resaid.bilgi.org.tr

resaid

CREATING SOCIETAL COGNITIVE RESILIENCE AGAINST INFORMATION DISORDERS

JEAN MONNET CENTRE OF EXCELLENCE

COGNITIVE SECURITY IN THE AGE OF AI: BUILDING NATIONAL RESILIENCE AGAINST SYNTHETIC INFLUENCE

Policy Paper No 4 | 2025

Co-funded by the European Union

İstanbul Bilgi University

CREATING SOCIETAL COGNITIVE RESILIENCE AGAINST INFORMATION DISORDERS

resaid

JEAN MONNET CENTRE OF EXCELLENCE

COGNITIVE SECURITY IN THE AGE OF AI: BUILDING NATIONAL RESILIENCE AGAINST SYNTHETIC INFLUENCE

Salih Bıçakcı

Policy Paper No 4 | 2025

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by the European Union

İstanbul Bilgi University

Cognitive Security in the Age of AI: Building National Resilience Against Synthetic Influence

Salih Bıçakçı

Introduction

In the twenty-first century, power has undergone a profound transformation, encompassing a shift in its exercise, contestation, and defense. While earlier periods prioritized physical security, encompassing territory, borders, resources, and military force, the contemporary era is characterized by a novel and intangible yet equally significant domain: the human mind. Attention, trust, memory, and decision-making processes have emerged as pivotal targets of influence operations, orchestrated by hostile states, political movements, and profit-driven platforms. This new area of debate is commonly known as "cognitive security" among scholars and practitioners.

The rise of cognitive security as a concept is inextricably linked to the broader transformations associated with the Fourth Industrial Revolution. Artificial intelligence, machine learning, ubiquitous connectivity, and algorithmic mediation have reshaped the information environment, producing an unprecedented density and velocity of content. This environment is not merely a neutral platform through which information flows; it actively shapes how individuals perceive reality, filter truth from falsehood, and make collective decisions. In this sense, cognitive processes have become a new kind of critical infrastructure, no less vital to national security and democratic resilience than energy grids or financial systems.

The problem, however, is that conventional approaches to information security are ill-equipped to address the challenges of this techno-driven new era. The dominant frameworks of the past two decades—centered on cybersecurity and disinformation—have focused on either protecting technical infrastructure or ensuring the accuracy of information-based content. Cybersecurity has primarily concentrated on data integrity rather than on trustworthiness or coherence. Cybersecurity has developed to defend networks, data, and digital systems against malicious intrusion. Efforts against disinformation have primarily revolved around fact-checking, content moderation, and the policing of false claims. Both approaches are necessary, but neither is sufficient in AI-enabled manipulation.

This insufficiency stems from the way of modern threats bypass conventional defenses. Generative artificial intelligence can produce synthetic images, voices, and texts with such realism that they erode the epistemic foundation of “seeing is believing.” Deepfakes of political leaders can circulate widely before they are debunked. At the same time, AI-driven bots and chat agents can engage individuals in persuasive interactions at scale, tailoring their messages to personal vulnerabilities. At the same time, algorithmic curation by social media platforms amplifies content based not on truthfulness but on emotional resonance, often intensifying polarization and distrust. These dynamics mean that when fact-checkers have disproven a claim, the damage to public trust is already done.

This paper, therefore, argues that the focus must shift from defending information as content to defending cognition as process. Cognitive security is not about policing truth or suppressing speech; rather, it is about ensuring that human beings can exercise autonomy, judgment, and resilience in the face of manipulation and deception. To secure democratic societies in the age of artificial intelligence, cognitive security must become a central pillar of national resilience.

The implications of this argument are profound. If decision-making under uncertainty is a core function of democracy, defense, and crisis management, exploiting cognitive vulnerabilities directly threatens sovereignty and stability. A society in which citizens cannot distinguish between authentic and synthetic messages, or in which trust in institutions has collapsed, is a society unable to govern itself effectively. The erosion of cognitive security thus translates into a crisis of democratic legitimacy and strategic resilience.

Accordingly, the scope of this paper encompasses both conceptual and policy-oriented aspects. It begins by defining cognitive security and situating it within the broader landscape of security studies, highlighting how it extends beyond cybersecurity and information security to address the vulnerabilities of human cognition itself. It then examines the role of artificial intelligence as a transformative disruptor of the information environment, illustrating how synthetic media and algorithmic persuasion exacerbate the risks of manipulation. The third section examines the shortcomings of existing policy tools in NATO, the European Union, and Türkiye, identifying the key gaps that leave democracies vulnerable to threats. The fourth develops a framework for cognitive resilience, integrating technical, educational, institutional, and societal measures. Finally, the conclusion reflects on the broader implications of this paradigm shift, calling for democracies to invest in cognitive security as urgently as they do in cyber defense.

By combining conceptual analysis with policy recommendations, this paper aims to enhance the understanding of cognitive security as both a scholarly field and a strategic imperative. It draws on emerging literature in psychology, security studies, and technology policy, while situating the argument in a comparative global context with particular attention to NATO, the European Union, and Türkiye. The ambition is not only to diagnose the problem but to propose a path forward that strengthens democratic resilience in an age of synthetic influence.

1. The Emergence of Cognitive Security as a Strategic Domain

The emergence of cognitive security as a distinct domain of inquiry and practice signifies a paradigm shift in how states, organizations, and scholars perceive risk in the twenty-first century. Historically, security frameworks were centered on physical domains, such as land, sea, and air, and subsequently augmented by space and cyberspace. The advent of the digital environment and the proliferation of networked infrastructures gave rise to cybersecurity as a primary concern. However, cybersecurity has increasingly revealed its limitations: it safeguards the technical foundation of digital systems but fails to address how information, once transmitted, is processed, interpreted, and acted upon by human cognition.

