Supplementary Materials

for the article: A Conceptual Framework for a Morphological Scenario Library and Playbook Mapping in Cognitive Warfare Defense

Dojin Ryu

Department of Hacking & Security, Far East University, Eumseong-gun 27601, Republic of Korea

Email: 2022080@kdu.ac.kr

Table S1. Literature search sources, full query strings, eligibility criteria, screening counts, and screening procedure.

S1A. Search sources and coverage

Source	Type	Coverage/Filter	Search date
Scopus	Bibliographic database	2018–2025; English; articles/reviews	20 December 2025
Web of Science Core Collection	Bibliographic database	2018–2025; English; articles/reviews	20 December 2025
IEEE Xplore	Digital library	2018–2025; conferences/journals	20 December 2025
ACM Digital Library	Digital library	2018–2025; conferences/journals	20 December 2025
Institutional reports (e.g., NATO, RAND, GCSP)	Grey literature	2018–2025; English; public reports	20 December 2025

S1B. Full query strings (Boolean search strings used in this study)

Query ID	Query string (Boolean)	Notes
Q1	("cognitive warfare" OR "cognitive domain operations" OR "information operations" OR "influence operations") AND (AI OR "artificial intelligence" OR "machine learning" OR deepfake* OR "synthetic media")	Core concept + AI
Q2	("disinformation" OR propaganda OR "psychological operations" OR psyops) AND (deepfake* OR "voice cloning" OR "synthetic audio" OR "synthetic video")	Synthetic media focus
Q3	("coordinated inauthentic behavior" OR "information manipulation" OR "online influence campaign") AND (detection OR mitigation OR	Detection/defence focus

	countermeasure* OR defence OR defense)	
Q4	("Telegram" OR "encrypted messaging" OR "private messaging") AND (extremis* OR radicali* OR propaganda) AND (network OR ecosystem OR channel*)	Closed ecosystem focus
Q5	("risk scoring" OR "early warning" OR "threat prioritization") AND (information operations OR influence operations OR disinformation)	Risk/early warning focus

S1C. Eligibility criteria

Criterion	Include	Exclude
Language	English	Non-English (except foundational method sources)
Period	2018–2025 for domain literature; any year for GMA/CCA methods	Outside period for domain literature
Document type	Peer-reviewed articles; public institutional reports	Blog posts; opinion pieces; non-public/internal documents
Topic scope	Cognitive warfare / influence operations / disinformation AND AI-enabled tactics or countermeasures	Purely technical cyber intrusions not linked to influence operations
Evidence utility	Defines/operationalises actors, tactics, media, targets, or objectives; or proposes detection/response approaches	No extractable constructs for taxonomy or playbooks
Accessibility	Full text available	Full text unavailable
Duplicates	Keep most complete record	Remove duplicates

S1D. Screening procedure and counts

Stage	Records (n)	Notes
Records identified (databases)	782	Sum across databases before deduplication
Records identified (reports/grey literature)	41	Public institutional reports
Duplicates removed	163	Deduplication across all sources
Records after duplicates removed	660	Unique records taken forward
Title/abstract screened	660	Screened against eligibility criteria
Records excluded (title/abstract)	536	Out of scope; insufficient relevance; duplicates discovered during screening
Full-text assessed for eligibility	124	Full-text retrieved and evaluated
Records excluded (full-text)	78	Out-of-scope (n=34); insufficient construct detail

		(n=27); inaccessible methods/data (n=17)
Studies included in qualitative synthesis	46	Used for taxonomy + indicator/playbook extraction

Figure S1. PRISMA-style flow diagram for study selection.

