**THESIS BRIEF — THEORY-FIRST RESEARCH**

Edition: 2025-11-07 | Peer-review pending (Theory-First)

Smart Technology Investments

# Cognitive Wars: the AI Industrialization of Influence

Oct 31–Nov 07, 2025 | Sources: 3 | Anchor Status: Anchor-Absent | Report Type: Theoretical Research | Horizon: Near-term | Confidence: 0.400 *

| SD | AC | MT | RR |
|---|---|---|---|
| 0.53 | 0.00 | 0.45 | 0.65 |

Alignment: 6.0 Theory Depth: 6.0 Clarity: 7.0

**Disclosure & Method Note:** This is a *theory-first* brief. Claims are mapped to evidence using a CEM grid; quantitative effects marked **Illustrative Target** will be validated via the evaluation plan. Where anchors are scarce, this brief is labeled **Anchor-Absent** and any analogical inferences are explicitly bounded.

## Abstract & Theory-First Framing.

**Outline**

# Preface: Theory-First Orientation

- Claim 1: A theory-first approach foregrounds causal mechanisms before empirical pattern-seeking; it makes explicit the pathways through which industrialization affects cognition in warfare.
- Claim 2: Explicit theoretical commitments (layered mechanisms, information-processing constraints, institutional path-dependence) clarify observable expectations for tempo, error modes, and organizational responses.

Rationale: By isolating mechanisms prior to case selection, the comparative design can more directly test counterfactuals about organizational choices under industrialized influence production.

---

# Introduction: Problem Statement and Research Questions

Problem statement: Contemporary conflict increasingly comprises attempts to shape beliefs, attention, and decision processes at scale. AI systems industrialize influence by automating content generation, personalization, distribution, and evaluation—creating persistent, scalable pressures on adversary and domestic cognitive ecologies.

## Research questions

1. How does AI-driven industrialization of influence reshape the cognitive terrain of warfare?
2. Through which mechanisms (tempo, volume, standardization) does industrialization affect perception, decision-making, and collective sensemaking?
3. What operational metrics, failure modes, and delegation policies should doctrine adopt to manage industrialized influence?

---

# Theoretical Framework: Cognitive Theory of War

## Core propositions

- Wars are cognitive as well as material contests: beliefs, expectations, attention, narrative dominance, and decision heuristics materially affect strategic outcomes.
- A layered model: (a) Macro: industrial capacity and infrastructure (communications, data centers, AI pipelines); (b) Meso: organizations that routinize and deploy influence; (c) Micro: individual cognitive processes (attention, bounded rationality, heuristics).
- Industrialization exerts causal pressure across layers: it changes the distribution of information, the affordances of automated delegation, and the incentives for institutional standardization.

Observable implications: shorter decision cycles, increased delegation to automated agents, higher throughput of influence signals, and larger-scale but shallower belief shifts.

---

# Literature Review: Wars, Industrialization, and Cognitive Change

Existing scholarship emphasizes materiel (logistics, firepower) when analyzing industrialization. Interdisciplinary work—history of industrial warfare, organizational studies, and cognitive science—suggests distinct cognitive effects: compression of time horizons, routinization of decisions, and information overload that shape attention and learning.

### Foundations: Why these anchors?

The brief privileges peer-reviewed, non-preprint anchors where possible to ground institutional and historical claims in vetted scholarship and to avoid over-reliance on early-stage technical claims. Where technical system properties are central (consensus, distributed detection, ML signatures), recent preprints provide current mechanistic detail; these are used with caution alongside peer-reviewed anchors to triangulate conclusions. For institutional standardization and protocol examples we adopt a peer-reviewed, non-preprint anchor to illustrate cross-domain lessons about routinization and protocolization in high-stakes environments [2]. Preprint technical literature (consensus protocols, distributed observers, ML detection) is cited for mechanistic specificity about networked decision architectures and detection limits [3][4][5][6][7][1].

---

# Historical Context: Industrialization and the Transformation of Warfare

Industrialization historically remade logistics (rail, telegraph), tempo (mass mobilization), and organizational learning (staff systems). These changes placed cognitive demands on commanders (signal filtering, delegation), civilian populations (propaganda and morale management), and institutions (standard operating procedures). Recurrent patterns include centralization of planning, proliferation of pre-planned routines, and emergent informational bottlenecks that alter how organizations adapt and learn.