Cognitive security builds upon but also departs from these earlier paradigms. It recognizes that human cognition—our capacity for judgment, trust, memory, and decision-making—has become a strategic target in its own right. As James Bone (2017) has argued in his pioneering work on "cognitive hacking," adversaries exploit technical vulnerabilities and the predictable biases and heuristics of human psychology. Similarly, Huang and Zhu (2023) frame cognitive security as a system-scientific challenge, emphasizing that the human mind is embedded in networks of information, technology, and social interaction that can be manipulated to destabilize societies.

The Fourth Industrial Revolution has accelerated this transformation. The rapid diffusion of artificial intelligence, machine learning, big data analytics, and ubiquitous connectivity has created an information environment in which cognitive vulnerabilities are magnified. The very features that enable efficiency and personalization—algorithmic curation, predictive analytics, and real-time communication—also create opportunities for exploitation. As Bone and Lee (2023) highlight in their recent work on cognitive risk, organizations and governments must reconceptualize their understanding of risk to account for these novel threats.

It is beneficial to analyze the historical evolution of information and security paradigms to comprehend the emergence of cognitive security. During the Cold War, information was acknowledged as a domain of contention, primarily in the context of propaganda and ideological competition. The focus was on message content—who could present more compelling narratives. With the advent of the internet in the 1990s, attention shifted to cybersecurity, which involves safeguarding networks, data, and digital infrastructure against unauthorized access and sabotage. In the 2000s and 2010s, the proliferation of social media and the weaponization of information flows by state and non-state actors raised concerns about "information warfare" and "disinformation campaigns."

Each paradigm captured an essential facet of the evolving landscape, but each was limited in its scope. Propaganda analysis focused on overt messaging but underestimated the role of cognitive biases. Cybersecurity focuses on technical defenses but overlooks the vulnerabilities of human interpretation. Disinformation frameworks concentrated on content, assuming that falsehood could be countered by fact-checking and content moderation. These paradigms overlooked the fact that cognition—the mental processes through which information becomes belief and action—can be targeted directly.

Cognitive security, therefore, represents an attempt to integrate insights from psychology, neuroscience, and behavioral economics into the field of security studies. Research on cognitive biases, from Kahneman and Tversky's (1974) work on heuristics to contemporary studies of motivated reasoning, has shown that human beings are not rational information processors. We rely on mental shortcuts that make us vulnerable to manipulation. Adversaries exploit confirmation bias by amplifying content that aligns with prior beliefs or availability bias by emphasizing vivid but unrepresentative events. The vulnerability lies not only in the content of information but in the architecture of cognition itself.

This recognition has profound strategic implications. It suggests that adversaries no longer need to control territory, destroy infrastructure, or even produce persuasive narratives in the traditional sense. Instead, they can destabilize societies by sowing doubt, confusion, and mistrust. As Pomerantsev (2019) has observed, the goal of modern information warfare is often not to persuade people of a particular falsehood but to erode their confidence in the possibility of truth. In this environment, the cognitive domain becomes the decisive terrain of conflict.

NATO has begun to recognize this reality. Its "NATO 2030" agenda highlights the importance of addressing disinformation and hybrid threats, and there is growing discussion of whether cognition should be conceptualized as a new "domain" of warfare alongside land, sea, air, cyber, and space. The European Union has also invested in monitoring and countering disinformation through initiatives such as the European External Action Service's East StratCom Task Force. Yet both NATO and the EU have struggled to move beyond reactive measures. Fact-checking, counter-narratives, and platform regulations are essential, but they focus on information content rather than the cognitive process.

Türkiye offers a particularly instructive case. Positioned at the intersection of Europe, the Middle East, and Eurasia, it is exposed to multiple impact ecosystems: Russian disinformation campaigns, Western narratives, regional sectarian propaganda, and domestic polarization. Therefore, the challenges of cognitive security in Türkiye are both external and internal. Externally, hostile actors exploit religious, ethnic, and geopolitical cleavages to weaken cohesion. Internally, political polarization creates fertile ground for manipulative narratives to take root. Türkiye's experience illustrates the dual nature of cognitive security: it is both an external defense issue and a matter of domestic resilience.

The academic study of cognitive security is still in its early stages, but it is growing rapidly. Huang and Zhu (2023) propose a system-scientific approach, treating cognition as part of a broader socio-technical system. They argue that cognitive security requires interdisciplinary collaboration, integrating insights from neuroscience, psychology, computer science, and political science. From a risk management perspective, Bone and Lee (2023) emphasize treating cognitive vulnerabilities as organizational risks that must be systematically identified, assessed, and mitigated. Both approaches commonly underline that cognition is a strategic asset to be protected.

This section has sought to map the emergence of cognitive security as a strategic domain. It has been argued that existing information security and cybersecurity paradigms are necessary but insufficient in the context of AI-enabled manipulation. Cognitive security extends the field by focusing on the vulnerabilities of human cognition itself. It is grounded in insights from psychology and behavioral science, and it reflects the transformations of the Fourth Industrial Revolution. The cases of NATO, the EU, and Türkiye illustrate the global relevance of this paradigm.

As we progress, the challenge will be to translate these conceptual insights into effective policy frameworks and practical strategies. Cognitive security must be institutionalized as a pillar of national resilience, comparable to cybersecurity and physical defense. However, it must also be pursued in ways that respect democratic principles and human autonomy. Therefore, the following sections of this paper examine how artificial intelligence disrupts cognitive security, where current policies fall short, and how a resilience framework can be developed to address these challenges.

2. The Dynamics of Information Disorder and Cognitive Vulnerability in the Post-Truth Era

In the contemporary information environment, disorder is increasingly described as a condition characterized by the blurred boundaries between truth and falsehood, fact and opinion, knowledge and belief. This so-called "post-truth era" does not merely imply misinformed individuals; it suggests a profound transformation in the relationship between information, cognition, and politics. The defining characteristic of post-truth is not the absence of truth but its delegitimization: truth becomes merely one narrative among many, stripped of its privileged authority. In this context, the central question for security studies shifts beyond correcting falsehoods to safeguarding the cognitive processes through which societies establish shared understandings of reality. Even slight alterations in the information presented during the decision-making process can significantly impact the decisions made, without any intervention.

