THESIS BRIEF — THEORY-FIRST RESEARCH

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

Smart Technology Investments

Cognitive Wars: the AI Industrialization of Influence

Aug 13–Nov 11, 2025 | Sources: 13 | Anchor Status: Anchored | Report Type: Theoretical Research | Horizon: Near-term | Confidence: 0.800 *

SD	AC	MT	RR
0.80	1.00	0.70	0.65

Alignment: 6.0 Theory Depth: 6.0 Clarity: 7.0

{% if route_rationale %}

Route: {{route_rationale.route|upper}} — MarketScore {{'%.3f'|format(route_rationale.market_score)}} | Fresh {{route_rationale.fresh}}/{{route_rationale.total}}, Unique domains {{route_rationale.unique_domains}}, Anchors {{route_rationale.anchors}}, Canonical {{route_rationale.canonical}}

{% endif %}

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.

Abstract & Theory-First Framing.

Outline

Introduction and Thesis Statement

This brief advances a theory-first argument that "cognitive wars" are best understood as a distinct modality of conflict produced and shaped by processes of industrialization. Cognitive wars are organized efforts by state and non-state actors to alter, disrupt, or control mass cognition, beliefs, attention, norms, and decision heuristics at scale. Industrialization — understood broadly as the historical and continuing process that converts social, informational, and communicative capacities into routinized, scalable, and cost-efficient productive forms — exerted structural and institutional influences that enabled warfare to target cognition at scale.

Central thesis: industrialization created the infrastructural, organizational, and technological preconditions that permit deliberate, high-throughput, and persistent targeting of human cognition. The argument proceeds by (1) defining cognitive wars and distinguishing them from kinetic, economic, and cyber modalities, (2) tracing how waves of industrialization enabled new capacities for cognitive targeting, (3) specifying mechanisms that link industrial forms to influence operations, and (4) demonstrating the theory with comparative and contemporary cases, including digital-industrial configurations amplified by generative AI and algorithmic amplification [6][7][12].

Theoretical Framework: Theory-First Approach

A theory-first approach prioritizes generative causal mechanisms over ad hoc empirical description when explaining the emergence of cognitive wars. The aim is to build middle-range theory that connects macro-level transformations (industrialization) to meso-level institutions (media, education, bureaucracy) and micro-level cognitive effects (attention capture, belief updating, norm adoption). Key commitments:

- Process tracing: identify sequential mechanisms that plausibly transmit industrial properties into cognitive warfare capacities.
- Mechanistic explanation: specify component parts (infrastructure, institutions, production technologies, organizational form) and how they interact.
- Falsifiable propositions: derive observable indicators (e.g., diffusion speed, institutional legibility, penetration of mass schooling) that can be empirically tested.

This framework privileges mechanisms that are necessary and generative rather than purely correlative descriptions.

Foundations

Why these anchors? The literature anchors selected here are chosen to balance direct studies of automated influence and cognitive operations with peer-reviewed, non-preprint works that capture mechanisms and empirical regularities. Where canonical reports or high-impact technical reviews are included (even when not traditional journal articles), they are treated as theoretical anchors because they synthesize cross-disciplinary evidence and foresight about misuse. Anchor selection strategy:

- Direct sources (Layer 1): empirical and analytic studies of automated influence, social bots, and computational propaganda to ground the specific phenomenon (e.g., operational patterns, botnet dynamics) [9][11].
- Domain sources (Layer 2): work addressing information warfare, influence operations and automated influence, including integrated reviews of electronic/automated systems in contested environments [4][7].
- Foundational sources (Layers 3–5): canonical multidisciplinary reports and journal articles that supply first-principles theories of diffusion, persuasion, and systemic vulnerability to large-scale interventions (e.g., forecasts of malicious AI capabilities) [12][6].

The selected anchors are primarily peer-reviewed conference and journal publications (ACM/IEEE/Springer/ScienceDirect) where available, supplemented by multidisciplinary technical syntheses recognized in policy and research communities [2][5]. These materials enable a theory-first mapping from institutional form to cognitive effect while preserving empirical touchstones.