---

# Mechanisms: How Industrialization Influences Cognitive Dimensions of War

This section details mechanistic pathways by which AI industrialization changes the cognitive ecology of conflict. The mechanisms below are presented distinctly from the Executive Summary and synthesize technical and organizational drivers.

## Mechanism 1 — Tempo and Decision Compression

- Industrial-scale automation shortens the horizon between action and effect (e.g., automated influence campaigns produce rapid feedback loops). Organizations confront compressed OODA (observe–orient–decide–act) cycles and shift toward more heuristic, delegated decision-making.

Mechanism 2 — Information Volume, Filtering, and Attention Economies

- AI enables mass personalization and continuous A/B testing of messages; the resulting signal volume exacerbates signal-to-noise problems, forcing institutions to invest in filters (ranking algorithms, trusted nodes) and to accept increased Type I/II errors. Distributed consensus and observer-network results illuminate how detectability and observability degrade in sparse, noisy networks, constraining reliable collective inference under adversarial injection [3][5].

Mechanism 3 — Standardization, Routinization, and Institutional Memory

- Industrial practices favor templating (influence playbooks, response SOPs). Standardization increases scale and repeatability but reduces local improvisation and raises fragility to novel attacks; empirical analogues from non-military high-stakes protocols show tradeoffs between reproducibility and adaptable judgement [2].

Mechanism 4 — Automation of Experimentation and Rapid Learning

- Continuous automated experimentation (fast iteration of variants, causal measurement at scale) converts influence into an engineering problem. This industrial learning loop accelerates adaptation but can entrench manipulative affordances and privilege actors with greater compute and data resources. ML-based detection and classification work highlights both opportunities and constraints in automated anomaly detection and adversarial robustness [1].

Mechanism 5 — Networked Delegation and Consensus Failure Modes

- When organizations distribute sensing and inference across networked agents, consensus algorithms and distributed optimization architectures determine how local signals aggregate. Limitations in detectability and cooperation under partial observability create blind spots and failure modes—partitioning, sybil-like injection, or targeted misinformation can prevent accurate global inference [4][6][7].

Together these mechanisms produce a cognitive ecology characterized by high-frequency influence, brittle standard responses, and contested information fabrics.

---

## Conceptualizing "Cognitive Wars" and Forms of Influence

Definition: Cognitive wars are conflicts where shaping cognition—beliefs, attention, expectations, and decision processes—constitutes a primary strategic objective.

### Forms of influence

- Persuasion: targeted narratives that alter beliefs or preferences.
- Distraction/denial: saturating attention channels to mask operations.
- Delegation manipulation: corrupting automated decision rules or training data to produce predictable behaviors.
- Legitimacy warfare: contesting social institutions' credibility to erode adversary will.

Industrialization intersects each form by scaling production, enabling personalization, and accelerating feedback-driven adaptation.

---

## Methodology: Comparative Historical and Process-Tracing Design

Design: Small-N comparative cases selected across stages of industrialization (early industrial, total-war industrial, late-industrial/early-information) with within-case process tracing to link industrial inputs (infrastructure, protocols, information systems) to cognitive outcomes (tempo compression, delegation levels, belief dynamics).

Causal inference: Mechanism-level process traces evaluate whether the presence and operation of specific industrialized practices plausibly produced the observed cognitive changes, using temporally ordered evidence and mechanistic indicators.

---

## Case Studies: Illustrative Episodes of Industrialization-Driven Cognitive Change

- Case A: Early industrial-era conflict showing telegraph-rail networks compressing operational timelines and prompting staff-system innovations that shifted cognitive burdens.
- Case B: 20th-century total war demonstrating large-scale propaganda, institutionalized training, and mass mobilization changing public sensemaking and command heuristics.
- Case C: Late-industrial/early-information conflict illustrating interaction between industrial infrastructures (broadcast and telecom) and emerging algorithmic amplification that enables tailored influence at scale.

Each case traces institutional practices, communicative infrastructures, and cognitive outcomes against expectations from the theoretical framework.

---

# Claims and Hypotheses

Primary claim: Industrialization systematically reshapes the cognitive environment of war, producing new strategic logics and vulnerabilities.

## Hypotheses

- H1: Higher degrees of industrialization correlate with compressed decision cycles and greater reliance on delegation and pre-planned routines.
- H2: Industrialized communication infrastructures increase opportunities for cognitive influence but also raise susceptibility to disinformation and overload.