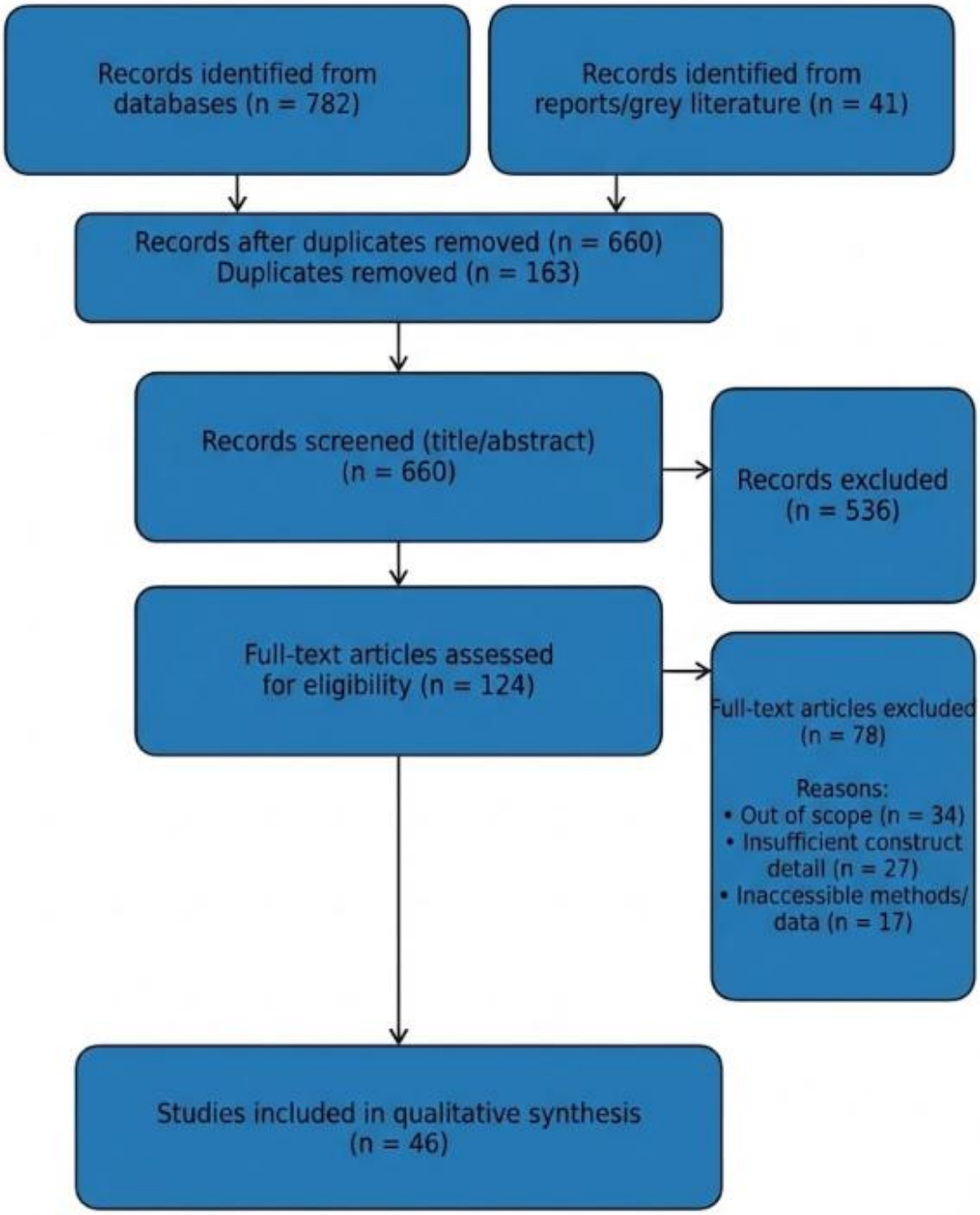


Table S2. Cross-consistency assessment (CCA) blocked state-pairs and pruning summary.