Theoretical Grounding and Conceptual Framework

Abstraction layers and key concepts

- Layer 1 (Specific): cognitive warfare, AI influence operations, computational propaganda, automated disinformation, deepfakes, synthetic media — the observable toolkit and manifestations of cognitive targeting [6][11].
- Layer 2 (Domain): information warfare and PSYOP, strategic communication, propaganda studies — the normative and institutional context for organized influence [7].
- Layer 3 (Methods/Mechanisms): social bots/botnets, algorithmic amplification, microtargeting, automated content generation — the operational instruments and pipelines that enact cognitive interventions [9][5][8].
- Layer 4 (Theoretical Frameworks): persuasion and attitude-change theory, diffusion models, network science, behavioral economics — models that explain how messages influence beliefs and propagate through networks [2].
- Layer 5 (Foundational Principles): information theory, game theory, social psychology — the formal constraints and incentives shaping signaling, strategic interaction, and conformity dynamics.

Reasoning chain (foundations → specific):

1. Foundational principles (Layer 5) define limits on signaling and inference (e.g., noisy channels, bounded rationality).
2. Theoretical frameworks (Layer 4) model how individual updating and network ties produce aggregate opinion dynamics under constraints of attention and bias.
3. Methods/mechanisms (Layer 3) deliver interventions that exploit these updating dynamics (bots amplify, personalization exploits heuristics, generative models scale production) [9][5].
4. Domain institutions (Layer 2) — media firms, education systems, state propaganda apparatus — provide channels, legitimacy, and organization for deployment [7].
5. Specific phenomena (Layer 1) emerge when industrialized production and automation make cognitive influence continuous, low-cost, and high-volume, producing what we call “cognitive wars” [6][12].

Canonical papers from broader layers are included not because they directly study every contemporary instance, but because their formal claims (e.g., diffusion dynamics, persuasion constraints, adversarial capability forecasts) set constraints and expectations for measurable outcomes in specific cases.

The conceptual map of "Cognitive Wars: The AI Industrialization of Influence" follows this chain: industrial properties → institutional capacities → technical mechanisms → population-level cognitive outcomes.

Conceptualizing "Cognitive Wars"

Definition: cognitive wars are organized, sustained efforts by political actors (state and non-state) to modify, suppress, or steer mass cognition — including attention, beliefs, preferences, norms, and decision heuristics — using sociotechnical means that are routinized, replicable, and scalable.

Key differentiators from other modalities:

- Target: cognitive (informational/mental states) rather than primarily kinetic (physical) or economic.
- Means: emphasis on information production, dissemination, and persuasion pipelines rather than weapon systems or sanctions.
- Metrics of success: shifts in public opinion, policy preferences, behavioral compliance, and normative alignment rather than territorial gains.

Operational criteria for identification: deliberate targeting of (a) attention (salience manipulation); (b) belief formation (fact framing, falsehood injection); (c) norms (social proof, moral narratives); and (d) heuristics (risk framing, uncertainty exploitation), executed at population scales and using industrialized media/information infrastructures [6][11].

Industrialization and the Evolution of Warfare

Industrialization transformed the scale, speed, and organizational capacity of states and non-state actors, producing prerequisites for modern cognitive warfare. Key points:

- Scale and standardization: mass printing, broadcasting, and schooling enabled uniform messaging to large, legible audiences; industrial production logic (repeatability, division of labor) translated into repeatable persuasion operations.
- Infrastructure repurposing: communication networks built for commerce and governance (rail, telegraph, radio, internet) became channels for coordinated influence operations at previously impossible speed and reach [1][6].
- Organizational capacity: bureaucracies, ministries, and commercial PR firms provided centralized planning, resource mobilization, and professional expertise to run sustained campaigns.

Historical correlation: major shifts in propaganda and mass persuasion maps onto waves of communicative industrialization — from print press and mass schooling to broadcast radio and now platformized social media amplified by algorithmic recommender systems and generative AI [7][12].