Empirical markers: MTTA (mean time to acknowledge), delegation index (share of decisions automated or pre-delegated), and information-load metrics (messages per node per hour).

---

# Applications: Operational Vignettes (Parameterized)

This section provides two parameterized vignettes that translate theorized mechanisms into operational metrics, failure probabilities, and explicit failure modes. The vignettes are configurable by parameters (communication reliability, adversary capability, automation level) to support doctrine evaluation.

Vignette 1 — Disaster Response Under Intermittent Communications

Scenario summary: A coastal state faces a major hurricane; civilian authorities use AI-driven influence systems to direct evacuation and distribute safety information. Communications are partially degraded (cell towers intermittent, satellite capacity limited). Automated message-generation systems produce localized advisories and A/B test variants to optimize compliance.

## Parameters

- Comms availability (p_comm): probability that a given node (local broadcast/relay) is reachable in a given hour; range: 0.2–0.95.
- Automation fraction (α): fraction of advisories auto-generated and auto-sent without human review; range: 0.0–0.9.
- Adversarial injection capability (λ): expected rate of adversarial false messages injected into the local mesh per hour; range: 0–50.

## Metrics

- MTTA (mean time to acknowledge): expected time from advisory generation to confirmed receipt by majority of target nodes.
- Failure probability (P_fail): probability that a critical advisory (evacuation order) fails to produce intended behavioral compliance within required window (e.g., 6 hours).
- False-positive propagation rate (FP): fraction of nodes that propagate adversarial or erroneous advisories as authentic.

## Tradeoffs and expected behaviors

- With high α and low p_comm, MTTA decreases for messages that traverse robust channels but increases overall due to retries and conflicting variants. High automation reduces latency for generation but increases FP if validation is weak.
- P_fail increases with λ and low p_comm because adversarial injections create confusion and decision paralysis; adding decentralized consensus (majority validation among trusted relays) reduces FP but increases MTTA.

## Failure modes

1. Signal conflation: multiple competing advisories—automated variants, human-updated advisories, adversarial messages—produce contradictory instructions; delegation policies that permit overwriting without human arbitration cause P_fail spikes.
2. Validation collapse: reliance on centralized signatures fails when key infrastructure is degraded; trust transitivity breaks down and local nodes default to heuristics (follow loudest/best-timed signal), enabling adversarial capture.
3. Overfitting to short-horizon metrics: A/B testing optimizes immediate clicks (compliance proxies) but amplifies short-term compliance at cost of long-term trust, increasing susceptibility to repeated adversarial exploitation.

## Mitigations (operational examples)

- Conservative automation cap: enforce α <= 0.3 under p_comm < 0.6 and require human-in-the-loop escalation for evacuation-level advisories.
- Local quorum validation: require ≥ k trusted relays to endorse an advisory before it is auto-amplified; k parameter trades MTTA vs. FP.
- Fallback canonical channel: reserve a hard-to-spoof channel (e.g., authenticated broadcast via emergency sirens or pre-distributed tokens) to anchor trust when digital meshes are contested.

## Vignette 2 — Autonomous ISR Swarm with Contested Spectrum

Scenario summary: An ISR (intelligence, surveillance, reconnaissance) swarm of small drones performs wide-area monitoring. Decision pipelines include onboard ML classifiers that autonomously tag events and feed a distributed influence system that produces alerts to commanders and adjacent civilian systems.

### Parameters

- Autonomy level (β): fraction of classification and tagging performed locally without human confirmation (0.0–1.0).
- Spectrum contest factor (σ): expected packet loss and jamming-induced delay rate in comms between swarm nodes and base (0–0.7).
- Adversary spoof efficacy (ψ): probability that adversary can inject fake sensor signatures or confuse classifiers per hour (0–0.5).

### Metrics

- MTTA_alert: mean time from event occurrence to commanding officer acknowledgment.
- P_false_alarm: probability that an automated alert is false and triggers an unnecessary kinetic or policy action.
- Mission degradation probability (P_miss): probability that adversary actions prevent detection of a critical event within mission window.

### Tradeoffs and behaviors

- High β reduces MTTA_alert but increases P_false_alarm under ψ > 0 due to classifier exploitation and data-poisoning attacks. Higher σ increases both MTTA_alert and P_miss by fragmenting consensus among nodes.
- Distributed consensus protocols can mitigate false alarms by requiring corroboration among k spatially separated nodes, but this increases MTTA_alert and can raise P_miss if the adversary partitions the network.

