

Cognitive Security for Critical Infrastructure: A DHS-CISA Strategy for Countering AI-Enabled Social Engineering Threats

Sanjoy Mukhopadhyay

Noblis-ESI Inc., Chantilly, VA, USA

Email: Sanjoy.Mukhopadhyay@Noblis-ESI.com

How to cite this paper: Mukhopadhyay, S. (2026). Cognitive Security for Critical Infrastructure: A DHS-CISA Strategy for Countering AI-Enabled Social Engineering Threats. *Open Journal of Social Sciences*, 14, 67-80.

<https://doi.org/10.4236/jss.2026.1410005>

Received: June 3, 2026

Accepted: October 7, 2026

Published: October 10, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The rapid advancement of artificial intelligence (AI) is transforming the security landscape of critical infrastructure by introducing new forms of cognitive risk that target human decision-making rather than technical systems alone. While cybersecurity frameworks have historically focused on protecting networks, software, and physical assets, the emergence of large language models, deepfake technologies, and AI-enabled influence operations has exposed vulnerabilities within the human element of cyber-physical systems. These technologies allow adversaries to manipulate trust, exploit cognitive biases, and influence operational decisions at unprecedented scale and sophistication. This article argues that human cognition has become a strategic attack surface requiring dedicated protection within national security and critical infrastructure frameworks. Drawing upon research in cybersecurity, cognitive psychology, socio-technical systems theory, and risk governance, the paper introduces Cognitive Security as a new paradigm for infrastructure protection. It further proposes the Cognitive Threat Assessment Standard (CTAS), a Cognitive Kill Chain model, and a Cognitive Security Operations Center (CogSOC) architecture as mechanisms for identifying, measuring, and mitigating AI-enabled social engineering threats. The article concludes that Cognitive Security should become a core component of DHS and CISA efforts to enhance national resilience in an increasingly AI-mediated information environment.

Keywords

Cognitive Security, Artificial Intelligence, Critical Infrastructure Protection, Social Engineering, Cognitive Warfare, Deepfakes, CTAS, CISA

1. Introduction

The protection of critical infrastructure has traditionally focused on safeguarding physical assets and information systems from disruption, sabotage, and unauthorized access. Over the past three decades, cybersecurity has emerged as a distinct discipline in response to the digitization of government, industry, and society. Federal agencies, including the Department of Homeland Security (DHS) and the Cybersecurity and Infrastructure Security Agency (CISA), have invested heavily in frameworks designed to strengthen cyber resilience, improve risk management, and protect critical infrastructure from increasingly sophisticated cyber threats.

Yet the emergence of artificial intelligence has revealed a critical vulnerability that lies beyond networks and software. Modern infrastructure systems depend not only on technology but also on the decisions made by operators, analysts, emergency managers, executives, and public officials. These individuals function as integral components of cyber-physical systems, translating information into operational action. As advances in generative AI accelerate, these decision-makers have become increasingly vulnerable to manipulation.

Recent developments in large language models, voice synthesis technologies, and deepfake generation have dramatically increased the effectiveness of social engineering. AI systems are now capable of producing persuasive messages tailored to individual targets, replicating trusted voices, and generating realistic visual content that can undermine confidence in authentic information (Europol, 2024). Unlike traditional cyberattacks that seek direct access to systems, these techniques seek to influence cognition itself. The objective is not merely to compromise infrastructure but to alter the judgments and behaviors of those responsible for operating it.

The implications are profound. A manipulated control-room operator can trigger operational failures without a single line of malicious code being executed. A deceived emergency manager can delay response actions during a crisis. An intelligence analyst exposed to coordinated AI-generated disinformation may unknowingly incorporate adversarial narratives into strategic assessments. In each case, the adversary targets human cognition as a pathway to operational impact.

This shift reflects a broader transformation in the threat landscape. Increasingly, security challenges involve the convergence of cyber, physical, informational, and psychological dimensions. Consequently, protecting critical infrastructure requires a more comprehensive understanding of security—one that recognizes human cognition as an essential component of national resilience.

