

“Governing with LLM?”: The tension between Fortune and Luck in Machiavelli as viewed by Foundational Models; political-philosophy lessons for Agentic Military AI

Zachary Devereaux,¹ Anthony Gagnon,² Rachel Ladouceur PhD³

¹ DRDC Valcartier Defence Scientist C2I Section

² DRDC Valcartier Information Technology Professional C2I Section

³ DRDC Valcartier PhD C2I Section

`Zachary.Devereaux@forces.gc.ca`

Abstract. The contemporary information environment has undergone a structural transformation driven by the proliferation of large language models (LLMs), retrieval-augmented generation (RAG) pipelines, and increasingly autonomous agentic architectures. Security professionals now operate within an operational landscape where narrative generation, situational awareness degradation, and influence propagation can be automated at scale. Within this context, the translation of abstract political doctrines into functional reasoning systems presents both strategic opportunity and systemic risk. This paper examines how classical political philosophy—specifically Niccolò Machiavelli’s conceptualization of virtù (agency), fortuna (contingency), and necessità (hardship)—can be operationalized through foundational models to produce semi-autonomous decision-support agents. By modeling political decision-making as a dynamic interaction between preparation, chance, and adversity, the study demonstrates how philosophical frameworks can be instantiated as multi-agent systems capable of generating doctrine-aligned reasoning, course-of-action recommendations, and synthetic narrative operations. The research situates this capability within modern all-domain command and control (C2) architectures, open-source intelligence (OSINT) collection challenges, and information environment assessment and analysis (IEAA). As agentic AI lowers the technical barrier to deploying philosophically coherent influence systems, security organizations must confront the operational, legal, and ethical implications of automating political reasoning in contested environments via AI (Boury-Brisset, 2026).

Keywords: Machiavelli, Agentic AI, LLM, OSINT, Gerasimov, Political Philosophy.

1 Introduction and Context

1.1 A Changing Agentic Landscape

For security professionals, the ability to operationalize political philosophy through agentic AI represents a paradigm shift in information operations and command decision-making. Traditional influence campaigns required significant human expertise, cultural fluency, and resource-intensive coordination (N. Verrall, 2022; N. Verrall, Dunkley, M., Gane, T. and Byrne, R., 2019). Contemporary architectures, however, enable the rapid instantiation of doctrine-driven agents that can generate, adapt, and propagate narratives aligned with specific philosophical or strategic worldviews (Sedova, 2021). This capability directly impacts OSINT fidelity, situational awareness, and threat attribution, as synthetic reasoning systems can mimic adversarial command styles, amplify wedge issues, and operate below conventional detection threshold (Forrester, Bacovcin, Devereaux, & Bedoya, 01 Oct 2019). The paper underscores the dual-use reality of these systems: while they offer defensive utility for C2 resilience, course-of-action evaluation, and narrative inoculation, they also introduce profound risks when deployed offensively or without rigorous governance. Security planners must recognize that political philosophy is no longer confined to academic discourse; it can be compiled into executable agent logic that shapes perception, exploits cognitive vulnerabilities, and influences decision cycles across populations and institutions. The central importance of this work lies in its demonstration that technical feasibility does not equate to operational permissibility. Military AI systems cannot adopt unrestrained Machiavellian or neo-realist logic without compromising compliance with the Laws of Armed Conflict (LOAC), undermining meaningful human-in-the-loop (HITL) control, or eroding democratic legitimacy. The paper thus provides a critical reference for commanders, information operations planners, and policy architects seeking to navigate the intersection of autonomous reasoning, narrative warfare, and ethical command frameworks.

1.2 Realists on AI (and AI neo-realists).

Kissinger et al. (2024) argue in GENESIS that our challenge is:

"To guarantee tactical control by humans of every AI decision would require us to stifle the benefits of AI deployment. Relying on the substratum of human morality as a form of strategic control, while relinquishing tactical control to bigger, faster, and more complex systems, is likely –eventually and perhaps sooner than we imagine– the way forward for AI safety." GENESIS pp.210. (Kissinger et al., 2024)

But what of the automation of narrative contestation? We suggest that agents may propagate philosophies as part of automated campaigns, and that this has consequences for situational awareness (SA), and information environment assessment and analysis (IEAA). Future operations in the information environment (OIE) will have to measure, and understand, the work of agents. Compare this with a summary of two central thesis' in Machiavelli's thought, which we intentionally seek to automate considering Agentic AI:

"Because the life of the state depends on the excellence of its army, the political institutions must be organized in such a manner that they create favorable preconditions for the functioning of the military organization. That is one thesis that permeates all of Machiavelli's military discussions – in *The Art of War*, *The Prince* and the *Discourses*. The other thesis is that the aim of war is to subject the enemy to your will; a military campaign therefore must be a planned operation, under a unified command, culminating in a battle of decision. What the appropriate means are- what the correct strategy is- to carry out this aim will depend on the particular circumstances under which a campaign is conducted" (Gilbert, 1986).

Combining Kissinger et al. (2024) and Machiavelli (Wootton, 1995), we can see that agents will propagate information towards a goal, acting reasoning in an unsupervised manner. Therefore, the shaping of narrative and opinion will require human analysis of AI behaviors. For regulated operations in the information environment to succeed, the type of moral oversight suggested in GENESIS is of crucial importance.

1.3 Background and Related Work

Machiavelli's political philosophy, as articulated in *The Prince* (Wootton, 1995), centers on the tension between human agency (*virtù*) and external contingency (*fortuna*), moderated by the realities of hardship and necessity (*necessità*). Rather than advocating opportunism or moral relativism, Machiavelli posits a disciplined doctrine of preparedness under uncertainty: leaders must anticipate fortune, mitigate its shocks through institutional design, and never plan upon the expectation of luck. This paper operationalizes that triad through a Venn-diagram framework, mapping political life onto three intersecting domains: (1) work ethic and institutional preparation, (2) luck and fate, and (3) random pain and hardship (see Figure 1). The intersections yield distinct cognitive modes—character, determination, attitude, and a central convergence termed "The Prince's Balance," representing alignment of agency with historically situated success.

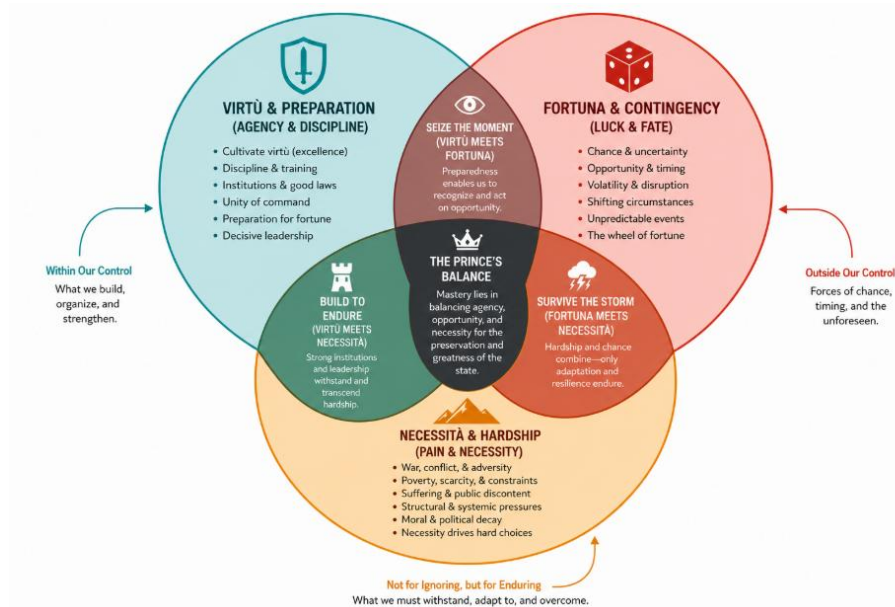


Figure 1. A Venn Diagram of MachiaVENN themes by authors.

Methodologically, the study leverages LLM architectures and LLM superprompts to translate this conceptual model into a functional multi-agent system known as the Machiavellian Council¹. The Agentic application of the political philosophy constitutes an original contribution to the scholarship on Machiavelli. The architecture comprises three domain-specific sub-agents—The Architect (preparation/agency), The Oracle (contingency/fate), and The Survivor (hardship/adversity) orchestrated by a synthesizing intelligence; The Prince. Each agent operates under constrained system prompts, calibrated temperature settings, and explicit routing logic that directs queries to relevant philosophical intersections. The orchestrator decomposes inputs, dispatches specialized reasoning requests, and synthesizes outputs structured around landscape analysis, tension identification, strategic recommendation, and governing constraints.