S2A. Blocked state-pairs (M = 7) and rationale

Pair ID	Dimension 1 (state)	Dimension 2 (state)	Rationale (brief)	Anchor ref.
P1	Target: Leadership elites	Objective: Polarisation	Polarisation is operationalised as population-level fragmentation across broad audiences; elite-targeted operations are better captured as decision manipulation or erosion of trust in this taxonomy.	9;10
P2	Target: Military personnel	Objective: Polarisation	Polarisation is defined as a mass-audience objective; when the primary target is military personnel, dominant objectives are more plausibly demoralisation or decision manipulation.	9;10
P3	Target: General public	Objective: Demoralisation	Demoralisation is operationalised as reduced morale/cohesion particularly in security-relevant communities; for general-public targeting, dominant objectives are polarisation or erosion of trust.	9;10
P4	Tactic: Social engineering	Medium: Traditional media	Social engineering is interaction-driven by definition; traditional media is institutionally gatekept one-to-many distribution with limited interactive affordances.	7;8
P5	Tactic: Social engineering	Objective: Polarisation	Social engineering seeks to induce specific actions/disclosures; as a dominant objective this aligns with decision manipulation rather than population-level polarisation.	7;10
P6	Medium: Private messaging	Objective: Polarisation	Closed messaging ecosystems primarily support community-bounded influence	8;10

			loops; population-level polarisation requires mass-reach channels in this simplified state space.	
P7	Actor: AI-assisted operator groups (attribution-unresolved)	Medium: Traditional media	AI-assisted operator groups are defined by automation and platform-scale content generation/distribution; traditional media gatekeeping makes it structurally implausible as a dominant medium.	6;8

S2B. Pruning statistics (as reported in Sections 2.3 and 3.2)

S2B. Pruning statistics (overlap-aware enumeration; to be used consistently in Sections 2.3 and 3.2)

Unfiltered configurations (|Ω| = 4^5) 1024

Dimensions (D) 5

States per dimension (|Sd|) 4

Blocked state-pairs (M) 7

Pruned configurations (|∪ Vi|) 331

Remaining configurations after pruning (N = |Ω| − |∪ Vi|) 693

Quantity Value

S2C. Derivation of pruned configurations with overlap accounting and sensitivity analysis

Let Ω be the full configuration space with D = 5 dimensions and |Sd| = 4 states per dimension, so |Ω| = 4^5 = 1024. Each blocked state-pair i defines a violation set Vi ⊂ Ω that contains all configurations simultaneously matching the two blocked states. For a single blocked pair, |Vi| = 4^(D−2) = 64, because two dimensions are fixed and the remaining D−2 dimensions are free.

Because blocked pairs can overlap, the total number of pruned configurations is the union size |∪ V_i| rather than the simple sum Σ|Vi|. Therefore, the union size is computed |∪ Vi| by exhaustive enumeration over Ω and evaluating membership in each Vi. Under the blocked pairs listed in S2A (M = 7), the overlap-aware union size is |∪ Vi| = 331, yielding N = 1024 − 331 = 693 remaining configurations.

Overlap diagnostics. Among the 331 pruned configurations, 235 violate exactly one blocked pair, 77 violate exactly two blocked pairs, 17 violate exactly three blocked pairs, and 2 violate four blocked pairs. This shows that overlap is non-trivial and must be accounted for when reporting pruning totals.

Sensitivity analysis (single-pair removal). To assess robustness, one blocked pair is removed at a time, the remaining library size N is recomputed, and the marginal number of restored configurations ΔN is recorded. Results are:

Pair ID Isolated |Vi| Marginal restored ΔN when pair i is removed

P1 64 33

P2 64 33

P3 64 57

P4 64 33

P5 64 16

P6 64 24

P7 64 39

The marginal restored counts are smaller than 64 because some configurations are simultaneously ruled out by multiple blocked pairs. In our experiments, removing any single pair changes N by at most 57 configurations, and the conceptual archetype coverage remains stable because the pruning rules target implausible dominance combinations rather than eliminating entire regions of the state space.

Table S3. Case instantiation evidence sheets, coding template, and test–retest agreement.

