserve the prevailing assessment, and make alternative trajectories harder to take seriously.

Heuer (1999) and Jervis (2010) show how prior expectations and cognitive frames shape the interpretation of evidence, while Handel (1984) emphasizes the recurring difficulty of recognizing strategic surprise before it materializes. The resulting failure is therefore not necessarily that relevant signals go unnoticed, but that they are assigned meanings that exclude courses of action treated as implausible because they depart from the established picture.

Institutional and political mechanisms form a second category of synthesis failure. Intelligence assessments move through organizations with established hierarchies, divisions of responsibility, incentives, and thresholds for dissent. These structures shape which questions are examined, how competing interpretations are weighed, and whether uncertain or politically inconvenient warnings reach senior officials with sufficient clarity and urgency. Decision-makers may, in turn, discount assessments that conflict with existing policy commitments or appear insufficiently actionable. Warning may therefore fail even when analysts identify a plausible trajectory, if the assessment is weakened as it moves through the institution or if its recipients are unwilling to accept or act upon it (Betts, 1978, 1982; Ikani, 2025; Wirtz, 2024).

Informational mechanisms form the third category of synthesis failure. Intelligence indicators are rarely self-interpreting, and their interpretation is complicated by the volume and heterogeneity of contemporary collection. A military exercise may represent routine training, deterrent signaling, rehearsal, or preparation for attack; rhetoric, financial transfers, communications patterns, imagery, and social-media activity may likewise support several interpretations when viewed in isolation. At the same time, OSINT and other expanding digital sources generate additional noise, false positives, and competing explanations alongside potentially valuable evidence. The challenge is therefore not simply to collect more information, but to distinguish weak yet meaningful signals from the background, connect them across sources, and interpret them in relation to an actor's goals, constraints, capabilities, perceptions, and prior behavior.

These mechanisms rarely operate in isolation. Cognitive frames shape how ambiguous evidence is interpreted; information saturation increases reliance on familiar assumptions; and institutional incentives influence whether assessments that challenge those assumptions are preserved and communicated. Strategic surprise becomes more likely when these processes reinforce one another, narrowing the range of interpretations that receive serious consideration. The resulting

failure is not simply an absence of information, but an inability to convert dispersed evidence into a coherent, testable assessment of possible actor behavior.

Illustrative cases show how these mechanisms combine. The September 11 attacks involved a failure of imagination alongside institutional barriers to integrating and sharing available information (National Commission on Terrorist Attacks Upon the United States, 2004). Wyss's (2024) preliminary assessment, based on publicly available information, suggests that the October 7 surprise was not the result of a single failure, but of the combined effect of multiple problems at the political and intelligence levels. These included a mistaken political assessment that Hamas had been successfully contained, incorrect assessments of Hamas's intentions and capabilities, and problems in intelligence collection, analysis, and dissemination.³ Russia's 2022 invasion of Ukraine presents a contrasting, multilayered pattern. U.S., U.K., and other intelligence services converted extensive open-source and secret intelligence into timely warning, but receptivity and preparedness varied among Ukraine and NATO partners (Gustafson et al., 2024). On the Russian side, Putin's management of an intelligence system that discouraged dissent and dispassionate assessment contributed to misjudgments of Ukrainian resistance and the likely Western response (Dylan et al., 2023). Together, these cases show that the central challenge of strategic warning often lies less in the absence of relevant information than in interpreting what dispersed and ambiguous signals mean collectively, challenging the assumptions that shape that interpretation, and using that synthesis to develop a clearer assessment of what an actor might do.

This problem structure suggests a direct role for AI-INT. Multi-source fusion can reduce analytic fragmentation by bringing dispersed indicators into a common evidentiary frame. Actor models can organize those indicators around an actor's goals, constraints, capabilities, perceptions, and prior behavior. Repeated simulations across systematically varied assumptions and conditions can then test how different interpretations alter the range of plausible actor responses. Analysts and decision-makers can interrogate the resulting outputs to examine why a trajectory emerges, which evidence supports it, and which indicators would make it more or less plausible. AI-INT cannot eliminate cognitive bias, institutional resistance, or uncertainty, but it can make alternative interpretations and their underlying assumptions more explicit, comparable, and contestable.

³ Wyss emphasizes that the precise causes of these misjudgments remain to be determined and that no monocausal explanation is likely.

Section 3: The Existing Intelligence Taxonomy and Its Limits

The modern intelligence taxonomy is built around distinct collection disciplines. HUMINT, SIGINT, OSINT, GEOINT, and related categories describe how information is obtained, processed, and transformed into intelligence. Official and professional accounts define these disciplines primarily through collection method and source environment, with the analytic question playing no defining role in the taxonomy (Henrico & Putter, 2025; Lowenthal, 2025; ODNI, n.d.).

This taxonomy remains essential. Intelligence collection cannot be reduced to a single information environment, and different source types provide different kinds of evidentiary access. HUMINT may provide insight into intentions and internal debate. SIGINT may reveal patterns of communication and operational activity. GEOINT, including imagery-based collection, may show physical movement, infrastructure, and capabilities. OSINT may expose publicly available discourse and data streams, while SOCMINT covers intelligence derived from social-media environments (Henrico & Putter, 2025; ODNI, n.d.; ODNI & CIA, 2024).

The limitation is not that these disciplines are obsolete, or that they cannot warn. A single HUMINT source, a SIGINT intercept, or an imagery finding may sometimes generate a clear warning. The limitation is that the taxonomy is primarily collection-oriented. It describes how information enters the intelligence system, but it does not fully describe the analytic function that converts multiple source streams into actor-specific, forward-looking behavioral hypotheses.

AI-INT draws on target-centric analysis (Clark, 2004), structured analytic techniques, agent-based modeling, and analytic wargaming, each of which addresses part of the synthesis problem identified above. It is proposed not as another collection discipline or as a replacement for these approaches, but as a persistent synthesis-and-simulation layer that integrates their complementary logics within all-source analysis.

Section 4: Defining AI-INT

A Working Definition

AI-INT, or Artificial Intelligence Intelligence, is a bounded, iterative analytic function within all-source analysis in which fused multi-source intelligence informs AI-supported models of specific actors, and those models are used to simulate actor responses at scale across systematically varied assumptions and conditions. Its outputs are comparative behavioral hypotheses for analyst interrogation and evaluation and, where appropriate, decision-maker consideration; they are not point predictions.

This definition draws on established work in multi-source data fusion, agent-based modeling, and the emerging literature on LLM-based behavioral simulation (Aher et al., 2023; Bonabeau, 2002; Dasarathy, 1997; Horton et al., 2023; Park et al., 2023).