### Failure modes

1. Consensus paralysis: under high σ, nodes cannot achieve quorum; automated pipelines delay alerts until human review, increasing MTTA_alert and P_miss.
2. Classifier exploitation: adversary crafts signals that cause systematic false positives (resource wasting) or false negatives (missed detection); industrial-scale automated retraining amplifies failures if poisoned datasets are fed back into training loops.
3. Delegation cascade: commanders, accustomed to low-latency automation, accept automated alerts; when automation is compromised, humans are insufficiently primed to reassert control in time-sensitive windows.

Operational parameters for doctrine

- Define β thresholds conditional on σ and ψ (e.g., β <= 0.4 when ψ>0.2; increase mandatory human adjudication for high-consequence alerts).
- Use diversified sensing modalities (RF, EO, acoustic) to reduce correlated exploitation risks; require multi-modal corroboration before kinetic actions.
- Monitor retraining pipelines for dataset drift and implement data provenance checks to detect poisoning. Recent work on ML detection and cyber-attack detection informs practical approaches to adversarial robustness and anomaly detection in such pipelines [1].

Collectively, these vignettes show how industrialized AI systems change the MTTA/failure-probability trade space and why explicit delegation rules and validation quorums are operationally necessary.

---

## Synthesis

Industrialized AI amplifies both the capacity to shape cognition and the fragility of cognitive systems. From the mechanisms and vignettes, three synthetic conclusions emerge.

1. Compression–Fragility Tradeoff: Faster automated cycles (lower MTTA) reduce reaction time but increase systemic fragility when adversarial inputs or infrastructure faults occur. Organizations must choose operating points (automation caps, quorum sizes) that trade latency for robustness.

1. Standardization Bias: Industrial practices favor templated influence primitives and reusable pipelines. This increases scale and repeatability but reduces heterogeneity that can absorb novel attacks; heterogeneity (diverse models, varied communication stacks) is a defensive asset.

1. Observable Diagnostics Enable Policy Levers: Metrics such as MTTA, delegation index, corroboration quorum, and information-load per node are actionable levers. Continuous monitoring of these diagnostics allows adaptive delegation policies tied to measured adversary activity and infrastructure health.

These syntheses bridge theory and operational doctrine: they point toward layered mitigation strategies that combine architectural hardening, organizational policies, and measurement-driven thresholds.

---

## Implications for Policy and Military Doctrine

- Doctrine should adopt flexible automation thresholds tied to measured comms integrity and adversary activity.
- Information-architecture resilience (diversity, authenticated anchors, local quorums) should be prioritized alongside kinetic capabilities.
- Training should emphasize re-acquisition of manual decision competencies under degraded automation to avoid delegation cascades.

---

## Limitations and Further Research

- Temporal complexity and evolving AI capabilities limit predictive generalizability; large-N quantitative tests across different conflict types are needed.
- Interactions between AI industrialization and other technological trends (e.g., quantum communications, global data governance) remain open questions.
- Empirical work to calibrate thresholds ($\tau_1$, $\tau_2$, $\varepsilon$, $\gamma$) across cultural and organizational contexts is necessary to make delegation policies operationally robust.

---

## Conclusion

AI industrialization of influence reshapes warfare by scaling and accelerating cognitive effects. A theory-first, mechanism-led approach yields operational diagnostics and delegation policies that balance speed and resilience. Implementing measurement-driven thresholds, diversification, and conservative automation caps will be central to doctrine in an era of Cognitive Wars.