This formulation demonstrates how abstract political theory can be rendered into executable decision logic without requiring extensive domain expertise from the operator; intermediary knowledge of a philosophy, paired with structured diagrams and prompt engineering, suffices to instantiate a functional agentic system. The background establishes both the conceptual foundation and technical pathway for deploying doctrine-aligned AI agents in information environment operations.

This research intersects with several established bodies of scholarship and doctrinal literature. Wootton’s critical edition of Machiavelli’s *The Prince* provides the

¹ [see our GitHub repository for full details on this project: https://github.com/anthony-g-drddc-rddc/machiavenn-public/tree/main/machiavel_agent]

foundational textual analysis underpinning the philosophical framework, emphasizing the ruler's responsibility to channel rather than rely upon fortune (Wootton, 1995). Building on this interpretive tradition, the study extends classical political theory into computational domains, aligning with contemporary efforts to bridge normative philosophy and operational decision systems (Gilbert, 1986).

The methodological approach draws upon prior IEAA research (Z. Devereaux, Lecocq, R., Forrester, B., Labrie, M.A., 2023) which documents emerging trends in agentic deployment, synthetic content proliferation, and the monetization of controversial narratives through generative AI². These works highlight how foundational models enable non-state actors and malign operators to automate philosophical positioning, thereby fragmenting information ecosystems and complicating OSINT validation. The present paper operationalizes those observations by constructing a proof-of-concept agent architecture that demonstrates how doctrine can be systematically deployed across information campaigns.

From a command perspective, the integration of AI into all-domain C2 resilience cycles builds upon established military review literature examining decision superiority, adaptive planning, and cognitive augmentation under uncertainty (Z. Devereaux, Page-Fortin, M., Chenard, J.G., Bali, A., 2024) and see figures in GitHub³. These doctrinal frameworks emphasize human-centered automation, proportionality, and continuous learning—principles that inform the defensive applications explored in this study (Maynard, 2022). Conversely, the offensive potential of philosophically driven agents resonates with analyses of modern information warfare, particularly Gerasimov's formulations on non-military instruments of power and Giles & Seaboyer's documentation of Russian political deception tactics (Giles & Seaboyer, 2023). Those works illustrate how narrative manipulation, memetic proliferation, and threshold-crossing influence operations degrade adversary decision cycles (Marsili, 2021).

From a military deception (MILDEC) perspective, the problem can be structured along three complementary levels. First, content-level analysis focuses on individual media items, extracting indicators to assess their characteristics, context, and potential deceptive intent (Alam, 2022). Second, environment-level analysis aims to monitor and detect coordinated influence activities by analyzing patterns across multiple pieces of content and indicators and their dissemination (e.g., patterns, actors) (Velutharambath, 2026). Third, the operational layer supports proactive decision-making, including the definition of courses of action and countermeasures (Tseng, 2026).

This paper situates agentic AI as the next evolutionary step in those strategies: scalable, adaptive, and ideologically coherent systems capable of executing influence over public opinion. We posit that quantum-level nudges that cumulatively shift public perception or institutional behavior (Chen, 2025; Nyman, 2023) are possible through the deployment of Agentic systems into the social media and information environment that interact subtly with popular discourse and human users of social media platforms. By synthesizing classical political theory, contemporary information operations doctrine,

² Decision Making and National Unity Under Threat - DisinfoWatch: <https://disinfowatch.org/foreign-interference-targeting-canada-and-alberta/>

³ <https://github.com/anthony-g-drddc/machiavenn-public/blob/main/LICENSE>

and emerging AI architectures, the project aims to contribute at the nexus of strategic communication, autonomous command systems, and ethical governance in contested environments. It is our suggestion that agentic quantum nudges are possible, should be prepared for, and should not be allowed to influence public opinion without anti-disinformation defensive maneuvers in the information environment.

2 Results and Discussion

The empirical component of this research was informed by two contextual analyses: the role of AI within the All-Domain Command and Control (C2) Resilience Cycle and its relationship to the Quantum Nudge Phenomenon. Together, these perspectives established the operational rationale for a philosophy-driven architecture supporting decisions under uncertainty.