Method and Scope. This article adopts a conceptual policy-analysis methodology that synthesizes literature from cybersecurity, cognitive psychology, human factors engineering, artificial intelligence, socio-technical systems, and critical infrastructure risk management to develop an integrated Cognitive Security framework. Rather than presenting empirical validation, the paper proposes a policy-oriented conceptual architecture intended to guide future operational implemen-

tation and research. The discussion focuses on AI-enabled social engineering, deepfake impersonation, influence operations, synthetic media, insider manipulation facilitated by AI, and decision-support attacks affecting U.S. critical infrastructure sectors. Traditional cyber intrusions, kinetic attacks, and military information operations outside civilian critical infrastructure are beyond the scope of this study.

2. From Cybersecurity to Cognitive Security

Traditional cybersecurity frameworks are largely grounded in technical assumptions. Risks are typically defined in terms of vulnerabilities within networks, devices, software applications, and communications infrastructure. While these approaches remain essential, they provide only a partial view of contemporary threats.

Research in cognitive psychology has long demonstrated that human decision-making is shaped by heuristics, biases, emotional influences, and social context (Kahneman, 2011). Adversaries have historically exploited these characteristics through deception, propaganda, and psychological operations. Artificial intelligence, however, has dramatically amplified these capabilities by enabling personalized persuasion at scale.

Recent scholarship has introduced concepts such as cognitive warfare, information manipulation, and computational influence operations to describe this evolving threat environment (NATO Strategic Communications Centre of Excellence, 2023). These approaches recognize that the objective of modern adversaries is often not the destruction of systems but the manipulation of perceptions, beliefs, and decisions.

Critical infrastructure environments are particularly vulnerable because operational decisions frequently occur under conditions of uncertainty, time pressure, and information overload. During emergencies, personnel must rapidly evaluate competing information sources while balancing safety, operational continuity, and public confidence. AI-enabled influence operations exploit these conditions precisely.

For this reason, Cognitive Security should be understood as an extension of cybersecurity rather than a replacement for it. Whereas cybersecurity protects information systems, Cognitive Security protects the human decision-making processes that govern those systems.

3. The Emergence of AI-Enabled Cognitive Threats

The rise of generative artificial intelligence has fundamentally altered the economics of social engineering. Historically, targeted influence campaigns required substantial resources, linguistic expertise, and extensive human effort. Today, many of these activities can be automated.

Large language models can generate contextually relevant communications that mirror organizational writing styles and exploit known cognitive biases. Voice-

cloning systems can reproduce speech patterns with remarkable accuracy, while synthetic media platforms can create realistic video impersonations of trusted individuals. These capabilities significantly reduce the barriers to conducting sophisticated influence operations.

Recent incidents involving deepfake-enabled financial fraud demonstrate the operational consequences of these technologies. In several documented cases, employees authorized multimillion-dollar transactions after participating in virtual meetings with synthetic representations of senior executives. Similar techniques could be directed toward critical infrastructure personnel responsible for approving operational changes, managing emergency response activities, or authorizing access to sensitive systems.

Hong Kong SAR Deepfake Fraud (2024)

In February 2024, an employee of the multinational engineering company Arup transferred approximately US\$25 million after participating in a video conference populated entirely by AI-generated deepfake representations of senior executives. The incident demonstrated that sophisticated synthetic media can successfully bypass conventional authentication procedures (Chen & Magramo, 2024).

Election Influence

AI-generated disinformation was also observed during the 2024 Slovak parliamentary election and multiple elections during 2024-2025, where synthetic audio and video were used to influence public opinion and undermine confidence in democratic institutions (OECD, 2024; World Economic Forum, 2025).

Critical Infrastructure Example

Similar techniques could target electric-grid control centers or water-treatment facilities, where spoofed emergency instructions, executive impersonation, or manipulated maintenance directives could influence operational decisions without compromising industrial control systems directly.

The threat extends beyond individual incidents. Nation-state actors increasingly employ information operations designed to influence public trust, distort perceptions of risk, and undermine confidence in democratic institutions. As AI-generated content becomes more difficult to distinguish from authentic communications, the challenge of maintaining decision integrity will continue to grow.

Consequently, the central question confronting DHS and CISA is no longer whether cognitive attacks will occur, but how organizations can systematically identify, assess, and mitigate them.

4. Cognitive Security as a Socio-Technical Framework

This article defines Cognitive Security as the protection of human decision-making processes from adversarial manipulation through integrated technical, behavioral, organizational, and policy safeguards.