The persistence of misinformation highlights the inadequacy of traditional fact-checking. Nyhan and Reifler (2010) demonstrate the "backfire effect," where corrections to false claims sometimes reinforce misperceptions instead of dispelling them. Similarly, Vosoughi, Roy, and Aral (2018) illustrate that false news spreads faster and more widely on social media than trustworthy news, primarily because it is more novel and emotionally engaging. These discoveries highlight a fundamental cognitive asymmetry: the mechanisms that render human cognition efficient, such as heuristics and emotion-driven attention, also render it susceptible to manipulation.

Cognitive asymmetry refers to the intentional disparity between the capacities of two actors to perceive, interpret, and make decisions. It manifests when one party systematically diminishes its own uncertainty and decision latency while simultaneously increasing the other party's confusion, verification expenses, and time to make a decision, thereby achieving an advantage without resorting to coercion.

In the post-truth condition, adversaries no longer need to convince audiences with a coherent falsehood. Instead, they can flood the environment with contradictory claims, erode trust in institutions, and exploit polarization. The Russian strategy of "firehose of falsehood," coined by researchers from RAND (2016), exemplifies this approach. The "firehose of falsehood" is characterized by four distinct traits: it operates at a high volume. It is rapid and continuous, utilizing multiple channels and lacking commitment to truth or internal consistency. Rather than employing a single credible narrative, it overwhelms audiences with numerous false, some partially true, and some mutually contradictory narratives disseminated through state-owned media outlets, proxy sites, social media platforms, bots, and influencers. Rather than securing a fact-check victory, the primary objective is to confuse, divert attention, and exhaust verification efforts. Consequently, the effect is not the promotion of belief in lies, but rather the erosion of belief in truth. In essence, these campaigns aim to manipulate the audience's perception of truth and confuse their cognitive processes, underscoring a shift in strategy.

This shift transforms the nature of cognitive vulnerability. It is not only about susceptibility to specific false claims but also about the erosion of epistemic authority. At the culmination of the operation lies a semantic assault that irrevocably erodes trust in fundamental pillars of democratic systems. When citizens no longer trust media, experts, or institutions, democratic processes are undermined. Elections depend on a shared acceptance of legitimate outcomes; public health relies on trust in scientific expertise; crisis management requires collective belief in authoritative information. Hence, the collapse of epistemic trust leads to political instability directly. It also affects the social solidarity of the population, and the resilience of society directly.

The appearance of artificial intelligence provided an effective platform to amplify these operations with reduced effort and enhanced persuasive capability. Generative AI technologies, such as deepfakes, voice cloning, and large-scale text generation, increase the speed, scale, and believability of manipulative content. A deepfake video of Ukrainian President Volodymyr Zelensky urging surrender in 2022 illustrates how synthetic media can target cognitive trust in sensory perception (Smalley, 2022). Voice cloning technologies have already been employed in financial fraud and could readily be utilized for political disruption. AI-driven chatbots and persuasion engines can initiate personalized dialogues, exploiting individuals' cognitive weaknesses in real-time. Algorithmic curation ensures that such manipulations are precisely targeted, delivering emotionally resonant narratives to susceptible audiences(U.S. Federal Trade Commission, 2023, 2024; Damiani, 2019).

Ambient technologies extend these capabilities by embedding sensing and actuation in everyday environments and on the body, including smartphones and the Internet of Bodies (IoB), smart speakers and other always-listening microphones, Wi-Fi/RF sensing, and immersive extended reality (XR). At the sense stage, passive data from devices and infrastructure generate continuous signals about presence, movement, routines, and physiological proxies of stress or arousal. Systematic reviews demonstrate that smartphone- and wearable-based digital phenotyping can effectively track mobility and sleep regularity, which are associated with mood and stress (Bufano et al., 2023; Choi et al., 2024). IoB analyses highlight the policy and security implications of health-adjacent wearables and implantables feeding this data layer (Lee et al., 2020). In parallel, the IEEE 802.11bf protocol is an amendment to the Wi-Fi (802.11) standard that adds a formal, interoperable framework for Wi-Fi sensing—using ordinary Wi-Fi signals not just for plain data, but to detect presence, motion, gestures, range, and related features of objects and people in the environment. (Ropitault et al., 2023; Sahoo et al., 2024). The ubiquitous presence of smart devices in our daily lives has led to the capture of substantial amounts of data. This data serves as a valuable resource for data analysts in the field of profiling. Recent research on smart-speaker misactivations has highlighted the potential for wake-word errors to inadvertently capture ambient audio or metadata. This phenomenon expands the scope for behavioral inference, raising concerns about the privacy and security of user data (Dubois et al., 2020). In operational terms, ambient stacks lower the cost of finding the “right moment” and the “right micro-audience,” even before any explicit persuasive message is shown.

At the deliver stage, algorithmic curation (ranking, notification scheduling, geofenced prompts) sequences content to maximize salience at moments of predicted susceptibility. At the same time, XR enhances presence and identification compared to non-immersive formats, which can lead to stronger attitude or behavior change in specific contexts (Makransky & Petersen, 2021). Many systems incorporate affective computing, which involves adapting headlines, imagery, or tone based on estimated arousal/valence derived from video, audio, text, or physiological streams. However, leading reviews caution that inferring discrete emotions from facial expressions is often unreliable across contexts (Barrett et al., 2019).

Neurotechnology adds further levers at both the sensing and modulation ends of the pipeline. For sensing/segmentation, non-invasive brain–computer interfaces (BCIs) and consumer neuro-wearables (EEG/fNIRS) can monitor coarse attentional or workload states. The increasing number of brain-computer interface (BCI) companies, such as Neuralink, Synchron, and Paradromics (Dawson, 2022), underscores the profound impact of the future on cognitive security. On the other hand, China is implementing a detailed national strategy to become a global leader in the brain-computer interface (BCI) industry by 2030.