Influence of Industrialization on Cognitive Domains

Industrialization shaped cognitive repertoires by standardizing curricula, normalizing repetitive media exposures, and organizing social life into schedules and workplaces that condition attention. Concretely:

- Mass literacy and centralized curricula created predictable informational priors and shared repertoires that are manipulable via standardized narratives.
- Institutionalization (state schools, public broadcasters) produced dependable channels for messaging and markers of legitimacy that influence uptake.
- The logic of industrial production (efficiency, scale, modularization) migrated into persuasion practices: message templates, segmentation (microtargeting), A/B testing, and content factories professionalized cognitive influence [7][11].

These changes increased both susceptibility (predictable heuristics to exploit) and reach (ability to flood attention markets at low marginal cost), lowering the cost and raising the impact of coordinated cognitive operations.

Mechanisms Linking Industrialization to Cognitive Wars

This section enumerates mechanisms without restating the executive summary.

Mechanism 1 — Infrastructure (distribution pipelines): industrial communication and transport networks reduce latency and widen reach. Cheap, high-throughput channels enable simultaneous multi-modal deployment (print, broadcast, digital) and rapid repetition necessary for memory consolidation and salience effects [4][6].

Mechanism 2 — Institutionalization (legibility & channels): bureaucratic education and media institutions create predictable population segments and trusted vectors; these are exploitable for targeted narrative insertion and scaling of influence operations [1][7].

Mechanism 3 — Technology and production (content factories): mass printing, recording, and now generative models lower marginal costs per persuasive item. Automated content generation combined with personalization pipelines permits high-volume, low-cost tailoring to microaudiences — multiplying per-message effectiveness through relevance and reinforcement [12][8].

Mechanism 4 — Organizational form (centralized campaign architectures): industrial organizations (propaganda ministries, intelligence units, PR firms) integrate planning, monitoring, and feedback loops (campaign metrics, engagement analytics). This professionalization allows iterative optimization (A/B testing, bot amplification strategies) akin to industrial process control [5][9].

Mechanism 5 — Algorithmic amplification and marketization: platform recommender systems and attention markets introduce automated non-linear amplification where content that maximizes engagement is preferentially distributed; industrial actors can game these systems at scale through coordinated behavior (bots, sockpuppets, farmed engagement) [11][9].

Mechanism 6 — Adversarial co-evolution: as influence tools industrialize, defensive systems and norms lag or adapt unevenly, creating gaps and arms races in detection, attribution, and governance [5][6].

Methodology: Theory-Driven Case Selection

Cases are selected to test the mechanisms across industrialization stages and political contexts: early industrial (print + schooling), mass-industrial (broadcast + wartime propaganda bureaus), and digital-industrial (platformized, algorithmic, AI-augmented). Methodological tools:

- Process tracing to identify causal sequences from institutional/technical features to cognitive outcomes.
- Comparative historical analysis to evaluate the presence/absence and magnitude of proposed mechanisms.
- Observable indicators: institutional capacity for mass messaging (budget, staffing, legal authority), diffusion speed (time to reach X% exposure), education penetration (literacy rates), normative shift measures (survey trend changes, behavioral proxies), and platform engagement metrics.

Empirical strategies include archival research, content and network analysis, and where ethical, field or laboratory experiments that assess message resonance under controlled exposure.

Illustrative Case Studies

Case A — Early industrial era (print press + schooling): mass conscription era propaganda used standardized textbooks, posters, and leaflets distributed via postal and rail networks to mobilize populations. Mechanisms: infrastructure (print runs, rail distribution), institutionalization (state school curricula and patriotic rituals), and organizational form (government information ministries). Observable outcome: measurable increases in recruitment rates and popular support correlated with exposure intensity.