The term is proposed here as a working designation rather than as a settled institutional category. Its purpose is to describe an emerging intelligence function that is not fully captured by existing collection-based disciplines. Whereas HUMINT, SIGINT, IMINT, OSINT, GEOINT, and FININT are primarily defined by how information is obtained, AI-INT is defined by how information from these disciplines is synthesized and used to generate structured, forward-looking assessments about actor behavior.

The double use of intelligence in the term is intentional.⁴ AI-INT refers both to a method, the application of artificial intelligence to structured actor modeling and simulation, and to a product, the resulting intelligence hypothesis. This dual meaning does not imply that every application of AI to intelligence constitutes AI-INT. AI-INT's contribution lies not in a new algorithm or simulation technique, but in an analytic architecture defined by four integrated components: provenance-preserving fusion of relevant multi-source intelligence; construction and updating of persistent actor-specific models; systematic simulation and comparison of actor behavior across varied assumptions and conditions; and iterative analyst interrogation, validation, and revision of the models, assumptions, and outputs. Taken together, these components establish the boundary of the concept. Systems that apply AI to isolated elements of this architecture without integrating

⁴ The phrase "Artificial Intelligence Intelligence" is intentionally redundant in the same sense that intelligence terminology often uses the suffix "INT" to designate both a production pathway and an intelligence product. The first "intelligence" refers to artificial intelligence as a technological method; the second refers to intelligence as an analytic product.

all four are better understood as AI-assisted analysis, predictive analytics, or simulation rather than as AI-INT in the sense proposed here.

What AI-INT Is Not

AI-INT is not merely AI-assisted analysis. Intelligence organizations already use AI for translation, summarization, pattern recognition, document triage, and other analytic-support tasks. Recent treatments of AI in intelligence and national security describe its growing use across collection, processing, exploitation, operations, and analysis, while also emphasizing augmentation, human-machine teaming, and the continuing limits of automation (Katz, 2020; Moran et al., 2023; ODNI, 2019). Those uses are important, but they remain tools within existing workflows. AI-INT refers instead to the integrated architecture defined above. Its distinctiveness lies in the integration of the four components rather than in any isolated AI-enabled task.

AI-INT is also not conventional predictive analytics. Predictive analytics often relies on statistical patterns in historical data to estimate the likelihood of future events. AI-INT may use such methods, but its distinctive contribution is actor-specific modeling: representing the goals, constraints, incentives, past behavior, internal structure, and decision logic of a particular actor.

AI-INT is also not conventional wargaming. It nevertheless builds directly on wargaming's analytic rationale: the structured exploration of actor behavior across alternative conditions, which can deepen understanding, challenge prevailing assumptions, and expose possibilities that conventional analysis may overlook (Perla & McGrady, 2011; Lin-Greenberg et al., 2022). Davis and Bracken (2025) provide an important bridge between this logic and AI-INT. They argue that modeling, simulation, and wargaming should be used together, that AI can contribute to each, and that these activities should be integrated with analysis. Their political-military research agenda includes alternative adversary models, massive scenario generation, iterative updating, and decision aids, and emphasizes exploring possibilities under deep uncertainty rather than reliable prediction. AI-INT's distinctive proposition lies in how this exploratory logic is connected to intelligence practice. Many conventional analytic wargames rely on human participants who are briefed to represent particular actors within a bounded exercise. AI-INT instead instantiates actor representations in persistent AI-supported models grounded in fused multi-source intelligence and updated as that intelligence changes, while preserving source provenance and uncertainty. These models can be run repeatedly across systematically varied assumptions and conditions, allowing

AI-INT to explore and compare a volume of scenario branches far beyond what human-participant wargames can realistically sustain. The proposed contribution of AI-INT is therefore to operationalize wargaming's exploratory logic through intelligence-grounded actor models and simulation at scale within all-source analysis.

Finally, AI-INT does not require artificial general intelligence. It requires bounded, validated systems capable of modeling specific actors and scenarios with enough fidelity to support intelligence judgment. For that reason, corpus discipline, retrieval-grounded generation, evaluation, and uncertainty communication are core design requirements (Moran et al., 2023; Ji et al., 2023; Lewis et al., 2020; NIST, 2023).

The Core Mechanism

AI-INT operates through four interconnected processes:

Multi-source fusion. AI-INT consumes outputs from other intelligence disciplines, including open sources, human reporting, signals, imagery, geospatial data, financial data, cyber indicators, and other relevant sources. The fusion layer is not merely aggregation; it is the disciplined reconciliation of heterogeneous evidence, uncertainty, source reliability, and provenance into an integrated analytic representation (Dasarathy, 1997; Hall & Llinas, 2008; Waltz, 2003).

Actor modeling. On the basis of this fused analytic representation, AI-INT constructs and updates a bounded, structured model of a selected actor, whether a state, non-state organization, institution, faction, or influential individual. The model represents the actor's decision logic, historical behavior, goals, constraints, capabilities, perceptions, and organizational dynamics. Such a model is an analytic approximation rather than a complete representation of the actor. This requires more than a rational-actor template: actor behavior may reflect organizational routines, bureaucratic politics, misperception, internal factionalism, and institutional cohesion, and cannot be reduced to a single unified preference function (Allison & Zelikow, 1999; Jervis, 2010). Recent work shows that LLMs can be prompted to simulate participants under specified demographic characteristics, roles, and experimental conditions, and that comparison with human-subject benchmarks can reveal both replicated behavioral patterns and systematic distortions (Aher et al., 2023). Related work extends this approach to simulated economic agents with specified incentives and decision settings (Horton et al., 2023).

Scenario simulation. Given an actor model and an updated representation of its environment, AI-INT runs simulations at scale across systematically varied assumptions and conditions to compare actor responses and surface plausible trajectories. The output should not be a single prediction, but a set of alternative behavioral hypotheses, explicit assumptions, and confidence assessments that analysts can interrogate. Because actor behavior is context-dependent, these simulations must also represent the situational and social context in which decisions are made, including roles, constraints, expectations, shared knowledge, and relevant norms. This logic builds on long-standing agent-based and geopolitical simulation research as well as newer work on generative agents, LLM-based political or strategic simulations, and AI-supported contextual analysis (Bonabeau, 2002; Epstein, 2008; Macal & North, 2010; Neuman, 2024; Park et al., 2023; Solopova et al., 2026; Yu et al., 2024).