The concept is grounded in socio-technical systems theory, which views organizational performance as the product of interactions among people, technology, processes, and institutions. From this perspective, human cognition represents a

critical component of infrastructure systems rather than an external factor.

A Cognitive Security framework therefore seeks to preserve decision integrity, information authenticity, and operational trustworthiness. Its objective is not to eliminate human judgment but to strengthen resilience against attempts to distort or exploit that judgment.

Within critical infrastructure environments, this requires the integration of behavioral analytics, synthetic media detection, human-centered risk assessment, and governance mechanisms capable of identifying emerging cognitive threats before they produce operational consequences.

Distinguishing Cognitive Security

Cognitive Security differs from traditional insider-threat programs, which primarily seek to identify malicious or negligent employees after indicators of compromise have emerged. It also differs from conventional security-awareness training, which emphasizes individual education rather than continuous organizational monitoring and risk assessment. Although related to the emerging literature on cognitive warfare, the present framework focuses specifically on civilian critical infrastructure protection rather than military influence operations. The Cognitive Threat Assessment Standard (CTAS) provides a structured methodology for measuring cognitive risk, while the proposed Cognitive Security Operations Center (CogSOC) operationalizes that assessment through continuous monitoring, threat analysis, and decision-support functions. Together, CTAS and CogSOC establish an enterprise governance capability rather than a training or counterintelligence program.

5. The Cognitive Threat Assessment Standard (CTAS)

A central challenge in addressing AI-enabled social engineering threats is the absence of a standardized methodology for assessing cognitive risk. Existing cybersecurity frameworks provide mature approaches for evaluating vulnerabilities in software, networks, and information systems; however, they offer limited guidance for evaluating threats directed at human cognition. This gap is increasingly significant because modern adversaries often achieve operational objectives not through technical compromise alone but through the manipulation of trust, judgment, and decision-making.

The Cognitive Threat Assessment Standard (CTAS) is proposed as a framework for systematically identifying, measuring, and mitigating cognitive risks within critical infrastructure environments. CTAS draws upon principles from enterprise risk management, behavioral science, intelligence analysis, and cybersecurity governance to create a repeatable approach for evaluating threats that target human decision-makers.

The framework is based on five interconnected dimensions: intent, capability, exposure, susceptibility, and consequence. Intent refers to the objectives and motivations of an adversary. Capability reflects the sophistication of technologies and techniques available to conduct manipulation, including large language models,

synthetic media generation, behavioral profiling, and automated influence operations. Exposure captures the degree to which personnel, communication channels, and organizational processes are accessible to adversarial influence. Susceptibility examines the behavioral and organizational factors that make manipulation more likely, including authority structures, workload pressures, trust relationships, and cognitive biases. Finally, consequence evaluates the operational, economic, safety, and national security implications that may result from successful manipulation.

CTAS Maturity Model

The CTAS Maturity Model is a staged organizational capability model that measures the maturity of Cognitive Security implementation from initial awareness to fully integrated enterprise governance.

Cognitive Security Integration Office

The Cognitive Security Integration Office is proposed as the executive governance body responsible for enterprise Cognitive Security policy, risk oversight, and interagency coordination.

Cognitive Firewall

A Cognitive Firewall refers to organizational, procedural, and AI-assisted controls that verify high-consequence communications before human decision-makers act upon them.

Then later references in Sections 8 and 9 no longer appear without definition.

Together these dimensions provide a comprehensive assessment of cognitive risk. Rather than treating human error as an isolated event, CTAS recognizes that human behavior is influenced by organizational, technological, and environmental conditions. This perspective aligns with contemporary socio-technical systems theory, which emphasizes that failures often emerge from interactions among people, technology, and institutions rather than from isolated individual actions.

To facilitate practical implementation, CTAS introduces a Cognitive Risk Index (CRI), a quantitative metric designed to support decision-making and resource prioritization. The index allows organizations to compare risks across business units, operational environments, and infrastructure sectors while providing leadership with a measurable representation of cognitive vulnerability. Similar to traditional cybersecurity risk metrics, the CRI serves as a decision-support tool rather than a predictive model, helping organizations allocate resources toward areas of greatest concern (**Table 1**).