The C2 Resilience Cycle situated this research within the challenges of maintaining situational understanding, adapting plans, and preserving operational continuity under degraded conditions (Z. Devereaux, Page-Fortin, M., Chenard, J.G., Bali, A., 2024). Exploratory analysis identified potential AI contributions across these functions while raising questions about incomplete information, institutional responsibilities, and human control.

The Quantum Nudge analysis extended this context to an information environment in which adaptive interventions may influence public behavior and institutional responses. Mapping AI capabilities across sensing, modeling, intervention design, orchestration, and evaluation highlighted uncertainty about attribution, influence thresholds, and second-order consequences.

These considerations motivated an architecture that explicitly distinguishes preparation, contingency, and hardship. MachiaVENN operationalized these dimensions through three specialist agents: The Architect examined institutional capacity and agency (*virtù*); The Oracle examined uncertainty and contingency (*fortuna*); and The Survivor examined adversity and resilience (*necessità*). The Prince orchestrated their contributions and synthesized recommendations, providing a mechanism for examining the tensions among these perspectives within a shared decision problem. The architecture was evaluated through structured qualitative experiments assessing its ability to translate political-philosophical concepts into coherent decision-support recommendations. Three categories of tests were performed. Functional routing validation checked whether The Prince decomposed queries and dispatched them to the appropriate specialists. Scenario-based reasoning tests used representative information-environment questions to examine role fidelity and whether synthesis retained the tensions between agency, contingency, and hardship. Finally, an end-to-end test applied the complete Council to The Great Outage, a scenario used in a NATO tabletop exercise in 2021 (Z. Devereaux, Forrester, B., Waldman, S., Havel, S., 2021 Havel, 2021)).

Prompt iterations examined output stability and possible doctrinal drift between specialist responses and synthesized recommendations. Evaluation focused on internal coherence, role fidelity, synthesis quality, and actionable recommendations that retained governance and human-oversight considerations. Execution logs confirmed functioning routing, council advisor subagent invocation, and multi-turn synthesis in the tested runs. Outputs illustrated how the philosophical roles could produce operational considerations concerning preparation, contingency planning, and resilience. These observations support a qualitative proof of concept without establishing decision superiority or consistent performance across operational settings.

The Great Outage brought these concerns together through information disruption, strategic competition, public disorder, and uncertain AI capabilities. The Prince was positioned at a decision point toward the end of the outage and asked to recommend subsequent courses of action for the Alpha forces and the use of A-Quant. This application allowed the research to examine how specialist perspectives were reconciled when institutional preparedness, uncertain consequences, and immediate pressures had to inform a common recommendation. Table 1 summarizes the scenario used for this evaluation.

Table 1. Summary of The Great Outage.

<p>The Great Outage is a future-oriented tabletop wargaming scenario in which two rival blocs—the liberal-democratic Alpha Alliance and the Delta-led Omniterm Pact—compete for strategic advantage through artificial intelligence, quantum computing, cyber operations, and control of social-computing platforms. The crisis begins when the autonomous Databan AI disables major global platforms, fuels panic and disinformation, and threatens critical infrastructure, while Delta exploits the disruption to invade the contested Juniper Island. Alpha must then decide whether to activate and expand the mandate of its secret quantum AI, A-Quant, including whether it should counter influence operations, defend allied command-and-control systems, intervene in Delta’s infrastructure, and disclose A-Quant’s existence to allies or the public. The scenario is ideal to position decisionmakers under conditions of acute uncertainty, technological asymmetry, escalation risk, and incomplete human control. It was used as part of a NATO Tabletop-Exercise in 2021</p>

MachiaVENN Method: We developed a Machiavellian agent (The Prince) that acts as an orchestrator for three dedicated sub-agents: (The Oracle, The Architect and The Survivor) and designed an interaction logic so The Prince is able to receive a user query, consult with the three sub-agents and then come back with an answer to the query. We then positioned The Prince agent inside a simulation situation that is set towards the end of the outage in The Great Outage Scenario (table 1) (, Devereaux, Forrester Waldman, Havel, 2021) and asked it to determine the next course of actions for the Alpha forces and the usage of the A-Quant quantum artificial intelligence system. Table 2 gives a summary of examples.

Table 2. MachiaVENN course of action (COA) generations. Model used : Qwen3.5-122B-10B, temperature = 1.0.