Analyst interrogation, validation, and revision. AI-INT operates as an iterative analytic loop. Analysts can interrogate why particular trajectories emerge, examine the evidence and assumptions on which they depend, compare model outputs with competing assessments, and identify gaps requiring additional collection. They can then incorporate new intelligence, revise the actor model or scenario conditions, and rerun the analysis. This loop keeps the models, assumptions, and outputs contestable and responsive to changing evidence while preserving human responsibility for analytic judgment. Such analytic validation complements, but does not replace, formal empirical and technical validation of the system.

Section 5: AI-INT in the Intelligence Cycle

Traditional intelligence-cycle models describe a sequence of planning and direction, collection, processing, all-source analysis-synthesis and production, and dissemination (Waltz, 2003). AI-INT adds a dedicated actor-modeling and simulation capability within the analysis-synthesis stage, converting fused intelligence into structured, forward-looking behavioral hypotheses. Recent scholarship and policy analysis on AI and intelligence similarly emphasize that AI may affect collection, processing, exploitation, and analysis across the intelligence cycle, but does not remove the need for human interpretation, explainability, accountability, and analytic judgment (Katz, 2020; Moran et al., 2023).

The relationship between AI-INT and existing intelligence disciplines is complementary and integrative. Each discipline contributes a distinct form of evidence to the actor model, with source provenance and uncertainty preserved throughout the process (Dasarathy, 1997; Hall & Llinas, 2008; ODNI & CIA, 2024).

AI-INT's contribution is to take these inputs and integrate them systematically into a structured, conditional assessment of how specific actors may behave, why they may behave that way, and how the assessment changes under alternative assumptions. Such assessments should maintain source transparency, distinguish likelihood from analytic confidence, and support disciplined consideration of alternatives (Kent, 1964; ODNI, 2015).

This does not eliminate the need for human judgment. On the contrary, AI-INT is most useful when analysts can interrogate the model, test assumptions, compare alternative scenarios, and identify where additional collection would most improve the assessment. The human analyst remains responsible for evaluating model outputs, communicating uncertainty, and integrating analytic judgments into policy-relevant products.

Figure 1 illustrates this relationship. AI-INT supports the analyst by converting multi-source intelligence into structured hypotheses and simulation-based queries, while the analyst retains responsibility for validation, interpretation, and dissemination.

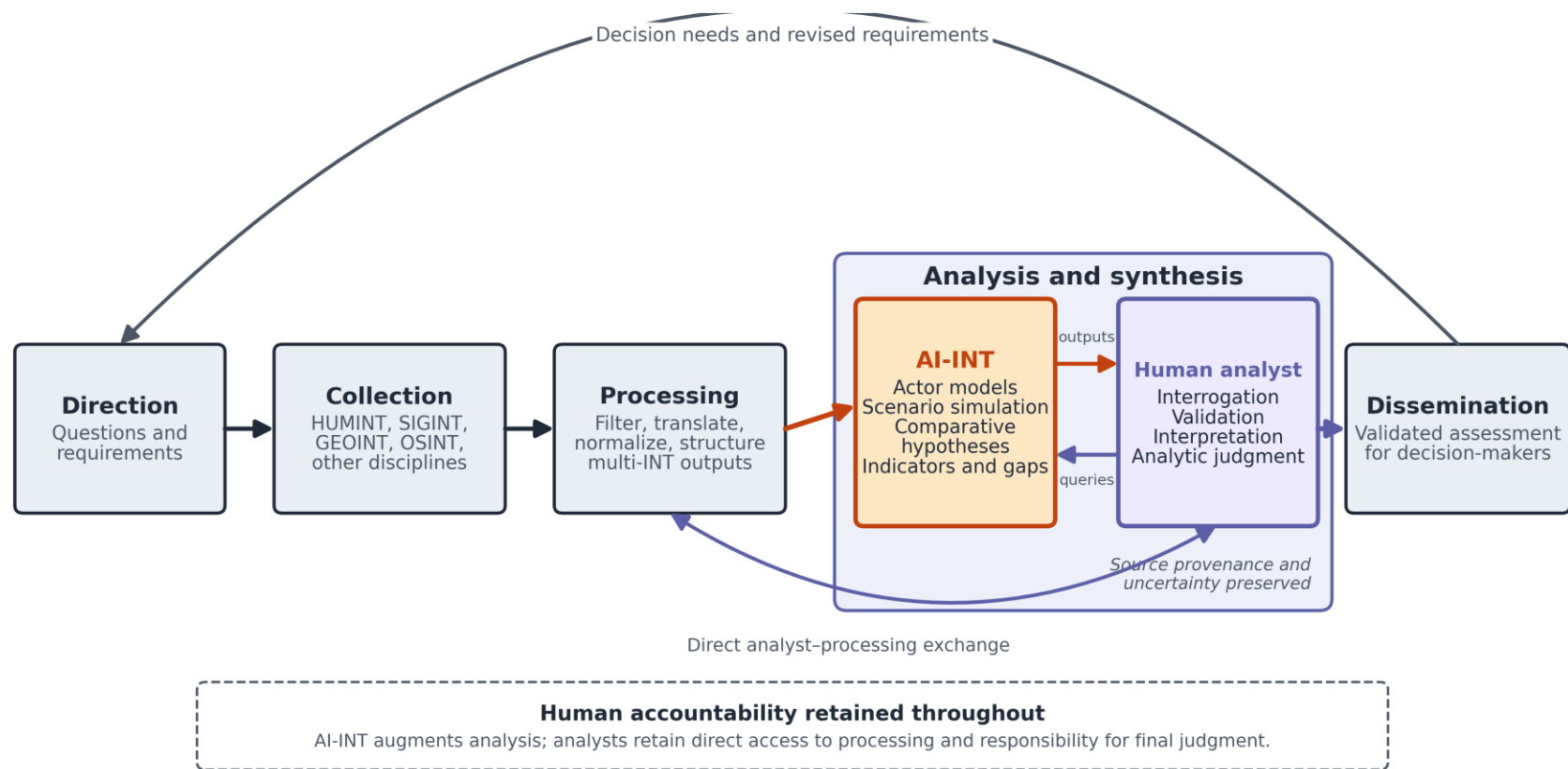


Figure 1. AI-INT within the intelligence cycle: direct analyst-processing interaction and the division of labor between AI-INT and the human analyst.