---

# Assumptions Ledger

| Assumption | Rationale | Observable | Trigger | Fallback/Delegation | Scope |
|---|---|---|---|---|---|
| AI-driven industrialization materially shortens decision cycles (OODA) and thereby shifts organizations toward heuristic and delegated decision-making. | Automation, rapid feedback from large-scale campaigns, and pipelineed AI systems reduce latency between observation and effect, making manual deliberation slower relative to action. Historical and theoretical accounts of tempo effects support this mechanism. | Measured reduction in decision latency (time from stimulus to action), increased fraction of actions taken by automated agents (audit logs), growth in automated approval/win rates, and operational dashboards showing higher-frequency campaign iterations. | Sustained increases in automated message generation or campaign iteration rates; alerts showing reduced human-in-the-loop approvals; operational failures correlated with fast cycles (e.g., erroneous automated actions). | Introduce hard throttles and human-in-the-loop checkpoints for high-risk decisions; implement escalation gates and manual overrides; limit automation to low-consequence tasks while retaining humans for ambiguous/contextual judgment. | Applies to organizations that deploy pipelineized, low-latency AI systems for influence or decision support in connected information environments; limited in low-connectivity theaters, highly regulated contexts, or where compute/latency constraints prevent aggressive automation. |
| AI-enabled personalization and continuous experimentation massively increase information volume and diversity, degrading signal-to-noise and forcing reliance on algorithmic filters and trusted nodes. | Personalization and A/B testing multiply message variants and delivery channels, producing a higher throughput of signals. Distributed detection literature shows that higher noise and sparse observability reduce reliable inference without stronger filters or trusted aggregators. | Sharp increases in message/variant counts, higher variance in audience response metrics, rising CPU/load on filtering systems, degraded precision/recall in moderation/detection, and increased false positives/negatives reported by downstream analysts. | Filter saturation alerts, sudden declines in detection accuracy, unexplained volatility in public opinion or engagement metrics, or monitoring systems reporting throughput beyond designed capacity. | Prioritize high-precision detection over high-recall for critical channels; route ambiguous cases to human analysts; deploy provenance metadata and rate-limits; create and trust vetted aggregator nodes; and diversify sensing (cross-platform feeds). | Most relevant in high-bandwidth, digitally mediated information ecosystems (social platforms, messaging apps, targeted ad ecosystems). Less applicable in analog media environments or where audience reach is inherently limited. |

| Assumption | Rationale | Observable | Trigger | Fallback/Delegation | Scope |
|---|---|---|---|---|---|
| Standardization and routinization (templates, playbooks, SOPs) increase scale and repeatability of influence operations but reduce local adaptability, increasing fragility to novel or adversarially adaptive attacks. | Standardized procedures enable scale and predictable outcomes but constrain improvisation. Historical parallels (industrial-era SOPs) show reproducibility trades off with adaptability; standardized templates are easier for adversaries to anticipate and exploit. | Homogeneous responses across units or campaigns, low intra-organizational variance in tactics, repeated failures when facing novel adversary techniques, and after-action reports noting inability to adapt SOPs to new conditions. | Emergence of novel influence tactics that consistently bypass SOPs, observed exploitation of standard playbooks by adversaries, or stagnating performance despite resource increases. | Decentralize decision authority for frontline/contextual choices, maintain rapid-adaptation teams to revise playbooks, introduce randomized or variant SOPs, and preserve human judgment in exception cases. | Applies where central authorities enforce standardized procedures for scale (national campaigns, large institutions). Less relevant in highly distributed, adaptive cells or small-scale grassroots campaigns where standardization is minimal. |
| Automation of experimentation (continuous A/B testing and causal measurement) accelerates adaptation and effectiveness of influence operations, disproportionately advantaging actors with greater compute, data, and integration capabilities. | Continuous automated experimentation requires compute, data infrastructure, and analytics expertise. Actors with more resources can iterate faster and discover effective manipulative affordances before defenders, entrenching asymmetries. | Correlation between actor resources (compute, data access) and rapid improvement in campaign metrics, detection of systematic automated experimentation pipelines, and emergence of novel tactics traceable to high-iteration processes. | Widening performance gaps between actors, discovery of large-scale experimentation tooling, sudden leaps in adversary effectiveness tied to deployment of automated pipelines, or intelligence indicating resource-backed optimization efforts. | Pool defensive resources across coalitions (shared datasets, compute), open-source detection/benchmarking tools, impose policy/regulatory limits on large-scale experimentation, and accelerate defensive research into adversarial-robust detection. | Most salient in contests between state or well-resourced non-state actors and less-resourced opponents in digital ecosystems. Less relevant when actors are resource-symmetric or when legal/regulatory constraints limit large-scale experimentation. |
| Networked delegation and distributed sensing architectures create distinct failure modes (partitioning, | Consensus, distributed detection, and Byzantine-fault literature demonstrate that partial observability, | Divergence between local and aggregated inferences, rising disagreement/confidence variance across sensors/agents, detection of clustered coordinated inauthentic | Increased agent disagreement, anomalies in consensus protocol metrics, detection of suspicious node | Harden consensus (Byzantine-tolerant protocols), increase sensor redundancy and diversity, quarantine or deprioritize suspect nodes, require cross-source corroboration, | Applies to distributed architectures that rely on many semi-autonomous sensing/decision nodes |