The plan focuses on advancing core technologies and developing high-performance products, while establishing a robust industrial ecosystem and adhering to strict ethical guidelines. Its primary objective is to utilize BCI for therapeutic and medical applications, while also encouraging consumer-facing product development (Ministry of Industry and Information Technology, 2025)

At the clinical frontier, invasive speech neuroprostheses have decoded attempted speech in near real time, illustrating rapid (though medical) progress in brain-signal decoding (Moses et al., 2021; Card et al., 2024). For deliver/modulate, transcranial magnetic stimulation (TMS) is clinically authorized for specific psychiatric indications (e.g., FDA marketing authorization for obsessive–compulsive disorder). In contrast, transcranial electrical stimulation (tDCS/tACS) exhibits small, variable, and context-dependent effects in healthy individuals, with more reliable gains typically associated with structured training regimens (U.S. Food and Drug Administration, 2018; Horvath et al., 2015; Price et al., 2015). Defense R&D has explicitly pursued nonsurgical high-performance interfaces (e.g., DARPA's N3), highlighting the dual-use trajectory of neurotech even as current noninvasive systems remain low-bandwidth and noisy (DARPA, 2019). The immediate security concern, therefore, is less about mind control and more about state-aware operations—specifically, targeting, timing, and personalization informed by neural/physiological indicators.

The optimization stage completes the cycle: multivariate testing and reinforcement of high-performing combinations iteratively refine which narrative, format, and timing are compelling for various micro-audiences. When combined with ambient sensing (IoB, Wi-Fi, smart speakers), affect proxies, and neuro-adjacent signals, optimization can produce verification-cost asymmetries, making it cheap for operators to iterate and expensive for the public and institutions to verify and counter. From a policy perspective, this argues for (i) data-protection that explicitly covers neural/affective inferences and cross-context portability, (ii) auditable provenance for state-relevant information flows and platform ranking inputs, and (iii) early participation in standards (e.g., 802.11bf) to embed privacy-by-design and rate-limiting into the infrastructure (Lee et al., 2020; OECD, 2019, 2025).

These abilities, merged with AI-enabled disinformation, distinguish them from earlier forms of propaganda. Traditional disinformation on the production of persuasive stories and the dissemination of stories through the mass media. AI allows mass customization: individuals can receive a narrative tailored to their cognitive biases, emotional triggers, and social networks. The challenge is not only that false content looks more authentic but also that it is delivered with greater relevance and resonance. The result is a new level of cognitive penetration and persuasion.

Military doctrines always acknowledge the presence of information operations (Department of the Army, 2023). The global landscape demonstrates the strategic use of these techniques. Russia and China have systematically weaponized information disorder as part of their hybrid warfare strategy, using disinformation to shape perceptions in Ukraine, Europe, and beyond. China has increasingly deployed synthetic narratives to influence diasporas, undermine democratic institutions, and promote its model of governance (Hao, 2018). Non-state actors, from extremist groups to profit-driven disinformation entrepreneurs, exploit the same tools for recruitment and impact. The privatized architecture of global communication platforms amplifies these efforts, as algorithms prioritize engagement over accuracy.

The European Union has responded with initiatives such as the Code of Practice on Disinformation and the Rapid Alert System, seeking to hold platforms accountable and coordinate responses. NATO has begun to incorporate cognitive resilience into its hybrid warfare doctrine, recognizing that adversaries exploit physical and cognitive vulnerabilities. Türkiye, meanwhile, has faced both external disinformation campaigns and internal polarization, highlighting the dual nature of the threat. During elections, synthetic narratives exploiting religious and national identities have circulated widely, testing the resilience of democratic processes. In crises like the 2023 earthquakes, disinformation spread rapidly online, complicating relief efforts and undermining trust in official communication.

These cases illustrate that cognitive vulnerability is not evenly distributed. Social, political, and institutional contexts shape it. Highly polarized societies are more susceptible to manipulative narratives, as adversaries can exploit existing divisions within them. Societies with low trust in institutions are more vulnerable to epistemic erosion. And societies with limited media literacy may struggle to recognize and resist synthetic manipulation. Cognitive security is therefore not a uniform condition but a relational one, dependent on the resilience of democratic institutions, the robustness of civil society, and the integrity of information ecosystems.

The post-truth condition also poses profound ethical dilemmas for policy responses. Efforts to protect cognitive security can easily slide into paternalism or censorship. Authoritarian regimes often justify information control in the name of security, suppressing dissent and monopolizing narratives. Democracies must avoid replicating these practices. The challenge is to defend cognition without undermining freedom. This requires transparency, accountability, and a commitment to human rights. Cognitive resilience must be built not by shielding citizens from exposure but by equipping them to navigate complexity and resist manipulation.

One promising approach is the concept of psychological inoculation, also known as "prebunking." Research by van der Linden and colleagues (2017) shows that exposing individuals to weakened forms of misinformation and explanations of manipulative techniques builds resistance to future attempts. Narrative resilience is another critical dimension: strengthening trustworthy narratives rooted in democratic values can counteract manipulative alternatives. These approaches move beyond reactive fact-checking toward proactive cognitive empowerment.

Ultimately, the dynamics of information disorder and cognitive vulnerability in the post-truth era reveal that the decisive battle is not over content but over process. It is not enough to identify and remove false claims. The deeper challenge is to preserve the cognitive conditions of autonomy, trust, and deliberation. As Rini (2020) argues, misinformation is fundamentally a crisis of trust: without confidence in epistemic authorities, citizens cannot coordinate around shared truths. Cognitive security thus requires strategies that rebuild trust, strengthen resilience, and preserve the integrity of democratic processes.

This chapter has argued that the post-truth condition represents a qualitative transformation in the nature of information disorder. By eroding epistemic authority and exploiting cognitive vulnerabilities, adversaries can destabilize societies without firing a shot. Artificial intelligence amplifies these dynamics by enabling speed, scale, and believability. Existing policy responses in NATO, the EU, and Türkiye remain focused on content, leaving the deeper vulnerabilities of cognition unaddressed. In this sense, the next chapterunderlines the question of how policies might be reframed to build cognitive resilience in an age of synthetic influence.