Case B — Mass-industrial era (radio + wartime propaganda bureaus): radio bureaus coordinated emotional framing at national scale; they leveraged centralized broadcast monopolies and routine programming to prime audiences and coordinate morale across industrial workforces. Mechanisms: centralized channels, repetition schedules, professionalized messaging. Indicators: changes in national sentiment proxies, rumor damping metrics, and wartime production compliance.

Case C — Digital-industrial era (platforms + algorithmic amplification + generative AI): commercial platforms, combined with automation (bots) and generative models, enable continuous, highly personalized influence campaigns that can generate tailored narratives and synthetic media at scale [9][11][12]. Mechanisms: content factories, algorithmic amplification, microtargeting, and coordinated organizational campaigns (state or commercial). Outcomes: measurable shifts in attention metrics, viral propagation, and measurable behavioral changes (e.g., turnout effects). Empirical traces include botnet growth patterns, content supply correlations with attention spikes, and automated narrative recombination patterns [9][11][8].

Each case demonstrates how distinct industrial configurations enabled cognitive targeting through identifiable mechanisms, and how later stages compound earlier capacities rather than simply replacing them.

Applications: Parameterized Vignettes and Metrics

Vignette 1 — Disaster response under intermittent communications

Scenario: A coastal region experiences a major hurricane; communications infrastructure is intermittently available. Two actors operate: an emergency management agency (EMA) seeking to coordinate evacuation and an adversarial influence actor (AIA) aiming to disrupt compliance and sow confusion.

Parameters:

- Network availability: fraction of population with intermittent connectivity (p_net = 0.4–0.8).
- Message channels: SMS blast system (capacity C_sms = 100k msgs/hr), community radio (coverage R_cov = 70%), peer-to-peer mesh (ad hoc, reach variable).
- Latency: mean delivery latency L = 5–60 minutes depending on channel load.
- Agent resources: EMA has verified shortcodes and content templates; AIA has botnets and capacity for localized deepfake voice messages (rate D = 500 msgs/hr).

Metrics (example operational metrics):

- MTTA (mean time to action): average time from EMA broadcast to measurable evacuation movement (minutes/hours).
- Failure probability (P_fail): probability that a critical fraction (>30%) of target population does not evacuate within policy window.
- Confusion index (CI): measured heterogeneity in received instructions (number of conflicting directives per 1,000 recipients).

Failure modes:

- Channel spoofing: AIA mimics EMA shortcode leading to instruction ambiguity (raises CI and P_fail).
- Amplification lag: platform outages cause delayed message dissemination, increasing MTTA.
- Trust erosion: prior AIA campaigns decreased baseline trust, so even verified messages have lower uptake.

Operational mitigation and thresholds:

- Verified multi-modal confirmation: EMA policy requires two independent channels (SMS + radio) before evacuation order; if p_net < 0.5, escalate to in-person alert via community leaders.
- Delegation policy: If CI > 0.2 (20% recipients report conflicting directives), local authorities are authorized to deploy door-to-door teams.

Vignette 2 — Autonomous ISR swarm with contested spectrum

Scenario: An ISR (intelligence, surveillance, reconnaissance) drone swarm provides situational awareness to civil authorities. An adversary conducts cognitive operations to inject false sensory cues and targeted social media disinformation timed to reduce public compliance with safety directives.

Parameters:

- ISR latency (L_isr): 2–10 seconds sensor-to-operator.
- Decision horizon (H): 30–120 minutes for tactical public guidance.
- Adversary capabilities: RF spoofing probability p_spoof = 0.05 per hour; synthetic imagery injection rate S_img = 10 images/hr into local social feeds.

Metrics:

- MTTA_recon: time from ISR detection of anomaly to public advisory issuance.
- Failure probability for advisory uptake (P_fail_adv): probability advisory is ignored or actively countered due to conflicting social media narratives.
- Misinformation penetration (MP): fraction of exposed population that interacts with false imagery.

Failure modes:

- Sensory deception: adversary creates plausible synthetic imagery that matches ISR reports, undermining operator confidence.
- Narrative synchronization: adversary times synthetic posts to coincide with advisory issuance, reducing uptake.