S3A. Fixed case-coding template (used for evidence sheets)

Field	Type	Description
Case ID	Text	C1–C3
Time window	Date range	Observation period reported in the public source
Actor (dominant)	Categorical	One of: State actors; Non-state groups; Insiders; AI-assisted operator groups (attribution-unresolved)
Tactic (dominant)	Categorical	One of: Disinformation; Propaganda; Social engineering; Deepfakes and synthetic media
Medium (dominant)	Categorical	One of: Social media; Traditional media; Private messaging; In-person and offline dissemination

Target (dominant)	Categorical	One of: General public; Military personnel; Leadership elites; Specific demographic communities
Objective (dominant)	Categorical	One of: Polarisation; Erosion of trust; Demoralisation; Decision manipulation
Indicators (by family)	Structured text	Content-centric; network/coordination-centric; media-forensics/channel-metadata; distribution/infra as applicable
Notes/annotations	Free text	Secondary attributes that inform indicator profiling and playbook selection

S3B. Case evidence sheets (auditable descriptors extracted from primary public sources)

Case	Public anchor	Time window	Dominant indicator families	Key auditable quantities	Scenario vector (Actor, Tactic, Medium, Target, Objective)	Notes/annotations
C1	Eady et al. (Nat. Commun., 2023) [11]	Apr 2016–Nov 2016 (exposure measured; pre-election period)	Content-centric; Network/coordination-centric	Potentially exposed U.S. users: ~32M; timeline deliveries: ~288M (≈2.5 months); concentration: top 1% ≈ 70% of exposures (reported in [11]).	(State actors; Disinformation; Social media; General public; Polarisation)	Secondary: exposure skewed toward strong partisans (reported in [11]).
C2	Clifford & Powell (GWU, 2019) [12]	Jun 1, 2017–Oct 24, 2018 (data window in report)	Network/coordination-centric; Distribution/infra-centric	Ecosystem size: 636 channels/groups; URLs: 46,579; unique base domains: 731 (reported in [12]).	(Non-state groups; Propaganda; Private messaging; Specific demographic communities; Decision manipulation)	Secondary: join-link dissemination and link-based migration mechanics described in [12].
C3	FCC News Release (23 May 2024) [13]	Jan 21, 2024 (robocalls; two days pre-primary)	Media-forensics/channel-metadata-centric; Incident-response-centric	FCC proposed fine: $6M; technique: AI voice cloning + caller-ID spoofing; target: New	(Non-state groups; Deepfakes and synthetic media; In-person and offline	Actor coded as non-state based on enforcement context; adjust only if your evidence log resolves

				Hampshire voters (reported in [13]).	dissemination ; General public; Decision manipulation)	attribution differently.

S3C. Test–retest agreement summary (two-pass coding; three cases; five dimensions)

Dimension	Agreement	Disagreement notes
Actor	66.7% (2/3)	Disagreement only for C3 attribution category
Tactic	100% (3/3)	—
Medium	66.7% (2/3)	Disagreement only for C3 channel coding
Target	100% (3/3)	—
Objective	100% (3/3)	—
Overall (all labels)	86.7% (13/15)	2 disagreements total (both in C3)

S3D. C3 resolution log (dimension-level reconciliation)

Dimension	Pass 1 label	Pass 2 label	Resolution rule applied	Final label
Actor	AI-assisted operator groups (attribution-unresolved)	Non-state groups	Use AI-assisted operator groups only when reliable attribution to a state actor, non-state group, or insider is not established and AI-driven automation is the defining observable actor attribute; otherwise code actor by attribution and capture AI via the Tactic state and/or annotations.	Non-state groups
Medium	Social media	In-person and offline dissemination	Robocall/telephony dissemination is coded under In-person and offline dissemination per the Medium operational definition.	In-person and offline dissemination

Table S4: Risk-score governance template, rubric levels, and deterministic scoring templates (S4A–S4C).