3. Building Cognitive Resilience: Policy Responses in a World of Synthetic Influence

The first two chapters have demonstrated that cognition is a newly emerged strategic domain, and artificial intelligence has transformed the nature of information disorder; the next step is to ask how societies can respond to this development. The challenge is not simply to detect or remove harmful content but to cultivate resilience in the face of synthetic influence. Cognitive resilience refers to the capacity of individuals and societies to withstand, adapt to, and recover from manipulative attempts to destabilize attention, trust, memory, and decision-making. It is the mental and institutional immune system of democracy, requiring a multi-layered policy framework.

In the history of security policy, resilience has often been framed in terms of physical infrastructure. Critical infrastructure protection involves redundancy, robustness, and the capacity to recover quickly from disruption. This logic can be extended into the cognitive domain. Just as energy grids require backup systems, societies require cognitive backups: credible narratives, resilient institutions, and educated citizens who can resist manipulation. Thus, the task of policyis to design measures that strengthen the cognitive immune system at multiple levels: technical, educational, institutional, and societal.

At the technical layer, the challenge is to develop tools that can detect and trace synthetic media. AI-generated content can often be identified through forensic analysis of artifacts, inconsistencies, or statistical signatures. Researchers have developed watermarking techniques that embed invisible signals into synthetic images or videos, allowing them to be authenticated. International collaboration is essential here: synthetic content circulates across borders, and detection tools must be interoperable. The European Union has already begun to explore provenance standards for digital content, and initiatives such as the Content Authenticity Initiative have brought together technology companies, media organizations, and civil society to develop shared protocols. NATO, too, has recognized the importance of rapid detection, particularly in crises where synthetic narratives can undermine military operations or alliance cohesion.

Yet detection alone is insufficient. As the previous chapter argued, the damage to trust may already be irreversible by the time a deepfake is debunked. Therefore, the technical layer must be integrated with proactive measures emphasizing cognitive resilience. One such measure is psychological inoculation, also known as prebunking. Research has shown that exposing individuals to weakened forms of misinformation and explanations of manipulative techniques builds resistance to future attempts. Google and Cambridge University have tested prebunking videos that explain common rhetorical tricks, such as scapegoating or emotional manipulation, and found that viewers subsequently became less susceptible to misinformation. National governments could build on these experiments to develop prebunking campaigns tailored to their specific cultural and political contexts.

Education forms the second critical layer of resilience. Traditional media literacy programs, which teach individuals to distinguish fact from falsehood, must evolve into what might be called “cognitive literacy.” Cognitive literacy goes beyond identifying reliable sources; it involves understanding the psychological processes of attention, bias, and persuasion. Citizens must be equipped to recognize misinformation and how their cognitive shortcuts can be exploited. Integrating cognitive literacy into school curricula, civil service training, and military education would build long-term resilience across society. or NATO and EU member states, such programs could be standardized and shared across borders, creating a collective baseline of cognitive resilience. Education fostering critical thinking and cross-cultural understanding is especially vital in Türkiye, where polarization and identity-based narratives create particular vulnerabilities.

At the institutional layer, the challenge is coordination. Cognitive security is currently fragmented across multiple domains: cybersecurity agencies focus on technical defenses, ministries of education handle literacy, health agencies address misinformation in public health, defense organizations monitor hybrid threats, and regulators oversee technology platforms. What is missing is a holistic strategy that integrates these efforts into a coherent whole. Some countries have begun establishing national centers for countering disinformation or hybrid threats; however, few have explicitly adopted cognitive security as an organizing concept. Establishing National Cognitive Security Centers could provide a focal point for coordination, bringing together intelligence agencies, educational institutions, civil society organizations, and technology companies. Such centers would monitor threats and design resilience programs, coordinate responses, and ensure that measures respect democratic values.

NATO and the EU also have roles to play at the institutional level. NATO has acknowledged the cognitive domain as an emerging area of competition but has yet to formalize strategies that operationalize this recognition. Developing a cognitive security doctrine—comparable to cyber or hybrid warfare doctrines—would provide a framework for member states to follow. For its part, the EU has advanced platform regulation through the Digital Services Act; however, it could further enhance its regulatory agenda by incorporating cognitive security considerations. Joint NATO-EU initiatives could establish cross-border rapid response mechanisms to synthetic influence campaigns, share best practices on cognitive literacy, and coordinate research on detection technologies.

The societal layer is the most challenging but also the most crucial one. Cognitive resilience cannot be imposed from above; it must be cultivated from below. Societies that are cohesive, inclusive, and characterized by high levels of trust are less vulnerable to manipulation. Conversely, societies marked by polarization, inequality, and distrust are fertile ground for synthetic narratives. Building societal resilience, therefore, requires investment in social cohesion, community-based trust networks, and transparent government communication. Independent fact-checkers, civil society organizations, and local media play vital roles in this ecosystem. Governments can support them through funding, capacity-building, and legal protections. But trust ultimately depends on performance: when institutions are perceived as corrupt, ineffective, or unresponsive, no amount of communication strategy can restore credibility.

Türkiye again provides a vivid example. During the 2023 earthquakes, disinformation spread rapidly on social media, undermining trust in official relief efforts.

Some of this misinformation was externally generated, but much of it reflected internal polarization and skepticism toward institutions. The episode demonstrates that cognitive resilience cannot be separated from broader questions of governance and legitimacy. Therefore, strengthening cognitive security in Türkiye requires technical detection tools, educational programs, and reforms that enhance institutional trust and democratic accountability.