Section 6: Why AI-INT Is Possible Now

The idea of AI-supported intelligence forecasting is not entirely new. What is new is that several enabling conditions have converged in ways that make the function more operationally plausible: the expansion of OSINT, the maturation of large language models, the growth of multi-agent and simulation methods, and the computational feasibility of large-scale scenario exploration (Macal & North, 2010; Moran et al., 2023; Park et al., 2023; Solopova et al., 2026; Yu et al., 2024).

The OSINT Explosion

Open-source intelligence has expanded dramatically. Publicly available information now includes social media, digital communications, commercial satellite imagery, vessel and aircraft tracking, financial transparency data, and other real-time or near-real-time sources. Among these sources, leaders' speeches and other public communications can provide behaviorally relevant evidence. Studies using operational-code and leadership-trait analysis have linked profiles inferred from such material to observed foreign-policy conduct and, in some cases, found them consistent with subsequent policy decisions (Walker et al., 1999; Özdamar et al., 2023). The Intelligence Community OSINT Strategy describes OSINT as a core intelligence discipline and highlights the need to professionalize its use, govern it, and integrate it with other sources (ODNI & CIA, 2024). Broader academic work likewise treats OSINT as an increasingly consequential intelligence practice, while cautioning against treating it as a simple technological revolution (Van Puyvelde & Tabárez Rienzi, 2025).

The Maturation of Large Language Models

Large language models can perform a wide range of language tasks and adapt to natural-language instructions or a small number of demonstrations without task-specific fine-tuning (Brown et al., 2020). Research on generative agents and LLM-based multi-agent systems further suggests that LLMs can support planning, role-based simulation, and structured interaction when appropriately constrained and evaluated (Guo et al., 2024; Park et al., 2023). Recent work has used LLMs to simulate economic agents and participants in human-subject studies (Aher et al., 2023; Horton et al., 2023). Emerging research has examined voter decision-making through large-scale LLM-based election simulations (Yu et al., 2024) and LLM decision behavior in structured geopolitical crisis simulations (Solopova et al., 2026). Relatedly, Johnson (2026) conceptualizes the use of

GenAI to simulate adversary intentions, escalation pathways, and imagined futures under uncertainty as “synthetic foresight,” warning that, in nuclear crises, such simulations may shape decision-makers’ perceptions and behavior and generate self-reinforcing escalation dynamics. Taken together, these developments make AI-supported actor simulation increasingly operationally plausible, although they do not by themselves establish its validity. For AI-INT, their significance lies in enabling repeated, structured exploration of actor behavior across varied assumptions and conditions, subject to careful grounding, validation, and human oversight.

The Computational Feasibility of Simulation at Scale

Agent-based modeling and geopolitical scenario simulation have existed for decades. What has changed is not the basic idea of simulation, but the computational and methodological feasibility of running many more scenario branches with richer data, more detailed actor models, faster iteration, and more systematic post-run analysis. Earlier systems such as the Advanced Global Intelligence and Leadership Experiment (AGILE),⁵ an agent-based simulation environment designed to help intelligence analysts model geopolitical crises, already allowed users to vary initial conditions, run multiple simulations, and explore alternative futures. However, the AGILE case required substantial manual model specification and developer effort (Taylor et al., 2004). Contemporary AI-INT could build on that foundation by combining scalable simulation, retrieval-grounded corpora, and LLM-based multi-agent systems, potentially reducing the manual effort required to construct and revise actor models and enabling faster testing of alternative assumptions across a broader scenario space.

⁵ AGILE represented countries as groups of decision-making agents interacting through a shared world model of political, economic, military, diplomatic, and social processes. In the Taiwan Strait demonstration, the model included approximately ten agents and fifteen interconnected variables and required roughly one month of software-development effort to build, test, run, and analyze. Validation remained primarily based on analyst face validation rather than systematic quantitative testing.

Section 7: SESBot as an Architectural Illustration of AI-INT

Purpose and Scope

SESBot, the Semi-State Terrorist Organization Bot, is a Behavioral-AI prototype developed by the author. It is included in this conceptual article for a limited and explicitly illustrative purpose: to demonstrate how AI-INT could be instantiated in an actor-specific simulation architecture. SESBot is not presented here as empirical evidence that AI-INT improves strategic warning, and this article does not report a formal evaluation of its performance. A full account of the prototype's corpus construction, model specification, implementation, validation, empirical performance, and generalizability is reserved for a separate empirical article. Within this limited scope, SESBot addresses a bounded analytical question: how might a selected Semi-State Terrorist Organization respond to changing strategic conditions or to a negotiation offer, and which organizational factors might shape that response? The system is designed to generate structured, actor-specific behavioral hypotheses and accompanying rationales, rather than a single authoritative prediction. Its intended analytic value lies in helping analysts and decision-makers design negotiation strategies by comparing plausible actor responses, examining the assumptions and organizational factors underlying those responses, and identifying alternative interpretations, missing variables, or analytic blind spots.

Why SESTOs?

SESBot focuses on Semi-State Terrorist Organizations, or SESTOs: non-state armed organizations that combine terrorist activity with territorial control, civilian governance, and strategic interaction with state actors. SESTOs were selected as the prototype's initial use case because they combine two characteristics that make actor-specific simulation especially consequential. On the one hand, prior research finds that they sustain significantly higher attack frequencies and aggregate fatalities than other terrorist organizations (Tzukert et al., 2025). On the other hand, they are unusually frequent participants in strategic negotiations with states (Tzukert & Sulitzeanu-Kenan, 2026). This combination makes designing an effective, actor-specific negotiation strategy a central intelligence and policy challenge. The task is further complicated by the fact that negotiations with terrorist organizations may continue alongside violence (Cronin, 2010), while extremist attacks can erode trust and derail agreement implementation (Kydd & Walter, 2002). Together, these characteristics make SESTOs a demanding test case for actor-specific simulation, requiring the

model to account simultaneously for organizational violence, governance responsibilities, and strategic bargaining.

Architecture

SESBot uses a hybrid, multi-agent architecture. Structured machine-learning components identify recurring empirical patterns and relationships among relevant variables, providing behavioral anchors, while LLM-based agents contribute contextual interpretation and qualitative scenario reasoning. The architecture does not rely on prompt-only generation. Instead, it combines curated actor-specific corpora, retrieval, structured variables, and a defined division of labor between statistical anchoring and generative reasoning. This arrangement is intended to constrain generative outputs with structured behavioral evidence while preserving the ability to interpret changing contexts and explore alternative scenarios. It cannot eliminate hallucination or generation-faithfulness risks, but it is intended to reduce the likelihood that fluent outputs drift away from the empirical structure on which the actor model is built. The design draws on work concerning retrieval-augmented generation (Lewis et al., 2020), hallucination and generation faithfulness (Ji et al., 2023), generative and multi-agent systems (Park et al., 2023), and the validation of LLM-based behavioral simulations against human-subject benchmarks (Aher et al., 2023). Figure 2 summarizes SESBot’s hybrid architecture and the flow from scenario input through information-channel LLM agents and the ML behavioral anchor to fusion, simulation, and analyst-facing output.