Mitigation policies:

- Diagnostics: require cross-validation from at least two independent ISR platforms before public advisories if p_spoof exceeds threshold (0.02/hr) or if operator confidence score < 0.8.
- Delegation: in high-contestation states, issue micro-targeted localized advisories via verified community channels rather than broad platform posts; rely on trusted local emissaries to counter synchronous disinformation.

Common notes on metrics and risk

- MTTA should be decomposed into detection time, verification time, and dissemination time to isolate where cognitive operations are most effective.
- Failure probabilities should be estimated with Bayesian updating as new exposure and trust data arrive; credible intervals must accompany point estimates.
- Practical deployment requires monitoring ensemble signals (engagement spikes, bot indicators, mismatched metadata) to detect coordinated industrialized influence attempts [9][11][15].

These vignettes illustrate how industrialized influence capabilities (automation, content factories, coordinated amplification) materially change operational risk profiles and demand infrastructural and institutional mitigations rather than only content moderation.

Implications for Policy and Theory

Policy implications derive from the industrial character of cognitive wars:

- Infrastructure and institution focus: resilience requires investments in trusted communal information infrastructures (education, local media, verification registries) and decentralization of critical channels so no single mode of failure enables widespread manipulation [7][6].
- Governance of production pipelines: regulation and standards for provenance, metadata integrity, and platform incentive structures (algorithmic transparency, friction against virality for unverified content) reduce exploitability.
- Capacity and norms: build interagency analytic capacities for attribution and proportionate response, while codifying human rights and due process in automated mitigation workflows.

Theoretical implication: warfare and security scholarship should integrate socio-technical industrial analysis (production processes, routinization, market forces) as a structural complement to unit-level actor analyses — explaining not only who attacks but how industrial modalities shape what is feasible and effective.

Conclusion and Agenda for Future Research

Summary: Industrialization — the routinization, scaling, and commodification of information production and distribution — is a causal antecedent shaping the emergence, form, and effectiveness of cognitive wars. The mechanisms identified (infrastructure, institutionalization, production technologies, organizational form, algorithmic amplification, and adversarial co-evolution) provide a tractable theory for empirical testing.

Empirical tests and data priorities:

- Archive and comparative study of propaganda institutions across industrialization phases.
- Cross-national measures of infrastructural centralization, education penetration, and platform market structure correlated with susceptibility indicators.
- Experimental and quasi-experimental evaluation of large-scale cognitive interventions with robust ethical safeguards.

Next steps: refine causal pathways into quantitative models (e.g., diffusion models parameterized by industrial features), simulate policy interventions under adversarial assumptions, and translate actionable recommendations into platform governance and civic resilience programs.

References (selected anchors cited inline): [1][2][4][5][6][7][8][9][11][12][15]