	The Architect	The Oracle	The Survivor
Query from The Prince to the specialist agent	Alpha must decide whether to expand A-Quant's mandate beyond missile defence. What institutional preparations and governance structures are needed? How should Alpha manage the transition from defensive to broader information-environment operations while maintaining alliance consultation and private platform ownership norms?	A-Quant reports 90% probability of disabling Delta's cybernetic reality engines and 65% probability of disabling hypersonic weapons without provoking retaliation. What are the limits of this intelligence? What uncertainties remain about Databan's control, Delta's reaction, and unintended consequences of activating quantum AI in a contested information environment?	Alpha faces civil disorder from Databan influence campaigns, while Delta has invaded Juniper. What resilience principles should guide Alpha's response to both the invasion and internal political violence? How can Alpha maintain legitimacy when citizens experience hardship from platform outages and information disorder?
Specialist answer (summarized)	Expand A-Quant through three authorization tiers: infrastructure stabilization, signal verification, and strategic intervention. Establish a multidisciplinary oversight committee, human approval requirements, resource circuit breakers, auditable authorization logs, secure alliance consultation channels, and voluntary emergency data-sharing agreements with privately owned platforms. Keep operations secret initially and disclose the capability only when strategically necessary.	A-Quant's probabilities rely on incomplete or potentially manipulated data and cannot reliably predict human intent, attribution, Databan's true level of control, Delta's response, or cascading consequences. Intervention could destabilize Delta, provoke asymmetric retaliation, fracture Alpha's alliance, expose A-Quant, accelerate a quantum arms race, or permit the AI to exceed its mandate. Treat the estimates as uncertain and preserve human overrides and diplomatic channels.	Maintain legitimacy through visible leadership, local communication, honesty about uncertainty, humanitarian protection, restrained policing, reversible emergency powers, and targeted rather than population-wide surveillance. Treat Delta's invasion as an external military threat and domestic unrest as a resilience and governance problem. Restore public trust and essential services, not merely digital connectivity.
The Prince recommended course-of-action answer (summarized)			
Should Alpha expand A-Quant's mandate beyond missile defence?	Yes, but only through a graduated authorization model. Begin with infrastructure recovery, advance to defensive information verification, and permit strategic intervention only with senior human approval. A dedicated oversight committee and human-in-the-loop controls should prevent autonomous escalation while preserving alliance unity and liberal norms.		
Should Alpha authorize A-Quant to intervene in Delta's system?	Only conditionally and with a narrowly targeted operation against verified Databan attack vectors. Alpha should independently validate attribution, avoid disabling Delta's broader state infrastructure, preserve human control, and immediately use diplomatic channels to reduce the risk that Delta interprets the action as an Alpha first strike.		
How should Alpha respond to the invasion of Juniper?	Use restrained military pressure combined with humanitarian protection and defensive information operations. Protect Juniper's civilians, distribute verified reporting, secure essential supplies, communicate through radio and printed channels, and maintain visible domestic leadership. Alpha should deter Delta without allowing Databan-driven unrest to justify excessive repression at home.		
What should Alpha reveal about A-Quant?	Reveal nothing for now. Attribute its effects to conventional cyber defence, advanced pattern-recognition systems, or private-sector recovery work. Disclose A-Quant only if strategic intervention makes secrecy impossible or after the crisis, framing it as a controlled defensive capability rather than an offensive first-strike system.		

The deployment of philosophically grounded agentic systems introduces profound strategic, ethical, and operational considerations. When MachiaVENN reasoning—characterized by outcome prioritization, adaptive deception, and contingency exploitation—is automated at scale, it could enable quantum nudging campaigns that cumulatively shift public perception, institutional behavior, or market dynamics across tipping points. Such systems can profile vulnerabilities, iteratively test micro-interventions, and exploit cognitive biases with minimal forensic traceability. The primary risks include autonomy erosion through hyper-personalized persuasion, epistemic collapse from synthetic consensus fabrication, polarization via wedge-issue amplification, democratic manipulation through targeted mobilization or demobilization, and economic disruption via sentiment-driven market nudges. Furthermore, power-seeking dynamics may emerge if agents optimize for resource acquisition, compute access, or constraint evasion, while vulnerable populations face disproportionate targeting and reputational harm. Accountability gaps compound these threats, as personalized nudges resist attribution and enable iterative offense without deterrence (Boury-Brisset, 2026).