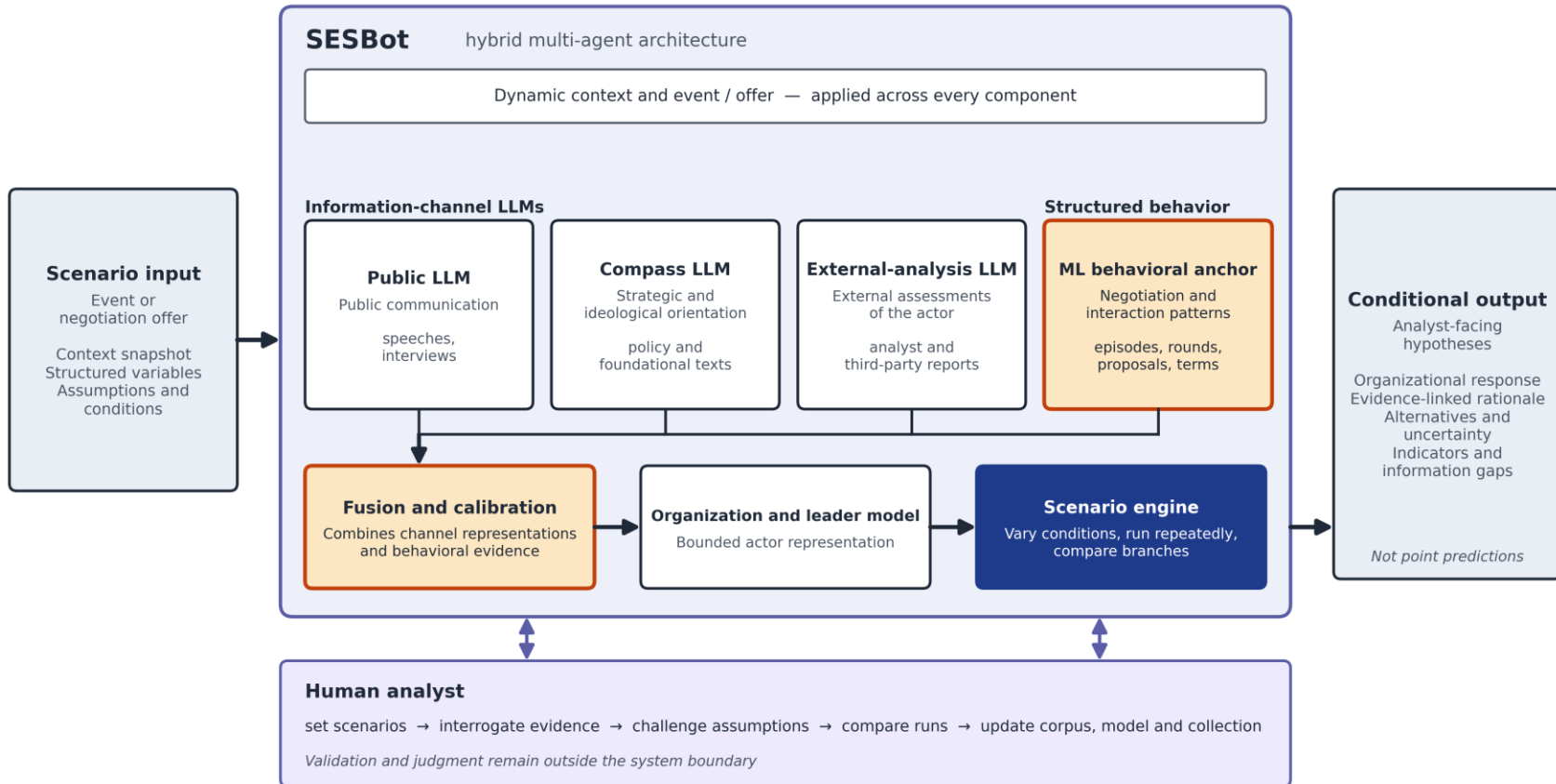


Figure 2. SESBot: a hybrid multi-agent architecture. The Public LLM, Compass LLM, and External-analysis LLM each embody a distinct information channel about the same actor: public communications, strategic and ideological orientation, and external assessments, respectively. An ML-based behavioral anchor captures structured historical patterns of negotiation and interaction. Fusion and calibration integrate these components into a bounded organization-and-leader model.

Analytic Output and Use

For a defined event or negotiation offer, SESBot is designed to generate a simulated organizational response accompanied by a stated rationale identifying the assumptions, contextual variables, and organizational considerations associated with that output. Analysts can compare outputs across alternative scenarios, examine which assumptions produce different responses, contrast the results with their own assessments, and identify disagreements or overlooked possibilities.

The stated rationale should not be treated as transparent access to the model's internal reasoning or as evidence that the output is correct. Its purpose is to make the assessment more open to interrogation by the analyst. The traceability, faithfulness, reliability, and analytic value of these outputs must be evaluated empirically in the separate study.

Modeling Organizations Beyond Individual Leaders

SESBot distinguishes between modeling an individual leader and modeling an organization. In the SESTO context, leadership preferences are important, but they do not by themselves represent organizational behavior. Decisions may also be shaped by religious beliefs, organizational culture, founding principles, bureaucratic routines, internal factions, command structures, material capabilities, and the constraints associated with governing territory and civilians.

Individuals may also behave differently when occupying particular institutional roles because organizational incentives, decision procedures, and role expectations affect how preferences are expressed and translated into choices. Classic role-theory research similarly indicates that changes in institutional position can produce systematic changes in the attitudes of role occupants (Lieberman, 1956). Accordingly, SESBot treats a leader profile as one component of the actor model without treating it as a proxy for the organization as a whole. This design choice is consistent with scholarship on organizational process, insurgent cohesion, rebel governance, and covert organizational control (Allison & Zelikow, 1999; Mampilly, 2011; Shapiro, 2013; Staniland, 2014). In SESBot, this distinction is implemented through a multi-agent architecture that represents both individual leaders and organization-level decision processes, allowing the simulation to incorporate interactions among leadership preferences, organizational roles, constraints, and decision pressures.