The ethical dimension remains central throughout—efforts to protect cognitive security risk sliding into paternalism or censorship if not carefully designed and implemented. Democracies must resist the temptation to emulate authoritarian information control. Transparency is essential: citizens must understand how and why content is being moderated, how algorithms curate their feeds, and what measures are being taken to protect them. Public-private partnerships with technology companies should be structured to ensure accountability, with precise oversight mechanisms and safeguards against abuse. The goal is not to shield citizens from exposure but to empower them to navigate complexity with autonomy.

This multi-layered technical, educational, institutional, and societal framework does not offer a final solution. Cognitive security is not a problem to be solved once and for all, but rather a condition that must be managed continuously. Just as cybersecurity requires constant adaptation to evolving threats, cognitive resilience is also required. The difference is that while cybersecurity protects networks and systems, cognitive security protects the very capacity of societies to deliberate, decide, and act collectively. It is therefore not a narrow technical issue but a foundational pillar of democratic resilience.

In a comparative perspective, the need for such a framework is evident. NATO's recognition of the cognitive domain highlights its strategic significance, but operational doctrines remain underdeveloped. The EU's regulatory initiatives represent an essential step, but enforcement is uneven, and the focus remains mainly on content rather than cognition. Türkiye's experience underscores the dual nature of the challenge, with both external manipulation and internal polarization creating vulnerabilities. Other states, from the United States to Asian democracies, face similar dilemmas. The global nature of synthetic influence means that no country can address the problem alone. Cognitive security must therefore be understood as both a national responsibility and a collective good requiring international cooperation.

This chapter has argued that building cognitive resilience requires a multi-layered approach. Technical tools such as watermarking and detection are necessary but insufficient.

Educational programs must evolve into cognitive literacy, equipping citizens to resist manipulation. Institutional coordination is crucial for preventing fragmentation and ensuring the implementation of holistic strategies. Societal resilience, rooted in trust, cohesion, and transparency, is the ultimate foundation for a stable society. Balancing these measures with respect for democratic values and human autonomy is the central normative challenge. The next chapter explores the development of a broader paradigm of cognitive security, examining how it can be integrated into the architecture of twenty-first-century security.

4. Toward a Cognitive Security Paradigm for the Twenty-First Century

The previous sections have explored the emergence of cognitive security as a distinct domain, the transformation of information disorder in the post-truth era, and the necessity of multi-layered policy responses. To consolidate these insights into a durable foundation for action, a broader paradigm shift is necessary. As Thomas Kuhn (1962) reminds us, a paradigm is more than a set of theories; it is a lens through which reality is perceived and organized. In this sense, cognitive security demands a reorientation in how states and societies understand the nature of security in the twenty-first century.

Traditional security paradigms were built around material threats. Military power was measured by the size and capabilities of armies, tanks, ships, and aircraft. Cybersecurity focused on networks and infrastructures, extending this logic into the digital realm. Information security, similarly, focuses on the integrity of data and content. These paradigms were adequate for an era in which threats were predominantly physical or technical in nature. However, they are insufficient for an age in which the decisive terrain of conflict is cognitive.

The concept of a “cognitive domain” is not merely metaphorical. As Huang and Zhu (2023) argue, cognition is part of a complex socio-technical system, subject to natural vulnerabilities and deliberate manipulation. The mind is not isolated; it is embedded in communication networks, social interaction, and technological mediation. In this sense, cognitive security is not simply an extension of cybersecurity but a distinct domain that entails its own doctrines, policies, and institutions.

One way to conceptualize this paradigm shift is based on the analogy of critical infrastructure. Just as energy grids, financial systems, and transportation networks are essential to the functioning of modern societies, so too is cognitive integrity. If citizens cannot trust their perceptions, memories, or judgments, the capacity of a society to deliberate and make collective decisions collapses. In this sense, cognition itself must be treated as critical infrastructure. Protecting it requires technical defenses and cultural, educational, and institutional investments.

Another way to frame the paradigm shift is through the lens of resilience. Traditional security approaches often prioritize prevention and deterrence: stop the attack before it happens or deter adversaries from acting through the threat of retaliation. Cognitive security, by contrast, emphasizes resilience. Manipulative narratives cannot be entirely prevented; synthetic media cannot be wholly eliminated; and algorithmic persuasion cannot be entirely neutralized. The aim is not to eliminate threats, but to build the capacity to absorb them without incurring a catastrophic loss of trust or autonomy. This is a shift from a logic of control to a logic of adaptation.

The implications for international security organizations are profound. NATO has long recognized land, sea, air, cyber, and space as distinct domains of warfare. Recognizing a cognitive domain would represent a significant expansion of this framework. It would require developing doctrines for cognitive defense, training military personnel in cognitive resilience, and integrating cognitive considerations into operational planning and execution. It would also require close cooperation with civilian institutions, as cognitive security extends beyond the battlefield into the fabric of society.

The European Union faces a similar challenge. Its regulatory initiatives, such as the Digital Services Act, represent necessary steps in holding platforms accountable for the harms of algorithmic curation. However, a cognitive security paradigm would require a broader orientation: not only regulating content and platforms but also investing in the cognitive resilience of citizens, supporting independent media, and fostering epistemic trust. It would mean treating cognitive security not as a niche issue of disinformation but as a core dimension of democratic resilience.

Türkiye's position at the intersection of Europe, the Middle East, and Eurasia highlights the global nature of the challenge. Cognitive security is not merely a Western concern but a universal one. Türkiye's exposure to multiple influence ecosystems—Russian disinformation, Western narratives, regional sectarian propaganda, and internal polarization—demonstrates that cognitive threats are multifaceted and crosscutting. A cognitive security paradigm would allow Türkiye to integrate its domestic resilience strategies with its role in NATO and its engagement with the EU, positioning it as a bridge between different regions and traditions.

The paradigm shift also has ethical dimensions. As Rouvroy and Stiegler (2016) argue, the digital era has given rise to new "algorithmic governmentality," in which power is exercised not through coercion or persuasion but through the modulation of attention, behavior, and choice. Cognitive security must therefore be understood not only in strategic terms but also in normative ones. Protecting cognition means protecting autonomy, dignity, and democratic agency. The risk is that states may be tempted to adopt paternalistic or authoritarian measures that undermine these values to defend cognition. Consequently, the cognitive security paradigm must be explicitly anchored in human rights and democratic principles.