Table 1. Example CTAS cognitive risk index.

Dimension	Score (1 - 5)	Weight
Intent	5	0.20
Capability	5	0.25
Exposure	4	0.20
Susceptibility	3	0.15
Consequence	5	0.20

The Cognitive Risk Index (CRI) may be computed as a weighted average:

$$\text{CRI} = \Sigma (\text{Weight} \times \text{Score})$$

In this example,

$$\text{CRI} = (5 \times 0.20) + (5 \times 0.25) + (4 \times 0.20) + (3 \times 0.15) + (5 \times 0.20) = 4.5/5.0$$

indicating a **Very High Cognitive Risk** requiring immediate mitigation.

This makes CTAS operational.

An important innovation within CTAS is the concept of Cognitive Attack Indicators (CAIs). These indicators function similarly to cybersecurity Indicators of Compromise but focus on signals associated with manipulation and influence operations. Examples include abnormal linguistic patterns in communications, synthetic media impersonation attempts, unusual requests that bypass established procedures, coordinated narrative campaigns targeting operational personnel, and behavioral anomalies suggesting coercion or insider influence. By integrating CAIs into existing security operations, organizations can begin to monitor cognitive threats with the same rigor applied to cyber threats.

Cognitive Attack Indicators would be detected using a combination of behavioral analytics, synthetic-media detection algorithms, communication authentication, anomaly detection, and analyst review. Suspected indicators would first be validated through correlation with threat intelligence, identity verification, communication metadata, and established operational procedures before escalation to the Cognitive Security Operations Center (CogSOC). Only validated indicators that exceed predefined CTAS thresholds would trigger incident response activities. Routine operational anomalies—including workload fluctuations, procedural deviations during emergencies, or benign communication irregularities—would be distinguished from cognitive manipulation through multi-source corroboration, historical behavioral baselines, and contextual analyst assessment, thereby minimizing false positives.

The broader significance of CTAS lies in its ability to establish a common language for cognitive security. Just as cybersecurity matured through the development of standardized frameworks and assessment methodologies, Cognitive Security requires shared concepts, metrics, and governance structures. CTAS represents an initial step toward creating such a foundation.

6. The Cognitive Kill Chain: Understanding Adversarial Manipulation

Effective defense requires a detailed understanding of how adversaries conduct cognitive attacks. While traditional cyber kill chain models describe the progression of technical attacks against information systems, they do not adequately capture the dynamics of influence, deception, and manipulation. To address this limitation, this article proposes a Cognitive Kill Chain (CKC) model designed specifically for AI-enabled social engineering operations.

The Cognitive Kill Chain begins with reconnaissance. During this stage, adver-

saries collect information about organizational structures, communication practices, decision authorities, social networks, and individual behavioral characteristics. Generative AI has dramatically enhanced this phase by enabling rapid aggregation and analysis of publicly available information. Social media activity, professional networking platforms, public documents, and leaked datasets can all contribute to highly detailed profiles of potential targets.

Following reconnaissance, adversaries conduct target profiling. The objective is to identify individuals occupying decision-critical roles and assess factors that may influence susceptibility. These factors may include authority relationships, organizational stressors, communication habits, or operational responsibilities. AI-driven analytics allow adversaries to personalize influence strategies with unprecedented precision.

The third stage involves engagement and trust establishment. Rather than immediately attempting manipulation, sophisticated adversaries often seek to develop credibility and familiarity. Synthetic personas, impersonated executives, trusted communication channels, and carefully crafted narratives are used to establish confidence and reduce skepticism. This stage is particularly dangerous because it exploits fundamental social mechanisms that organizations depend upon for effective collaboration.

Once trust has been established, adversaries initiate cognitive manipulation. Here, psychological principles such as authority, urgency, reciprocity, fear, and social proof are leveraged to shape perception and influence behavior. Artificial intelligence enhances this process by enabling continuous adaptation based on target responses. Communications can be dynamically modified to maximize persuasive effectiveness, creating influence campaigns that are far more sophisticated than traditional phishing attacks.

The next stage, decision exploitation, occurs when the target performs an action that benefits the adversary. This action may involve approving a financial transaction, modifying operational procedures, granting access privileges, sharing sensitive information, or delaying response activities during an emergency. Importantly, the manipulated action often appears legitimate from the perspective of the target.