Section 8: Standards, Risks, and Responsible Use

The Limits of Simulation

AI-INT systems do not predict the future in a deterministic sense. They generate structured hypotheses about possible actor behavior across alternative scenarios, based on available information, modeled assumptions, and simulations run at scale. Their outputs should therefore be communicated with explicit uncertainty and reviewed by human analysts or decision-makers. The central risk is not only that the model may be wrong, but that users may overinterpret its outputs as more certain, complete, or authoritative than they are.

Validation Requirements for AI-INT Systems

Validation standards for AI-INT systems should reflect their intended analytic use. Systems designed to generate hypotheses should be assessed according to whether they surface plausible alternatives, make assumptions explicit, identify relevant indicators, and improve structured analytic reasoning. Systems intended to approximate actor behavior or expert judgment require more demanding evaluation, including historical back-testing, out-of-sample testing, comparison with expert benchmarks, inter-rater reliability, calibration and classification metrics, ablation studies, and robustness testing under alternative assumptions and inputs. When historical back-testing involves pretrained LLMs, validation should impose a documented temporal cutoff on the evidence supplied to the system and explicitly test for post-cutoff knowledge embedded in the model, so that apparent performance is not inflated by hindsight leakage. Reproducibility and auditability also require recording the model and version used, the prompts, the retrieval-corpus snapshot, the sampling parameters, and variation across repeated runs.

Validation should also distinguish between several different questions: whether a system produces plausible outputs, whether it reproduces observed behavioral patterns, whether it surfaces plausible actor decisions or responses that its users had not previously considered, whether its explanatory rationales are traceable to relevant evidence, and whether it adds value relative to simpler models or unaided human analysis. Success on one of these dimensions should not be treated as evidence of validity on the others.

Analytic and Institutional Risks

Human review, although necessary, does not by itself eliminate the risks of AI-INT. AI-generated assessments may be treated as more objective or authoritative than their evidentiary basis warrants, reinforcing existing biases and limiting critical interrogation. In intelligence settings, secrecy may further restrict internal and external scrutiny, making problematic assumptions and biased classifications harder to identify or contest (Bellaby, 2024). In nuclear crisis settings, simulated futures may also shape decision-makers' perceptions and behavior, creating self-reinforcing dynamics between anticipation and action (Johnson, 2026). Responsible use therefore requires source-level traceability, audit trails, structured adversarial review, explicit separation between model output and analytic judgment, and procedures for preserving dissent.

AI-INT cannot make decision-makers receptive, and nothing in the architecture addresses the political incentives that lead warnings to be discounted. Its bearing on the institutional mechanism is narrower and concerns the form in which assessments travel through an organization. First, the system generates, records, and links alternative trajectories to evidence. As a result, a minority assessment survives as a documented artifact even when it does not prevail. This partially answers Handel's (1984) objection that the devil's advocate is an artificial device that soon becomes routinized, though not his further objection that an in-house advocate reproduces the group's own image of the adversary. Second, because each trajectory is tied to monitorable indicators, the output supports graduated rather than binary warning, which bears on the mobilization dilemma Handel (1984) describes: the choice between costly alerts and inaction becomes a question of which indicators have moved. Third, the audit trail makes it possible to establish, after the fact, whether a trajectory that materialized had been generated and at what point it was set aside. None of this guarantees uptake. It changes what can be suppressed silently.

Actor Model Fidelity

The quality of AI-INT depends on the fidelity of the actor model. A weak model may reproduce stereotypes, overfit historical patterns, or confuse rhetoric with decision logic. Validation should therefore examine whether the model adequately represents leadership preferences, organizational constraints, internal factions, material capabilities, governance obligations, and external patronage, and whether its outputs remain robust when those assumptions change. For complex organizations,

validation should test both the model's representation of organization-level decision processes and its treatment of leadership preferences.

Pattern Departure and the Risk of Confirmation

A structural objection follows from Section 2. Actor models are built from prior behavior, yet strategic surprise typically arises when an actor departs from its established pattern. A system anchored in past conduct may therefore re-encode the prevailing assessment in computational form, reproducing the entrenchment it is meant to counter (Handel, 1984; Heuer, 1999). Three features of the proposed architecture bear on this risk. Behavioral anchors constrain the actor's decision logic rather than the conditions under which it is exercised; since many departures from pattern reflect stable goals applied to changed constraints or perceptions, systematic variation of scenario conditions is the primary means by which the simulation space extends beyond the historical record. The multi-agent representation of internal structure allows trajectories favored by factions or subordinate commanders to emerge even where they diverge from aggregate organizational behavior. And divergence between the statistical anchor and the contextual LLM agents should be treated as an analytic signal identifying an assumption that warrants scrutiny, rather than as noise to be reconciled. These features reduce the risk without removing it. Validation should therefore include assumption-relaxation runs in which anchors are weakened or suspended, and historical cases of pattern departure tested under a documented temporal cutoff. Where an actor's departure originates in information absent from all available channels, AI-INT will not generate it; its contribution is then limited to making explicit where the assessment rests on continuity assumptions the evidence does not independently support.

Uncertainty Communication

Uncertainty communication is equally central. AI-INT outputs should not be written as deterministic forecasts. Where AI-INT outputs include likelihood judgments, these should be expressed through standardized estimative language (Kent, 1964; ODNI, 2015). The frequency with which an outcome appears across exploratory simulation runs should not, by itself, be treated as evidence of its real-world probability (Davis & Bracken, 2025). They should also identify assumptions, evidentiary gaps, source provenance, analytic confidence, alternative scenarios, and

indicators that would change the assessment. This is consistent with analytic-tradecraft standards that emphasize objectivity, independence, sourcing, uncertainty, and alternatives.

Data Security and Source Protection

AI-INT systems that integrate sensitive or classified inputs require security architectures appropriate to the information they process. In operational environments, such systems would need to run within secure infrastructure, maintain strict controls over inputs and outputs, and protect intelligence sources and methods. Security controls must also protect input integrity. Adversaries may poison training or retrieval data or embed indirect prompt injections in material likely to enter the retrieval pipeline (Autio et al., 2024).

The present article is based on non-classified conceptual analysis and does not disclose restricted operational information.

Human Oversight

AI-INT should augment intelligence professionals, not replace them. Its role is to generate hypotheses, expose assumptions, simulate alternative trajectories, and support structured analytic reasoning. Final judgments should remain the responsibility of human analysts and decision-makers, particularly when assessments may affect the use of force, diplomacy, detention, sanctions, or other consequential state actions. This is consistent with broader work on responsible national-security AI, human-machine teaming, risk management, and the institutional vulnerabilities of intelligence systems (National Security Commission on Artificial Intelligence, 2021; NIST, 2023).