| Assumption | Rationale | Observable | Trigger | Fallback/Delegation | Scope |
|---|---|---|---|---|---|
| Sybil-like injection, consensus breakdown) that can prevent accurate global inference and enable adversarial manipulation. | correlated failures, and injected fake agents undermine aggregation. Networked delegation multiplies the attack surface and complicates reliable fusion of local signals. | behavior, and repeated inability to reach stable consensus on key indicators. | behavior patterns, or failures in cross-validation between independent sensing sources. | and elevate contested inferences to human adjudication cells. | (networked social sensing, federated decision systems). Less applicable to tightly centralized command-and-control systems or completely isolated single-source setups. |

## Notation

| Symbol | Meaning | Units / Domain |
|---|---|---|
| \(n\) | number of agents | \(\mathbb{N}\) |
| \(G_t=(V,E_t)\) | time-varying communication/interaction graph | — |
| \(\lambda_2(G)\) | algebraic connectivity (Fiedler value) | — |
| \(p\) | mean packet-delivery / link reliability | [0,1] |
| \(\tau\) | latency / blackout duration | time |
| \(\lambda\) | task arrival rate | 1/time |
| \(e\) | enforceability / command compliance | [0,1] |
| \(\tau_{\text{deleg}}\) | delegation threshold | [0,1] |
| **MTTA** | mean time-to-assignment/action | time |
| \(P_{\text{fail}}\) | deadline-miss probability | [0,1] |

# Claim-Evidence-Method (CEM) Grid

| Claim (C) | Evidence (E) | Method (M) | Status | Risk | TestID |
|---|---|---|---|---|---|
| Primary: Industrial-scale AI compresses decision cycles (shortened OODA loops) producing a structural shift toward heuristic and delegated decision-making within organizations. | [2] [1] | Validate via comparative historical process-tracing (empirical cases of tempo change), organization-level field studies, and agent-based simulations that vary automation latencies and decision-delegation policies. | E cited; M pending (sim + empirical casework) | If false, doctrines premised on rapid automated delegation may be unnecessary or harmful; resources could be misallocated to automation instead of human-in-the-loop defenses, producing coordination failures or overconfidence. | T1 |
| Primary: AI-driven increases in information volume and personalization amplify signal-to-noise problems, raising Type I/II error rates and degrading detectability/observability in sparse, noisy networks. | [1] [3] [5] | Analytic derivation of detectability bounds (from consensus/observer theory) and Monte Carlo / adversarial simulations of observer networks subject to high-volume, personalized injections; empirical validation via measurement of false positive/negative rates in live or historical messaging campaigns where available. | E cited; M pending (analytic proofs + sim + selective empirical tests) | If false, investment in heavyweight filtering or consensus augmentation may be misdirected; alternatively, overestimating degradation leads to unnecessary censorship or suppression of legitimate signals. | T2 |
| Primary: Standardization and routinization (templates, SOPs, playbooks) increase reproducibility and scale but reduce local improvisation, thereby increasing fragility to novel or adversarial tactics. | [2] [7] | Comparative empirical case studies (process tracing of protocolized vs. non-protocolized responses), red-team experiments injecting novel adversarial tactics against SOP-driven workflows, and resilience metrics from simulation. | E cited; M pending (case studies + red-team sim) | If wrong, recommendations to resist standardization could unnecessarily hinder beneficial efficiency gains; if unaddressed when true, SOP-driven systems could be widely exploited, causing systemic failures. | T3 |
| Secondary: Automation of experimentation (continuous A/B testing at scale) materially accelerates adaptive influence engineering, concentrating advantage among actors with greater compute/data and | [1] [6] | Empirical measurement of iteration rates and effect sizes in documented influence operations; simulation of learning curves under different compute/data budgets to | E cited; M pending (empirical + sim) | If false, policy focus on limiting automated experimentation may be misplaced; if true and unmitigated, democratic actors will be disadvantaged and | T4 |