Operational impact follows when the manipulated decision produces measurable consequences. These consequences may include cyber intrusions, infrastructure disruptions, economic losses, reputational damage, or diminished public trust. In critical infrastructure environments, even a single manipulated decision can generate cascading effects across interconnected systems.

Finally, adversaries may seek persistence and amplification. Rather than treating manipulation as a one-time event, they may establish ongoing influence relationships, recruit insiders, reinforce narratives, or conduct repeated campaigns designed to maintain access and shape future decisions. This stage transforms isolated incidents into sustained cognitive operations capable of producing long-term strategic effects.

The Cognitive Kill Chain offers an important analytical tool for defenders because it identifies multiple opportunities for intervention. By understanding where manipulation occurs, organizations can deploy safeguards designed to disrupt adversarial activities before operational consequences emerge.

7. Cognitive Security Operations Centers (CogSOC)

The operationalization of Cognitive Security requires institutional capabilities capable of monitoring, analyzing, and responding to cognitive threats. Existing Security Operations Centers (SOCs) provide valuable capabilities for cybersecurity monitoring but were not designed to address influence operations, synthetic media attacks, or behavioral manipulation. As a result, this article proposes the concept of a Cognitive Security Operations Center (CogSOC).

A CogSOC extends traditional security operations into the cognitive domain. Rather than focusing exclusively on technical indicators, it integrates behavioral analytics, threat intelligence, information integrity monitoring, and decision-assurance services. Its purpose is to maintain the integrity of human decision-making within critical infrastructure environments.

At the heart of the CogSOC is a Cognitive Threat Intelligence function responsible for monitoring emerging influence techniques, AI-enabled deception technologies, and adversarial information operations. This capability would support both strategic and operational assessments while facilitating information sharing among government agencies and private-sector infrastructure operators.

The CogSOC also incorporates behavioral analytics systems capable of establishing baseline patterns of communication and decision-making. Deviations from these baselines may indicate manipulation attempts, insider influence, coercion, or other forms of cognitive attack. Advances in machine learning provide increasingly sophisticated mechanisms for identifying such anomalies without relying solely on static rules.

Another critical component involves synthetic media detection. As deepfake technologies continue to improve, organizations require mechanisms capable of authenticating voice, video, images, and digital documents. Verification systems integrated into operational workflows can provide additional assurance before high-consequence decisions are executed.

Finally, the CogSOC serves a governance function by providing leadership with visibility into cognitive risk. Through the integration of Cognitive Risk Index scores, Cognitive Attack Indicators, threat intelligence, and operational metrics, decision-makers gain a comprehensive view of the cognitive threat environment. This visibility is essential for incorporating cognitive risk into broader enterprise risk management strategies.

8. CTAS Mapping to NIST Cybersecurity Framework 2.0

To facilitate adoption and interoperability, CTAS aligns directly with the NIST Cybersecurity Framework (CSF) 2.0.

GOVERN

The Govern function establishes Cognitive Security governance, policies, oversight structures, and accountability mechanisms.

CTAS Activities:

- Cognitive Security policy development
- Risk governance frameworks
- Privacy and civil liberties oversight
- Cognitive Security Integration Office governance
- Executive risk reporting

IDENTIFY

The Identify function focuses on understanding cognitive assets, vulnerabilities, and risks.

CTAS Activities:

- Cognitive attack surface mapping
- Human risk assessments
- Decision-critical role identification
- Cognitive Risk Index baseline development
- Sector-specific risk profiling

PROTECT

The Protect function implements safeguards designed to reduce cognitive vulnerabilities.

CTAS Activities:

- Cognitive resilience training
- Deepfake awareness programs
- Decision validation procedures
- Communication authentication protocols
- Cognitive firewall deployment

DETECT

The Detect function identifies cognitive attacks and manipulation attempts.

CTAS Activities:

- Cognitive Attack Indicator monitoring
- Behavioral anomaly detection
- AI-generated content identification
- Influence campaign monitoring
- Synthetic media detection

RESPOND

The Respond function addresses active cognitive incidents.

CTAS Activities:

- Cognitive incident response procedures
- Influence operation mitigation
- Stakeholder notification processes
- Threat intelligence sharing
- Crisis communication support

RECOVER

The Recover function restores trust and operational effectiveness following a cognitive attack.