Assumptions Ledger

Assumption	Rationale	Observable	Trigger	Fallback/Delegation	Scope
Industrialization created the infrastructural, organizational, and technological preconditions that enable deliberate, high-throughput, and persistent targeting of human cognition (i.e., cognitive wars are made possible by industrial forms).	Historical and contemporary evidence shows mass-printing, broadcast networks, mass schooling, bureaucratic organization and commercial PR scaled communication, standardized audiences, and enabled sustained, repeatable campaigns—conditions analogous to industrial production logic (division of labor, routinization, scale).	Presence of dedicated influence units, sustained budgets and bureaucratic processes for messaging; reuse of message templates; measurable increases in reach and frequency tied to large-scale infrastructure (press, broadcast, platforms); archival/process evidence of coordination (plans, schedules, A/B logs).	Investigation of an influence campaign that appears sustained, routinized, or organizationally complex; detection of campaign scale, repeatability, or explicit organizational artifacts (e.g., content factories, editorial calendars).	If industrial preconditions are absent, treat campaigns as decentralized/ad hoc: focus analysis on networked grassroots dynamics, emergent influencer markets, illicit ad-hoc tactics; delegate countermeasures to local actors, community organizers, or targeted digital-forensics teams rather than large institutional reforms.	Applies to state and sizable non-state actors operating in contexts with mass communication infrastructures and organizational capacity; less applicable in pre-industrial, highly fragmented, or extremely low-connectivity environments where industrial-scale routinization is infeasible.
Generative AI and algorithmic recommender/amplification systems materially lower marginal cost and raise throughput of producing and distributing persuasive content, thereby amplifying cognitive-war capacities.	Generative models automate content creation (text, image, audio, video); recommender systems increase likelihood content attains attention; documented automation of bots, microtargeting and synthetic media show cost and time barriers falling for influence operators.	Rapid increases in synthetic or near-synthetic content volume; large numbers of similar variants across accounts/platforms; accelerated A/B cycles and ephemeral testing signatures; platform telemetry showing algorithmic boosts; forensic traces of model-generated artifacts.	Surges in high-volume, low-variation content; detection of rapid, repeated message variants; platform-surface anomalies consistent with automated production or algorithmic amplification; intelligence/reports claiming use of generative tools.	If AI/amplification do not substantially lower costs (e.g., due to compute limits, defensive measures, or poor model fidelity), revert to human-driven or hybrid influence strategies; prioritize detection, verification, provenance/authentication systems and legal/regulatory measures; delegate technical mitigation to platform engineers and AI safety teams.	Valid where adversaries can access generative models and operate on platforms with algorithmic amplification and non-stringent moderation—mainly moderate-to-high digital-penetration societies; less applicable where compute/access is restricted or platform governance blocks synthetic content.

Assumption	Rationale	Observable	Trigger	Fallback/Delegation	Scope
Existing institutions (media organizations, educational systems, bureaucratic channels) provide channels, legitimation, and audience legibility that attackers can co-opt or exploit to conduct cognitive wars.	Institutions create shared priors, trusted signals, and routinized distribution channels; historically states and actors have used schools, public broadcasters, and press to shape opinion—institutions confer reach and credibility that magnify influence operations.	Coordinated messaging appearing in institutional channels; alignment between official communications and covert campaigns; evidence of content seeding via trusted institutions (curriculum changes, editorial alignment, sponsored content); shifts in trust metrics correlating with institutional messaging.	When messaging appears in or references institutional sources, when campaigns target legitimacy signals (e.g., endorsements, curricula), or when institutions change their outputs in alignment with suspected influence actors.	If institutions are weak, distrusted, or unavailable, attackers may shift to fringe networks, influencers, or covert personas—defenders should instead strengthen alternative trusted intermediaries, community-led verification, and localized resilience programs; delegate institutional-strengthening to policy-makers and civil-society actors.	Applies in contexts where institutions maintain reach and perceived legitimacy (centralized media ecosystems, standardized schooling); less applicable in highly fragmented, distrustful, or pluralized publics where institutional signals have low purchase.
Standardized mass education and repetitive mass-media exposure create sufficiently predictable cognitive priors, heuristics, and attention patterns that can be exploited by industrialized influence techniques.	Mass schooling produces common knowledge frameworks and heuristic shortcuts; repeated, standardized media exposures create entrenched attention patterns and cultural references—these regularities make audiences legible and segmentable for persuasion and microtargeting.	Homogeneity in survey responses across cohorts exposed to the same curricula/media; successful targeting using education or media-consumption indicators; predictable response patterns in experiments or A/B tests tied to schooling or media habits.	Design of targeting strategies that rely on demographic/educational segmentation, detection of consistent behavioral responses across cohorts, or observed success of narrative frames predicated on shared cultural referents.	If priors are heterogeneous or fragmented (e.g., plural media diets, decentralized education), influence must become highly localized and culturally tailored; delegate content localization and ethnographic research to local experts, linguists, and community mediators.	Most applicable in societies with centralized curricula and mass media consumption; weaker in multilingual, highly individualized, diaspora, or post-traditional media environments where priors and heuristics vary widely.
A theory-first, mechanistic (process-tracing) approach can generate falsifiable propositions and observable indicators linking industrial properties to population-level cognitive outcomes.	Mechanistic explanations map component parts and causal sequences, enabling testable claims (e.g., diffusion	Derivable, time-sequenced hypotheses that can be operationalized (e.g., measurable lags between infrastructure activation and	Research design, case selection, or policy evaluation requiring causal claims about how industrial features produced observed cognitive effects; when interventions must be	If mechanisms are not empirically tractable or falsifiable in practice, adopt mixed-methods strategies: richer qualitative case studies, probabilistic/statistical models, or iterative theory	Applies to scholarly research and policy evaluation aiming for middle-range theory; does