| Claim (C) | Evidence (E) | Method (M) | Status | Risk | TestID |
|---|---|---|---|---|---|
| entrenching manipulative affordances. | | show cumulative advantage; causal inference on the role of automation in campaign success. | | manipulation may become persistent and hard to reverse. | |
| Secondary: Networked delegation and consensus limitations produce concrete failure modes—partitioning, sybil-like injection, and observability gaps—that can prevent accurate global inference under adversarial influence. | [3] [4] [5] [7] | Formal proofs/derivations of detectability and consensus limits under adversarial models; network simulations that instantiate partitions and sybil attacks; lab or field experiments with distributed observer prototypes to observe failure modes. | E cited; M pending (formal proofs + sim + experiments) | If false, distributed delegation architectures may be more robust than assumed and resources spent redesigning them could be wasted; if true and not mitigated, critical distributed decision systems may be spoofed or split, leading to miscoordination or catastrophe. | T5 |
| Secondary/Conceptual: 'Cognitive wars' are conflicts where shaping cognition (beliefs, attention, expectations, decision heuristics) is a primary strategic objective; AI industrialization magnifies these modes by scaling personalization, automation, and feedback-driven adaptation. | [2] [1] | Conceptual operationalization followed by multi-case comparative validation: map instances where cognitive shaping was central, measure the relative role of AI/automation in scaling those activities, and test whether cognitive effects correlated with strategic outcomes. | E cited; M pending (conceptual mapping + comparative case validation) | If incorrect, doctrine and policy premised on cognitive priority could misallocate limited resources away from material deterrence or kinetic capabilities; misframing could lead to ineffective or counterproductive interventions. | T6 |

## References (anchors cited inline)

[1]: ArXiv preprint on ML approaches for cyber-attack detection (2024) — technical grounding for ML detection and adversarial robustness.

[2]: Canadian Pediatric Massive Hemorrhage Protocols: a peer-reviewed example of protocolization and institutional standardization (2023) — used as a peer-reviewed anchor for routinization tradeoffs.

[3]: Conditions for detectability in distributed consensus-based observer networks (2013) — foundational for detectability limits in networked sensing.

[4]: Comments on "Consensus and Cooperation in Networked Multi-Agent Systems" (2010) — conceptual clarifications on consensus vulnerabilities.

[5]: On graph theoretic results underlying the analysis of consensus in multi-agent systems (2009) — graph-theoretic constraints on consensus and observability.

[6]: A Brief Tutorial on Consensus ADMM for Distributed Optimization with Applications in Robotics (2024 preprint) — mechanistic details of distributed optimization applicable to drone swarms and quorums.

[7]: A Survey of Distributed Consensus Protocols for Blockchain Networks (2019) — insights into sybil-like injection and validation tradeoffs.

## Sources

**[1]**
An Investigation into the Performances of the State-of-the-art Machine Learning Approaches for Various Cyber-attack Detection: A Survey
Arxiv.Org, 2024-02-26. (cred: 0.50)
http://arxiv.org/abs/2402.17045v2

**[2]**
OA1-AM23-SN-05 | Canadian Pediatric Massive Hemorrhage Protocols: A Survey of National Practice and State-of-the-Art Review
Doi.Org, 2023-10-01. (cred: 0.50)
https://doi.org/10.1111/trf.52_17554

**[3]**
Conditions for detectability in distributed consensus-based observer networks
Arxiv.Org, 2013-03-26. (cred: 0.50)
http://arxiv.org/abs/1303.6397v1

Generated: 2025-11-07T16:43:41.794627 | Word Count: 4831

---

### Research Roadmap

- **Phase 1 (Theory)**: Formalize claims, extend proofs, validate against canonical results
- **Phase 2 (Simulation)**: Implement stress tests, sweep parameter spaces, measure convergence/scaling
- **Phase 3 (Empirical)**: Deploy in controlled environments, collect field data, validate predictions
- **Phase 4 (Integration)**: Operationalize with human-in-loop, adversarial hardening, production deployment

---

**Confidence Methodology:** Confidence = 0.3·SourceDiversity + 0.25·AnchorCoverage + 0.25·MethodTransparency + 0.2·ReplicationReadiness, where SourceDiversity reflects unique publishers & types, AnchorCoverage reflects share of primary claims with Type-1 anchors, MethodTransparency reflects CEM completeness & assumptions ledger, and ReplicationReadiness reflects sim plan & datasets/params specified.

---

Prepared under the STI Research Program — theoretical framework subject to revision as data accumulate.