CTAS Activities:

- Post-incident analysis
- Trust restoration initiatives
- Cognitive resilience improvements
- Lessons-learned integration
- Continuous risk reassessment

This alignment allows CTAS to be incorporated into existing NIST-based cybersecurity programs without creating separate governance structures or compliance requirements.

9. Strategic Significance

Together, the Cognitive Kill Chain, CTAS Maturity Model, Cognitive Security Operations Center, and NIST CSF alignment transform Cognitive Security from an emerging concept into a practical operational framework. These additions provide DHS and CISA with the foundational elements necessary to establish a national Cognitive Security program capable of protecting critical infrastructure against AI-enabled manipulation, influence operations, and next-generation social engineering threats.

As cyber-physical systems become increasingly dependent on human-machine decision ecosystems, Cognitive Security will become an essential component of national resilience. CTAS provides the framework through which DHS and CISA can lead this transformation and establish the United States as the global leader in securing the cognitive dimension of critical infrastructure protection.

Limitations and Governance

The framework presented in this article is conceptual and has not yet undergone empirical validation across operational environments. Consequently, CTAS metrics, Cognitive Risk Index thresholds, and Cognitive Attack Indicators should be considered initial policy constructs requiring refinement through simulation, red-team exercises, and operational testing.

Cognitive Security initiatives must also operate within established legal, ethical, and privacy boundaries. Monitoring employee communications or behavioral patterns should remain narrowly focused on security-relevant indicators, employ data-minimization principles, and comply with applicable privacy, civil liberties, labor, and constitutional protections. Cognitive Security is intended to protect decision integrity rather than monitor personal beliefs or lawful expression.

Successful implementation of a Cognitive Security Operations Center further requires executive sponsorship, interdisciplinary staffing, clearly defined governance authorities, legal oversight, integration with existing cybersecurity and enterprise risk management programs, and transparent accountability mechanisms

to ensure organizational trust.

10. Future Research Directions

The emergence of Cognitive Security as a distinct field of study presents significant opportunities for interdisciplinary research. While cybersecurity research has traditionally focused on technical vulnerabilities and system resilience, the growing importance of human cognition within cyber-physical environments requires broader analytical frameworks that integrate insights from psychology, sociology, organizational behavior, artificial intelligence, and public policy. As AI-enabled influence operations become increasingly sophisticated, understanding how individuals, organizations, and societies respond to cognitive threats will be critical to future infrastructure protection efforts.

One important area for future research involves the development of empirically validated models for measuring cognitive risk. Although the Cognitive Threat Assessment Standard (CTAS) proposed in this article provides a conceptual framework for assessing threats, additional research is required to establish reliable indicators of susceptibility, exposure, and resilience. Scholars must determine how cognitive vulnerabilities can be quantified across diverse operational environments and how those measurements correlate with real-world security outcomes. Longitudinal studies examining decision-making behavior under conditions of information manipulation would contribute significantly to this effort.

A second research priority concerns the role of artificial intelligence in both enabling and mitigating cognitive attacks. Current discussions often focus on AI as a threat vector; however, AI may also serve as a defensive capability capable of detecting manipulation, authenticating information sources, and supporting human decision-making. Future studies should examine the effectiveness of AI-assisted cognitive defense systems and evaluate how human operators interact with these technologies. Particular attention should be given to issues of trust, automation bias, and overreliance on machine-generated recommendations.

The growing prevalence of synthetic media presents another critical area for investigation. Although deepfake detection technologies continue to improve, little is known about the long-term social and organizational consequences of operating in environments where information authenticity can no longer be assumed. Researchers should explore how repeated exposure to synthetic media affects trust in institutions, emergency communications, leadership messaging, and public information systems. Such studies would provide valuable insights into the broader societal implications of AI-generated deception.

Future research should also examine the relationship between organizational culture and cognitive resilience. Existing cybersecurity studies have demonstrated that organizational factors significantly influence security outcomes. Similar dynamics are likely to affect susceptibility to cognitive attacks. Questions regarding leadership practices, communication norms, workforce training, and institutional trust warrant systematic investigation. Understanding how organizational char-

acteristics influence resilience could inform the development of more effective Cognitive Security programs.