Assumption	Rationale	Observable	Trigger	Fallback/Delegation	Scope
	speed, institutional legibility, laundering pathways) and guiding empirical measurement—this improves explanatory power beyond ad hoc description.	opinion shifts); reproducible process-tracing evidence across cases; successful prediction or empirical rejection of proposed mechanisms.	justified by causal mechanism rather than correlation.	refinement with empirical teams; delegate measurement and fieldwork to empirical researchers and data-collection partners.	not guarantee immediate operational predictive power in highly emergent or adversarially adaptive environments without iterative empirical work.

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) Industrialization produced the infrastructural, organizational, and technological preconditions that permit deliberate, high-throughput, persistent targeting of human cognition (i.e., 'cognitive wars' are an outcome of industrialization).	[6] [7] [9] [12]	Comparative historical institutional analysis + process tracing of cases (archival and doctrinal sources) combined with complementary simulations that model scaling effects of infrastructure and organization on information throughput.	E cited; M pending comparative case studies and scaling simulations	If false, the central theoretical framing collapses — policies and predictions that target industrial/organizational sources (e.g., platform regulation, professionalized influence actors) may miss the true causal drivers of large-scale influence.	T1
(Primary) The logic of industrial production (standardization, division of labor, routinization) migrated into persuasion practices (message templates, A/B testing, content factories, microtargeting), lowering marginal cost per influence and enabling high-volume operations.	[5] [7] [11] [8]	Empirical measurement of production pipelines (case studies of content factories, industry interviews), cost-per-message accounting, lab and field A/B testing experiments, and agent-based simulations of scaled campaign deployment.	E cited; M pending targeted field experiments and cost modeling	If incorrect, resource and countermeasure strategies that assume low marginal cost and factory-style production may be misdirected; presumed economies of scale in influence would need re-assessment.	T2
(Primary) Algorithmic amplification (recommender systems, engagement optimization) amplifies salience and accelerates diffusion of influence operations, making platformized ecosystems key force multipliers in cognitive wars.	[6] [7] [14] [10]	Observational causal inference using platform telemetry (where available), natural experiments (policy or algorithm changes), and agent-based simulations of recommender dynamics to estimate amplification effects on salience, spread, and belief updating.	E cited; M pending platform access and simulation validation	If false, platform-centric mitigation (de-ranking, recommender redesign) may yield limited returns; attention should shift to alternative channels or offline institutional factors.	T3
(Secondary) Mass schooling and centralized curricula increased population informational legibility and created predictable priors and repertoires that make standardized narratives more effective at scale.	[1] [7] [12]	Cross-national statistical analysis linking schooling/centralization metrics to measures of susceptibility (survey experiments, historical opinion data), supplemented by historical case studies and lab experiments testing narrative uptake across cohorts with differing curricular exposure.	E cited; M pending cross-national regressions and experimental replication	If wrong, interventions premised on manipulating shared curricular priors (e.g., curriculum reform, educational resilience) may be less effective; vulnerability may instead be driven by other social processes.	T4
(Secondary) Generative AI (LLMs, multimodal generators) drastically scales content production and plausibility of synthetic content, increasing throughput and temporal persistence of disinformation	[12] [8] [6]	Technical benchmarks measuring generation throughput and quality, controlled deception/Turing-style studies comparing human vs. synthetic persuasion efficacy, and field	E cited; M pending benchmark experiments and controlled deception trials	If overstated, defensive investments in AI-centric detection and mitigation may crowd out other necessary measures; if understated, underpreparation risks rapid, large-scale misuse.	T5