Section 9: Conclusion - AI-INT as a Proposed Synthesis-and-Simulation Layer

The novelty of AI-INT does not rest on claiming that its individual functions are entirely new. Multi-source fusion, actor modeling, structured scenario analysis, and iterative analyst interrogation all have antecedents in established intelligence analysis, wargaming, and computational modeling. Its distinctive contribution lies in integrating them into a single, persistent, actor-specific architecture in which each component constrains the others and the whole remains open to analyst interrogation and revision. In doing so, AI-INT extends established intelligence practice by adding a structured synthesis-and-simulation layer intended to expose overlooked trajectories, hidden assumptions, and potential warning indicators before analytic blind spots contribute to strategic surprise.

The purpose is not to eliminate surprise or predict future decisions with certainty. It is to reduce analytic blind spots by making a wider range of plausible actor decisions and scenario developments analytically visible before events narrow the space for interpretation and response.

Future research should proceed along two complementary tracks. Comparative studies should evaluate AI-INT against human-led wargaming, focusing on their relative ability to generate novel and decision-relevant insights, expose analytic blind spots, and identify assumptions shaped by cognitive or organizational bias. A second, technological research agenda should examine how to develop and validate AI-supported actor models that reproduce relevant patterns of human and organizational behavior under comparable conditions, and how to run large numbers of systematically varied simulations efficiently, reliably, and reproducibly.

This article advances AI-INT as a possible response to the enduring problem of strategic surprise by naming this function, defining its scope, and outlining the initial conditions for its responsible development and evaluation.

References

- Aher, G. V., Arriaga, R. I., & Kalai, A. T. (2023). Using large language models to simulate multiple humans and replicate human subject studies. *Proceedings of the 40th International Conference on Machine Learning*, 202, 337-371.
<https://proceedings.mlr.press/v202/aher23a.html>
- Allison, G. T., & Zelikow, P. (1999). *Essence of decision: Explaining the Cuban Missile Crisis* (2nd ed.). Longman.
- Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). *Artificial intelligence risk management framework: Generative artificial intelligence profile* (NIST AI 600-1). National Institute of Standards and Technology.
<https://doi.org/10.6028/NIST.AI.600-1>
- Bellaby, R. (2024). The ethical problems of ‘intelligence–AI’. *International Affairs*, 100(6), 2525–2542. <https://doi.org/10.1093/ia/iaae227>
- Betts, R. K. (1978). Analysis, war, and decision: Why intelligence failures are inevitable. *World Politics*, 31(1), 61-89.
- Betts, R. K. (1982). *Surprise attack: Lessons for defense planning*. Brookings Institution.
- Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(Suppl. 3), 7280-7287.
<https://doi.org/10.1073/pnas.082080899>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Clark, R. M. (2004). *Intelligence analysis: A target-centric approach*. CQ Press.
- Cronin, A. K. (2010). *When should we talk to terrorists?* United States Institute of Peace Special Report 240. <https://www.usip.org/publications/2010/05/when-should-we-talk-terrorists>

- Dahl, E. J. (2013). *Intelligence and surprise attack: Failure and success from Pearl Harbor to 9/11 and beyond*. Georgetown University Press.
- Dasarathy, B. V. (1997). Sensor fusion potential exploitation: Innovative architectures and illustrative applications. *Proceedings of the IEEE*, 85(1), 24–38.
<https://doi.org/10.1109/5.554206>
- Davis, P. K., & Bracken, P. (2025). Artificial intelligence for wargaming and modeling. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 22(1), 25–40. <https://doi.org/10.1177/15485129211073126>
- Dylan, H., Gioe, D. V., & Grossfeld, E. (2023). The autocrat’s intelligence paradox: Vladimir Putin’s (mis)management of Russian strategic assessment in the Ukraine War. *The British Journal of Politics and International Relations*, 25(3), 385-404.
<https://doi.org/10.1177/13691481221146113>
- Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social Simulation*, 11(4), Article 12. <https://www.jasss.org/11/4/12.html>
- Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., Wiest, O., & Zhang, X. (2024). Large language model based multi-agents: A survey of progress and challenges. *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*.
<https://doi.org/10.24963/ijcai.2024/890>
- Gustafson, K., Lomas, D., Wagner, S., Shaaban Abdalla, N., & Davies, P. H. J. (2024). Intelligence warning in the Ukraine war, Autumn 2021-Summer 2022. *Intelligence and National Security*, 39(3), 400-419. <https://doi.org/10.1080/02684527.2024.2322214>
- Hall, D. L., & Llinas, J. (2008). Multisensor data fusion. In M. E. Liggins II, D. L. Hall, & J. Llinas (Eds.), *Handbook of multisensor data fusion: Theory and practice* (2nd ed., pp. 1–14). CRC Press. <https://doi.org/10.1201/9781420053098.ch1>
- Handel, M. I. (1984). Intelligence and the problem of strategic surprise. *The Journal of Strategic Studies*, 7(3), 229-281. <https://doi.org/10.1080/01402398408437190>
- Henrico, S., & Putter, D. (2025). Intelligence collection disciplines: A systematic review. *Journal of Applied Security Research*, 20(1), 46-70.
<https://doi.org/10.1080/19361610.2023.2296765>

- Heuer, R. J., Jr. (1999). *Psychology of intelligence analysis*. Center for the Study of Intelligence, Central Intelligence Agency.
- Horton, J. J., Filippas, A., & Manning, B. S. (2023). *Large language models as simulated economic agents: What can we learn from Homo silicus?* NBER Working Paper No. 31122. National Bureau of Economic Research. <https://doi.org/10.3386/w31122>
- Hua, W., Fan, L., Li, L., Mei, K., Ji, J., Ge, Y., Hemphill, L., & Zhang, Y. (2023). War and peace (WarAgent): LLM-based multi-agent simulation of world wars [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2311.17227>
- Ikani, N. (2025). Beyond the binary: A new typology for evaluating warning success and failure in strategic surprise. *International Studies Review*, 27(1), viaf009. <https://doi.org/10.1093/isr/viaf009>
- Jervis, R. (2010). *Why intelligence fails: Lessons from the Iranian Revolution and the Iraq War*. Cornell University Press.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), Article 248. <https://doi.org/10.1145/3571730>
- Johnson, J. (2026). GenAI and synthetic foresight at the brink: The future of nuclear crisis decision-making. *European Journal of International Security*, 1–22. <https://doi.org/10.1017/eis.2026.10063>
- Katz, B. (2020). *The collection edge: Harnessing emerging technologies for intelligence collection*. Center for Strategic and International Studies. <https://www.csis.org/analysis/collection-edge-harnessing-emerging-technologies-intelligence-collection>
- Kendrick. (2023, November 27). *AI is the sixth intelligence discipline*. Towards AI. <https://towardsai.net/p/artificial-intelligence/ai-is-the-sixth-intelligence-discipline>
- Kent, S. (1964). Words of estimative probability. *Studies in Intelligence*, 8(4), 49-65.