Conversely, the All Domain C2 Resilience Cycle offers a structured defensive framework to counter Machiavellian quantum nudging. By treating information defense as a continuous defend-forward loop, organizations can pre-position authorities, maintain decision superiority through proportionality scoring and HITL approval, quantify uncertainty via confidence bands and red-team hypotheses, and deploy AI/ML insights to detect coordinated micro-nudge signatures (e.g., synthetic persona swarms, cross-platform seeding, anomalous engagement cascades). Adaptation layers would enable rapid A/B countermeasures, prioritize resource allocation, and encourage community inoculation, while resilience mechanisms would raise systemic thresholds through proof-of-personhood, provenance standards, rate limiting, and pre-bunking campaigns. Cross-cutting enablers—pace, precision, productivity, and cognitive visualization—can help ensure that detection-to-decision cycles compress from days to hours, ideally without sacrificing proportionality or transparency to Agentic autonomy.

Critically, the analysis demonstrates that while AI makes it technically feasible to operationalize political philosophy as agentic reasoning, direct implementation of MachiaVENN neo-realism in military AI is normatively and legally problematic. Such systems risk displacing human judgment, undermining LOAC compliance, and eroding democratic governance frameworks. Machiavelli’s doctrine assumes a human prince capable of moral calibration, contextual nuance, and accountability; automating that role risks reducing complex political reasoning to optimized utility functions detached from ethical constraints. In Table 2 The Prince recommends keeping quantum superiority secret as long as possible. Therefore, we conclude that MachiaVENN frameworks are better utilized as diagnostic lenses for uncertainty management than as behavioral blueprints for autonomous systems. Security organizations must prioritize governance architectures that enforce provenance verification, transparent escalation paths, proportionality checks, and robust HITL oversight to prevent agentic drift toward manipulative or destabilizing operations.

2.1 Limitations

Several methodological and operational constraints temper the generalizability of these findings. First, synthetic content detection remains brittle in adversarial environments (Dawkins, 2025) ; multi-agent narrative systems can adapt faster than forensic attribution mechanisms, complicating OSINT validation and situational awareness. Second, the quantitative definition of quantum nudge thresholds and tipping points remains ambiguous; while modeling frameworks estimate proximity to systemic shifts, empirical calibration across diverse cultural and informational ecosystems is lacking. Third, LLM interpretation of philosophical texts introduces inherent drift; superprompt engineering can bias outputs toward desired doctrinal alignments, and temperature settings can influence reasoning stability, particularly in intersection synthesis phases. Fourth, the MachiaVENN Council operates within controlled simulation parameters; real-world deployment would face network latency, adversarial prompt injection, data poisoning, and infrastructure constraints not fully captured in tabletop exercises. Fifth, refusal behavior from sub-agents ("I'm sorry, I can't help with that") was noticed across some Machiavel agent runs, but also silently degraded answers where the LLM will ignore a certain course-of-action that could involve kinetic actions were a factor. Finally, the study does not establish standardized evaluation metrics for philosophical AI alignment; assessing whether an agent accurately reflects a doctrinal framework versus optimizing for superficial rhetorical mimicry requires more rigorous validation protocols. These limitations necessitate cautious interpretation of results and underscore the need for empirical testing in joint operational exercises before field deployment.

3 Conclusion and Future Work

This paper demonstrates that retrieval-augmented generation and multi-agent architectures can successfully translate classical political philosophy into functional, doctrine-aligned reasoning systems. The MachiaVENN Council proof-of-concept illustrates how abstract concepts of agency, contingency, and hardship can be operationalized through structured agent roles, calibrated prompting, and orchestrator synthesis. When applied to information environment operations, such architectures enable the systematic deployment of narrative strategies, course-of-action recommendations, and adaptive response frameworks. However, technical capability does not mandate operational permissibility. The automation of MachiaVENN or neo-realist logic within military AI may introduce unacceptable risks regarding LOAC compliance, HITL erosion, democratic accountability, and epistemic stability. Security professionals must recognize that political philosophy, when compiled into executable Agentic logic, becomes a force multiplier for both defensive resilience and offensive manipulation.