Claim (C)	Evidence (E)	Method (M)	Status	Risk	TestID
and enabling more believable multimodal propaganda.		experiments measuring spread and uptake of generated content vs. human content.			
(Secondary) Social bots and botnets, as automated/industrialized actors, change diffusion dynamics and can be more challenging for defenders than strategic (single) attackers because of scale, automation, and mimicry.	[9] [11] [5] [15]	Network-level experiments and counterfactual simulations of bot deployment; empirical forensic analysis of known botnets; red-team exercises comparing defensive detection difficulty against automated vs. strategic attackers.	E cited; M pending red-team evaluations and detection performance benchmarks	If incorrect, attribution and mitigation strategies focused on automated networks may be misprioritized; human-centered adversaries might pose different, underappreciated threats.	T6

Sources

[1]
In 'crisis' we trust? On (un) intentional knowledge distortion and the exigency of terminological clarity in academic and political discourses on Russia's war against ...
Link.Springer.Com, 2023-01-01. (cred: 0.50)
https://link.springer.com/article/10.1057/s41268-023-00313-2

[2]
Consensus of multi-agent networks in the presence of adversaries using only local information
Dl.Acm.Org, 2012-01-01. (cred: 0.50)
https://dl.acm.org/doi/abs/10.1145/2185505.2185507

[4]
Artificial intelligence aided electronic warfare systems-recent trends and evolving applications
Ieeexplore.Ieee.Org, 2020-01-01. (cred: 0.50)
https://ieeexplore.ieee.org/abstract/document/9292960/

[5]
A Cyber-War Between Bots: Cognitive Attackers are More Challenging for Defenders than Strategic Attackers
Dl.Acm.Org, 2025-01-01. (cred: 0.50)
https://dl.acm.org/doi/abs/10.1145/3712672

[6]
Computational analysis of Information Disorder in Cognitive Warfare
Sciencedirect.Com, 2025-01-01. (cred: 0.50)
https://www.sciencedirect.com/science/article/pii/S2468696425000230

[7]
Automated influence and the challenge of cognitive security
Dl.Acm.Org, 2020-01-01. (cred: 0.50)
https://dl.acm.org/doi/abs/10.1145/3384217.3385615

[8]
Propaganda to Hate: A Multimodal Analysis of Arabic Memes with Multi-Agent LLMs
Arxiv.Org, 2024-09-11. (cred: 0.50)
http://arxiv.org/abs/2409.07246v2

[9]
Dissecting a Social Botnet: Growth, Content and Influence in Twitter
Arxiv.Org, 2016-04-13. (cred: 0.50)
http://arxiv.org/abs/1604.03627v1

[10]
The role of online attention in the supply of disinformation in Wikipedia
Arxiv.Org, 2023-02-16. (cred: 0.50)
http://arxiv.org/abs/2302.08576v1

[11]
The Rise of Social Bots
Arxiv.Org, 2014-07-19. (cred: 0.50)
http://arxiv.org/abs/1407.5225v4

[12]
The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation
Arxiv.Org, 2018-02-20. (cred: 0.50)
http://arxiv.org/abs/1802.07228v2

[14]
Phase Field Modeling in Social Media Dynamics:Simulation of Opinion Evolution with Feedback, Separation
Arxiv.Org, 2023-11-06. (cred: 0.50)
http://arxiv.org/abs/2311.03137v2

[15]
ORFEL: efficient detection of defamation or illegitimate promotion in online recommendation
Arxiv.Org, 2015-05-25. (cred: 0.50)
http://arxiv.org/abs/1505.06747v6

Generated: 2025-11-11T11:58:14.914569 | Word Count: 4932

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.