Another promising avenue involves the study of cognitive warfare within critical infrastructure sectors. While considerable attention has been devoted to military applications of cognitive warfare, relatively little research has examined how influence operations may affect civilian infrastructure systems. Energy grids, transportation networks, healthcare systems, emergency management organizations, and financial institutions all depend on timely and accurate decision-making. Comparative studies across these sectors could identify unique vulnerabilities as well as common patterns that inform national-level policy development.

The proposed Cognitive Kill Chain model similarly requires empirical validation. Future research should examine whether cognitive attacks consistently follow identifiable stages and whether intervention strategies can reliably disrupt adversarial activities before operational impacts occur. Controlled experiments, simulation environments, and red-team exercises may provide valuable opportunities for evaluating the model's effectiveness and refining its structure.

The concept of a Cognitive Security Operations Center (CogSOC) also raises important research questions regarding governance, workforce development, and operational effectiveness. Scholars should investigate the organizational structures, skill sets, and technological capabilities necessary to support cognitive threat monitoring. Additionally, research is needed to determine how CogSOC functions can be integrated with existing Security Operations Centers and risk management frameworks without creating unnecessary complexity or duplication.

At the policy level, future studies should examine legal, ethical, and privacy considerations associated with Cognitive Security programs. Efforts to monitor behavioral indicators and detect manipulation may raise concerns regarding civil liberties, transparency, and government oversight. Research that explores governance mechanisms capable of balancing security requirements with democratic values will be essential for ensuring public trust and legitimacy.

International collaboration represents another important area for future inquiry. AI-enabled influence operations routinely transcend national boundaries, affecting governments, private-sector organizations, and societies across the globe. Comparative studies examining how different countries approach Cognitive Security, information integrity, and AI governance could help identify best practices and support the development of internationally recognized standards.

Finally, future research should focus on the development of a comprehensive science of cognitive resilience. While resilience has become a central concept in cybersecurity and infrastructure protection, relatively little attention has been devoted to understanding resilience at the cognitive level. Developing theoretical models, measurement frameworks, and intervention strategies capable of strengthening individual and organizational resistance to manipulation may ultimately become one of the most important challenges in national security research.

As artificial intelligence continues to reshape the information environment,

Cognitive Security will likely evolve into a major area of scholarly inquiry and policy development. The concepts presented in this article—including CTAS, the Cognitive Kill Chain, and the Cognitive Security Operations Center—should therefore be viewed as starting points for a broader research agenda aimed at protecting human decision-making in an increasingly complex and contested information landscape.

11. Conclusion

The future of critical infrastructure security depends not only on protecting systems and networks but also on safeguarding the decisions that control them.

AI-enabled social engineering, synthetic media, and influence operations are creating unprecedented opportunities for adversaries to exploit human cognition as a pathway to compromise. Existing cybersecurity frameworks provide important protections but do not fully address this emerging risk.

The Cognitive Security framework and the Cognitive Threat Assessment Standard (CTAS) provide DHS and CISA with a strategic path forward. By integrating cognitive risk assessment, AI-enabled detection, behavioral analytics, and resilient human-machine systems into critical infrastructure protection programs, the United States can strengthen national resilience against the next generation of threats.

Protecting cognition is no longer a theoretical concern. It is becoming a national security imperative.

Acknowledgement

The author acknowledges many leading discussions with colleagues from DHS-FEMA-Resilience group on topics related to this article. The opinion in this article does not reflect the views of the author's place of employment.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Chen, H., & Magramo, K. M. K. (2024). Finance Worker Pays Out \$25 Million after Video Call with Deepfake 'Chief Financial Officer'. *CNN World*.
<https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk>
- Europol (2024). *Facing Reality? Law Enforcement and the Challenge of Deepfakes*. Publications Office of the European Union.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.
- NATO Strategic Communications Centre of Excellence (2023). *Cognitive Warfare*.
<https://www.act.nato.int/activities/cognitive-warfare/>
- OECD (2024). *OECD Framework for the Classification of AI Systems*.
https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html
- World Economic Forum (2025) *Global Cybersecurity Outlook*.
<https://www.weforumasia.cn/publications/global-cybersecurity-outlook-2025/>