Supplementary Table S4A. Weight determination and governance template for policy-set weights (w_k)

Step	Purpose	Input	Output artifact	Audit trace
1	Define mission context and risk tolerance	Mission type, protected assets, legal constraints, escalation policy	Mission profile and scoring scope statement	Signed mission brief; version ID
2	Elicit initial weights via structured expert process	Expert panel roster; elicitation protocol (e.g., Delphi or MCDM)	Candidate weight vector w^(0) with rationales	Meeting minutes; anonymized rating sheets
3	Normalize and validate constraints	Candidate weights w^(0)	Validated weights w with Σw_k=1 and w_k≥0; sensitivity notes	Computation log; checklist
4	Approval and change control	Validated weights w; governance review	Approved weights w* and approval record	Approval memo; approver identity
5	Versioning and lifecycle management	w*; change requests; incident learnings	Versioned weight table with effective date and change log	Repository history; diff log

Supplementary Table S4B. Rubric levels and default deterministic mapping to [0,1] for risk-score components (c_k)

Component	Very low (0.00)	Low (0.25)	Moderate (0.50)	High (0.75)	Very high (1.00)	Evidence requirement
c1 Scenario-match confidence	Poorly supported match; conflicting cues	Weak match; limited corroboration	Plausible match; partial corroboration	Strong match; consistent multi-signal support	Very strong match; multi-signal support with high	Provide evidence pointers supporting each selected dominant state

					calibration confidence	
c2 Reach and velocity	Minimal dissemination; slow growth	Limited dissemination; modest growth	Sustained spread; moderate growth rate	Large spread; rapid growth and cross-community diffusion	Explosive spread; high velocity relative to channel baseline	Report channel baseline reference and normalization rule used
c3 Coordination or automation intensity	No coordination indicators	Sparse coordination indicators	Multiple coordination indicators with moderate strength	Strong coordination signatures and/or automation cues	Highly coordinated, multi-account or multi-node automation pattern	List observable coordination cues (templates, timing, graph patterns)
c4 Content-harm potential	Benign or low-harm content	Potentially misleading but limited harm	Materially misleading or coercive content	High harm (impersonation, targeted manipulation, incitement risk)	Severe harm (e.g., election interference, safety-critical deception)	Map to a stated harm taxonomy; include justification snippet
c5 Target sensitivity	Low sensitivity population and timing	General audience; low consequence context	Moderate sensitivity or mild critical timing	High sensitivity (protected groups, leadership, or crisis timing)	Very high sensitivity (elections, military ops, emergency events)	Specify why the target/timing is policy-sensitive for the mission

Supplementary Table S4C. Example deterministic scoring templates for c2 and c3 under bounded observability

Component	Observable inputs	Computation template	Mapping to [0,1]
c2 Reach and velocity	Post/mention volume V in window; time-to-peak T; channel baseline distributions	Compute $x_V = \log(1+V)$; compute $x_T = 1 - \min(T/T_{max},1)$. Combine as $x = \alpha x_V + (1-\alpha) x_T$ (α set by policy).	Normalize x to [0,1] using channel-specific min/max or percentile mapping; document baseline period.

c3 Coordination or automation intensity	Template similarity score S; synchronization score Y; graph centralization G; observable automation cues A (binary or count)	Compute raw $z = \beta 1\ S + \beta 2\ Y + \beta 3\ G + \beta 4\ f(A)$ with β weights set by policy; f(A) is monotone.	Apply monotone transform c3 = min(max(z,0),1) or logistic transform; publish thresholds for Low/Moderate/High.

Worked numerical example (illustrative) for Case C3

The values below are illustrative (not empirical estimates) and are provided solely to demonstrate how the analyst-configurable weights and component scores produce a composite prioritisation score R(s,e) in [0,1].

Component	Weight w_k	Illustrative c_k	w_k × c_k	Notes (Case C3)
Scenario-match confidence (c1)	0.25	0.80	0.20	Robocall attributed and tactic known; partial taxonomy observed
Reach/velocity (c2)	0.20	0.55	0.11	Regional phone list; rapid distribution but bounded geography
Coordination/automation intensity (c3)	0.20	0.70	0.14	Automated dialling + synthetic voice generation
Content-harm potential (c4)	0.20	0.75	0.15	Election integrity risk; impersonation
Target sensitivity (c5)	0.15	0.65	0.10	General public during primary; elevated but not military target
Composite score R(s,e)	1.00		0.70	$R = \Sigma\ w_k\ c_k = 0.70$ (illustrative)