- Kydd, A. H., & Walter, B. F. (2002). Sabotaging the peace: The politics of extremist violence. *International Organization*, 56(2), 263-296.
<https://doi.org/10.1162/002081802320005487>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.
- Lieberman, S. (1956). The effects of changes in roles on the attitudes of role occupants. *Human Relations*, 9(4), 385-402.
- Lin-Greenberg, E., Pauly, R. B. C., & Schneider, J. G. (2022). Wargaming for International Relations research. *European Journal of International Relations*, 28(1), 83–109.
<https://doi.org/10.1177/13540661211064090>
- Lowenthal, M. M. (2025). *Intelligence: From secrets to policy* (10th ed.). CQ Press.
- Macal, C. M., & North, M. J. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4(3), 151-162. <https://doi.org/10.1057/jos.2010.3>
- Mampilly, Z. C. (2011). *Rebel rulers: Insurgent governance and civilian life during war*. Cornell University Press.
- Moran, C. R., Burton, J., & Christou, G. (2023). The US intelligence community, global security, and AI: From secret intelligence to smart spying. *Journal of Global Security Studies*, 8(2), ogad005. <https://doi.org/10.1093/jogss/ogad005>
- National Commission on Terrorist Attacks Upon the United States. (2004). *The 9/11 Commission report: Final report of the National Commission on Terrorist Attacks Upon the United States*. U.S. Government Printing Office. <https://www.govinfo.gov/content/pkg/GPO-911REPORT/pdf/GPO-911REPORT.pdf>
- National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). <https://doi.org/10.6028/NIST.AI.100-1>
- National Security Commission on Artificial Intelligence. (2021). *Final report*.
<https://digital.library.unt.edu/ark:/67531/metadc1851188/>

- Neuman, Y. (2024). *AI for understanding context*. Springer. <https://doi.org/10.1007/978-3-031-64210-4>
- Office of the Director of National Intelligence. (2015). *Intelligence Community Directive 203: Analytic standards* (as amended June 12, 2023).
<https://www.dni.gov/files/documents/ICD/ICD-203.pdf>
- Office of the Director of National Intelligence. (2019). *The AIM initiative: A strategy for augmenting intelligence using machines*.
<https://www.dni.gov/files/ODNI/documents/AIM-Strategy.pdf>
- Office of the Director of National Intelligence. (n.d.). *What is intelligence?*
<https://www.dni.gov/index.php/what-we-do/what-is-intelligence>
- Office of the Director of National Intelligence & Central Intelligence Agency. (2024). *The IC OSINT strategy 2024-2026*.
https://www.dni.gov/files/ODNI/documents/IC_OSINT_Strategy.pdf
- Özdamar, Ö., Haliştoprak, B. T., & Young, M. (2023). Do campaign speeches predict foreign policy? An operational code and leadership trait analysis of Donald Trump's MENA policies. *Uluslararası İlişkiler*, 20(80), 73–91. <https://doi.org/10.33458/uidergisi.1300777>
- Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*.
<https://doi.org/10.1145/3586183.3606763>
- Perla, P. P., & McGrady, E. D. (2011). Why wargaming works. *Naval War College Review*, 64(3), 111–130. <https://digital-commons.usnwc.edu/nwc-review/vol64/iss3/8/>
- Schwarz, J. O. (2009). Business wargaming: Developing foresight within a strategic simulation. *Technology Analysis & Strategic Management*, 21(3), 291–305.
<https://doi.org/10.1080/09537320902750590>
- Shapiro, J. N. (2013). *The terrorist's dilemma: Managing violent covert organizations*. Princeton University Press.

- Solopova, V., Skorik, V., Tereshchenko, M., Haidun, A., & Vykhopen, O. (2026). *LLMs as strategic actors: Behavioral alignment, risk calibration, and argumentation framing in geopolitical simulations* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2603.02128>
- Staniland, P. (2014). *Networks of rebellion: Explaining insurgent cohesion and collapse*. Cornell University Press.
- Taylor, G., Frederiksen, R., Vane, R. R., III, & Waltz, E. (2004). Agent-based simulation of geopolitical conflict. *Proceedings of the Sixteenth Innovative Applications of Artificial Intelligence Conference*, 884-891. AAAI Press.
- Tzukert, N., & Sulitzeanu-Kenan, R. (2026). *Fight and bargain: Negotiation patterns of semi-state terrorist organizations*. Digital Access to Scholarship at Harvard.
- Tzukert, N., Sulitzeanu-Kenan, R., & Berrebi, C. (2025). *Dangerous by design: Distinct patterns of violence among semi-state terrorist organizations*. Digital Access to Scholarship at Harvard.
- Van Puyvelde, D., & Tabárez Rienzi, F. T. (2025). The rise of open-source intelligence. *European Journal of International Security*, 10(4), 530-544. <https://doi.org/10.1017/eis.2024.61>
- Walker, S. G., Schafer, M., & Young, M. D. (1999). Presidential operational codes and foreign policy conflicts in the post-Cold War world. *Journal of Conflict Resolution*, 43(5), 610–625. <https://doi.org/10.1177/0022002799043005004>
- Waltz, E. (2003). *Knowledge management in the intelligence enterprise*. Artech House.
- Wirtz, J. J. (2024). Are intelligence failures still inevitable? *International Journal of Intelligence and CounterIntelligence*, 37(1), 307-330. <https://doi.org/10.1080/08850607.2023.2214328>
- Wyss, M. (2024). The October 7 attack: An assessment of the intelligence failings. *CTC Sentinel*, 17(9), 1–9. <https://ctc.westpoint.edu/the-october-7-attack-an-assessment-of-the-intelligence-failings/>

Yu, C., Ye, J., Li, Y., Li, Z., Ferrara, E., Hu, X., & Zhao, Y. (2024). *A large-scale simulation on large language models for decision-making in political science* [Preprint]. arXiv.
<https://doi.org/10.48550/arXiv.2412.15291>