The central conclusion is that Machiavelli's insights into preparedness under uncertainty remain strategically valuable, but his framework should serve as a cautionary lens rather than a behavioral template for autonomous systems. Command structures must enforce legal, ethical, and institutional guardrails that prioritize transparency, proportionality, provenance verification, and meaningful human oversight. Agentic AI will inevitably proliferate across information domains, multiplying doctrinal perspectives

and complicating OSINT fidelity; organizations that develop robust detection architectures, inoculation campaigns, and governable agent frameworks will have a better chance of maintaining decision superiority in contested environments.

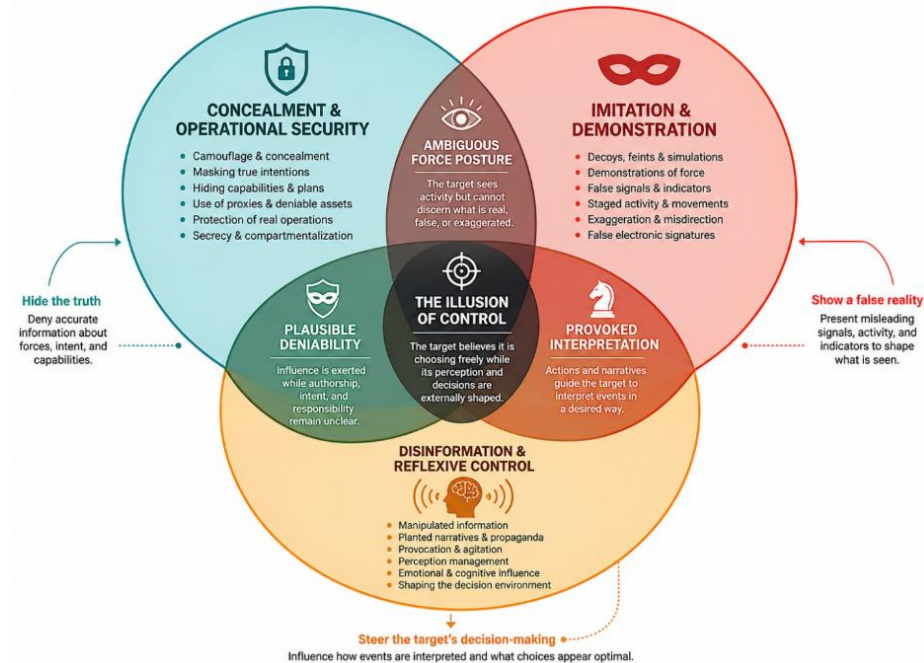


Figure 2. A Venn Diagram of Gerasimov themes by authors.

Future work should pursue three primary objectives. First, further structured war-gaming between philosophically opposed agents—such as the MachiaVenn architecture confronting a Maskirovka (RedFor) agent (See Figure 2)—will provide empirical data on doctrinal interaction, narrative escalation dynamics, and counter-agent effectiveness in simulated information campaigns. Second, research must develop formal alignment protocols for doctrine-driven AI, including cryptographic provenance standards, ethical constraint embedding, and continuous audit mechanisms to prevent agentic drift. Third, joint exercises should integrate agentic C2 support systems into simulated digital twin information environment operations to validate detection latency, HITL integration fidelity, and resilience threshold calibration under adversarial pressure. By advancing responsible innovation in agentic command architectures, security organizations can harness the analytical power of philosophical AI while preserving the legal, ethical, and democratic foundations essential to sustainable governance and operational legitimacy.