The conceptual foundations of this paradigm are still being developed. Scholars such as Bone (2017) and Bone and Lee (2023) emphasize cognitive risk management, treating cognitive vulnerabilities as organizational risks to be identified and mitigated. Huang and Zhu (2023) propose a systems-scientific approach, situating cognition within complex adaptive systems. Rouvroy and Stiegler (2016) highlight the implications of algorithmic governance for human autonomy. These perspectives, while distinct, converge on the recognition that cognition is both a strategic asset and a normative value.

Moving toward a cognitive security paradigm requires integrating these insights into policy and practice. This entails developing doctrines that view cognition as a domain of security, designing resilience frameworks that operate across technical, educational, institutional, and societal layers, and incorporating ethical safeguards that protect autonomy and democracy. It means recognizing that cognitive security is not a temporary problem of disinformation but a permanent condition of the digital age.

This chapter has argued that cognitive security represents a paradigm shift in security thinking. Traditional paradigms focus on physical and technical domains, while cognitive security focuses on the human mind as a strategic terrain. This shift requires treating cognition as critical infrastructure, emphasizing resilience over control, and integrating ethical considerations into strategic planning and decision-making. NATO, the EU, and Türkiye exemplify the urgency and global relevance of this paradigm. The following chapters will explore the future challenges and research agendas necessary to consolidate cognitive security as a durable pillar of twenty-first-century resilience.

5. Cognitive Security Futures: Challenges, Opportunities, and Research Agenda

As cognitive security increasingly becomes recognized as a defining domain of twenty-first-century conflict and governance, policymakers and scholars face the challenge of anticipating its future trajectory. The rapid pace of technological change, combined with the evolving strategies of adversaries, ensures that the challenges of today will not be identical to those of tomorrow. At the same time, new opportunities are emerging to strengthen resilience, develop international cooperation, and foster innovation in protective measures.

This chapter examines the future landscape of cognitive security, focusing on the significant challenges, the potential opportunities, and the research agenda required to navigate the age of synthetic influence.

The first challenge lies in the accelerating sophistication of artificial intelligence. Generative models are advancing at a pace that outstrips detection capabilities. Deepfakes that were once crude and easily identifiable are now increasingly seamless, able to replicate not only the visual appearance of individuals but also their voice, mannerisms, and even idiosyncratic speech patterns. Large language models are capable of generating persuasive texts that are indistinguishable from human writing. When integrated with personalization engines, they can simultaneously deliver tailored narratives to millions of individuals. The trajectory of technological development suggests that these capabilities will only improve, making synthetic influence harder to detect and more persuasive in its impact.

A second challenge is the asymmetry between offense and defense. Creating synthetic media is relatively cheap and scalable; while detecting and debunking it is resource-intensive and often slow. This asymmetry favors adversaries, who can flood the information environment with manipulative content at minimal cost. Defensive measures, by contrast, require sophisticated technologies, human expertise, and institutional coordination. The result is a structural imbalance that leaves societies perpetually vulnerable to being outpaced.

A third challenge is attribution. Identifying the source of synthetic influence campaigns is notoriously tricky. State and non-state actors can conceal their origins, route operations through multiple intermediaries, and capitalize on the anonymity of the digital environment. Without reliable attribution, deterrence becomes nearly impossible. International law, which relies on clear lines of responsibility, struggles to adapt to a world of plausible deniability and distributed manipulation. This creates a permissive environment in which adversaries can act with relative impunity.

A fourth challenge is the risk of excessive intervention. Efforts to protect cognitive security can easily slide into paternalism or censorship, especially in contexts where governments already face pressures toward authoritarianism. In attempting to defend citizens from manipulation, the danger is that states may restrict freedom of expression, suppress dissent, or monopolize narratives. This would not only undermine democratic values but also erode trust further, paradoxically weakening cognitive resilience rather than strengthening it.

Despite these challenges, the future of cognitive security also presents opportunities. The same technologies that enable synthetic influence can be harnessed for defense. Artificial intelligence can be used to detect manipulative content, trace its origins, and model its likely impact. Machine learning can identify emerging narratives before they spread widely, enabling the development of preemptive countermeasures. Natural language processing can support prebunking campaigns by generating tailored educational materials that inoculate individuals against manipulation. These opportunities require investment, but they also demonstrate that technological innovation is not inherently destabilizing; it can be harnessed for resilience and beneficial manipulation.

Another opportunity lies in international cooperation. Cognitive security is a transnational challenge: synthetic narratives transcend borders, and adversaries exploit global platforms. This creates a strong incentive for collaboration among states, international organizations, and technology companies. NATO has already taken steps to integrate cognitive resilience into its hybrid warfare strategy, and the European Union has developed mechanisms for platform regulation and cross-border alerts. Expanding these initiatives into a more comprehensive cognitive security framework could provide collective protection against synthetic influence campaigns. Türkiye, situated at the crossroads of multiple influence ecosystems, could bridge the gap in fostering dialogue between Western and non-Western approaches to cognitive security.

At the societal level, the future of cognitive security presents an opportunity to renew democratic resilience. By investing in cognitive literacy, supporting independent media, and promoting social cohesion, societies can enhance their ability to resist manipulation. The post-truth condition is not inevitable; it results from institutional failures, technological incentives, and cognitive vulnerabilities. Addressing these underlying conditions can rebuild trust and restore the mental foundations of democracy. In this sense, cognitive security is both a defensive necessity and a positive project, offering a vision of societies that are more informed, resilient, and capable of collective self-governance.

The research agenda for cognitive security is correspondingly broad. Several key areas demand attention. First, the mental dimension of security must be theorized more systematically. Huang and Zhu (2023) have begun this work by framing cognition as part of a socio-technical system. However, further research is needed to integrate insights from neuroscience, psychology, sociology, and political science. How do cognitive biases interact with algorithmic curation? How does social polarization amplify vulnerability? How does trust in institutions shape susceptibility to manipulation? These are questions that require interdisciplinary investigation.