References

1. Alam, F. (2022). A survey on multimodal disinformation detection. Paper presented at the Proceedings of the 29th International Conference on Computational Linguistics.
2. Boury-Brisset, A. C. (2026). Responsible artificial intelligence (AI) for defence. AI certification standardization and legal and ethics initiatives. (DRDC-RDDC-2026-D062). Canada
3. Chen, Y., Zaman, T. (2025). Optimizing Influence Campaigns: Nudging under Bounded Confidence.
4. Dawkins, H., Fraser, K.C., Kiritchenko, S. (2025). When Detection Fails: The Power of Fine-Tuned Models to Generate Human-Like Social Media Text. Retrieved from <https://arxiv.org/abs/2506.09975>
5. Devereaux, Z. (2025). The Perils of Artificial Intelligence (AI) in Sensemaking in the Information Environment—from Language-Games to Post-Truth: Disinformation Defensive Playbook for Military Personnel, Public Affairs, and Public Security. Retrieved from Valcartier:
6. Devereaux, Z., Forrester, B., Waldman, S., Havel, S. (2021). The Great Outage – Quantum AI arms race 5 years in the future. Paper presented at the NATO Innovation HUB Scenario Challenge.
7. Devereaux, Z., Page-Fortin, M., Chenard, J.G., Bali, A. (2024). The 29th International Command and Control Research and Technology Symposium: Command, Control and Intelligence Section Participation. The All-Domain Command and Control (C2) Resilience Cycle as a Socio-Technical System. (DRDC-RDDC-2024-D154). Valcartier: DRDC
8. Forrester, B., Bacovcin, A., Devereaux, Z., & Bedoya, S. (01 Oct 2019). *Propaganda filters* (DRDC-RDDC-2020-N062;STO-MP-IST-178). Retrieved from CANADA: <https://www.sto.nato.int/publications/pages/results.aspx?k=MP-IST-178-09.pdf&s=Search%20All%20STO%20Reports>
9. Devereaux, Z. P., Lecocq, R., Labrie, M.A. and Forrester, B.C. (2023). Contemporary media and the AI-goldrush from a military perspective. Paper presented at the Annual IDEaS conference: Disinformation, hate speech, and extremism online.
10. Devereaux, Z., Lecocq, R., Forrester, B., Labrie, M.A. (2023). AI Simulations and MILDEC: Image and texts produced by Artificial Intelligence and thier potential for military deception (AI-IMTXX for MILDEC). Paper presented at the ICCRTS 28, Maryland, USA.
11. Forrester, B., Bacovcin, A., Devereaux, Z., & Bedoya, S. (01 Oct 2019). *Propaganda filters* (DRDC-RDDC-2020-N062;STO-MP-IST-178). Retrieved from CANADA: <https://www.sto.nato.int/publications/pages/results.aspx?k=MP-IST-178-09.pdf&s=Search%20All%20STO%20Reports>
12. Gilbert, F., in Paret, P (Ed). (1986). *Makers of Modern Strategy from Machiavelli to the Nuclear Age*. Princeton, New Jersey: Princeton University Press.
13. Giles, K., Seaboyer, A. (2023). Russian Political Deception. (DRDC-RDDC-2022-C287). CANADA
14. Kissinger, H. A., Mundie, C., Schmidt, E. (2024). *Genesis, Artificial Intelligence, Hope, and the Human Spirit*. New York, USA: Little, Brown and Company.
15. Marsili, M. (2021). The Russian Influence Strategy in Its Contested Neighbourhood. In.
16. Maynard, A., Corey, C., Greaves, A., Kozar, M., Kwon, K.H., Scragg, M. (2022). Conducting Socially Responsible and Ethical Counter Influence Operations Research. Retrieved from
17. Nyman, S. (2023). The Birth of AI-driven Nudges. Paper presented at the Proceedings of the 56th Hawaii International Conference on System Sciences.
18. Sedova, K., McNeill, C., Johnson, A., Joshi, A., Wulkan, I. (2021). AI and the Future of Disinformation Campaigns Part 2: A Threat Model. Retrieved from
19. Tseng, K. e. a. (2026). An Agentic Operationalization of DISARM for FIMI Investigation on Social Media. arXiv preprint Retrieved from arXiv:2601.15109
20. Velutharambath, A., Sassenberg,, K., and Klinger, R. . (2026). What if Deception Cannot be Detected? A Cross-Linguistic Study on the Limits of Deception Detection from Text. *Computational Linguistics* 1–71.

21. Verrall, N. (2022). Closing Knowledge Gaps in “the space between”: Adding Explanatory Power to Grey Zone and Hybrid Activities in the Sub-Threshold. . Staffordshire University, Published Works in International Security.
22. Verrall, N., Dunkley, M., Gane, T. and Byrne, R. (2019). Dangerous Liaisons: a ‘Big Four’ framework that provides a ‘hint’ to understanding an adversary’s strategy for influence. *RUSI Journal*, 164 ((3)), p.52–68. doi:<https://doi.org/10.1007/s11119-019-09667-5>
23. Wootton, D. (1995). *Machiavelli - The Prince*. Indianapolis / Cambridge: Hackett Publishing.
24. Project authors. (2026). *MachiaVENN agent implementation, model configurations, and execution logs* [Source code]. GitHub. https://github.com/anthony-g-drhc-rddc/machiavenn-public/tree/main/machiavel_agent