Second, the effectiveness of resilience measures must be rigorously tested. Psychological inoculation and prebunking show promise, but their long-term effects remain unclear. Do inoculation techniques lose their effectiveness over time? Can they backfire under certain conditions? How can they be adapted to different cultural and political contexts? Large-scale field experiments, such as those conducted by the Cambridge Social Decision-Making Lab, provide a model for this research. Still, more systematic studies are needed across diverse societies.

Third, the ethical dimensions of cognitive security necessitate careful exploration. As Rouvroy and Stiegler (2016) argue, algorithmic governmentality represents a new form of power that operates by shaping attention and behavior rather than issuing commands. Protecting cognition raises questions about autonomy, consent, and the boundaries of state intervention. How can democracies defend citizens without undermining freedom of expression? What safeguards are necessary to prevent abuse of cognitive security measures? These questions must be addressed by policymakers, ethicists, legal scholars, and civil society.

Fourth, the role of technology companies must be scrutinized. Platforms such as Facebook, Twitter, and TikTok are not neutral information conduits; their algorithms shape what users see and believe. Their business models, which prioritize engagement, often amplify manipulative content. Holding them accountable is therefore essential to cognitive security. But what form should this accountability take? Regulation is one option, but it risks stifling innovation or driving platforms to jurisdictions with weaker rules. Public-private partnerships are another option, but they require trust and transparency to be effective. Research is needed to explore the most effective governance models.

Finally, the global dimension of cognitive security requires sustained attention. The dynamics of synthetic influence vary across cultural and political contexts. In liberal democracies, the challenge is to defend autonomy and trust without undermining freedom. In authoritarian regimes, the challenge is that cognitive security may be co-opted as a justification for repression. In regions such as the Middle East, where geopolitical rivalries intersect with religious and ethnic divisions, cognitive security assumes distinct forms. Comparative research is therefore essential, both to understand these variations and to develop context-sensitive strategies.

The future of cognitive security will not be determined solely by technology. The interaction of technological innovation, cognitive vulnerabilities, institutional responses, and normative choices will shape it. The challenges are formidable: accelerating AI, offensive-defensive asymmetries, attribution difficulties, and ethical dilemmas.

However, the opportunities are equally significant: harnessing technology for defense, fostering international cooperation, renewing democratic resilience, and advancing interdisciplinary research.

The stakes could not be higher. As Bone and Lee (2023) argue, cognitive risk is the new frontier of security that will define the resilience of organizations and societies in the digital age. The erosion of cognitive security poses a threat not only to political stability but also to the very possibility of democratic self-governance. Yet by investing in cognitive resilience, societies can turn vulnerability into strength. They can build the capacity to withstand manipulation, adapt to new challenges, and preserve the autonomy of their citizens.

This chapter outlines the future challenges, opportunities, and research agenda for cognitive security. It has been argued that the field must move beyond reactive measures toward proactive resilience, integrating technological innovation, interdisciplinary research, international cooperation, and ethical safeguards. The next chapter, the conclusion, reflects on the broader implications of this argument, emphasizing the urgency of embedding cognitive security at the core of national resilience strategies in the age of synthetic influence.

6. Conclusion – Securing the Cognitive Domain in the Age of AI

The preceding chapters have traced the conceptual emergence of cognitive security, the transformation of information disorder in the post-truth era, the disruptive potential of artificial intelligence, and the contours of a multi-layered resilience framework. Together, they reveal a profound shift in security: cognition, our attention, trust, memory, and decision-making have become a strategic domain. This recognition carries with it both a warning and a call to action. Without robust policies and practices to defend the cognitive domain, democracies risk entering an era of synthetic influence in which truth, trust, and self-governance are systematically eroded.

The first core argument advanced in this paper is that cognitive security must be understood as distinct from, though complementary to, traditional notions of information security and cybersecurity. Information security focuses on ensuring content integrity, which means data is accurate, protected, and reliable. Cybersecurity emphasizes the defense of technical infrastructures against intrusion, sabotage, and disruption. Both are necessary, but neither addresses the vulnerabilities of cognition itself. Human beings are not rational information processors; we rely on heuristics, emotions, and social cues that adversaries can manipulate. As Kahneman and Tversky (1974) demonstrated decades ago, cognitive biases are an inherent feature of human reasoning.

In the digital environment of the Fourth Industrial Revolution, these biases have become exploitable attack surfaces. The rise of algorithmically mediated communication has rendered the mind a contested domain, necessitating explicit protection.

The second core argument is that artificial intelligence has transformed the nature of disinformation from a problem of content to a problem of cognition. Earlier forms of propaganda sought to persuade audiences of particular narratives. Contemporary AI-enabled manipulation, by contrast, often aims to erode the possibility of truth itself. Deepfakes blur the boundary between reality and fabrication; voice cloning undermines trust in sensory perception; large language models can flood the information environment with persuasive but misleading text at scale. As Pomerantsev (2019) has argued, many modern influence campaigns aim not to replace truth with falsehood but to create a state of epistemic chaos in which citizens no longer know what to believe. This shift from persuasion to confusion, from narrative control to narrative overload, represents a qualitative transformation like information disorder.

The third core argument is that current policy tools are inadequate to meet these challenges. Fact-checking and content moderation, while valuable, are fundamentally reactive in nature. They address the content of misinformation after it has already spread, but they do not address the cognitive vulnerabilities that make individuals susceptible in the first place. Legal frameworks often lag behind technological innovation, with limited regulation of synthetic media and inadequate accountability mechanisms on platforms. Institutional responsibilities are fragmented across various sectors, resulting in gaps in coordination and effectiveness. Democracies remain vulnerable not only to external manipulation by hostile states but also to internal polarization, which is amplified by AI-driven narratives.

The fourth core argument is that true resilience requires defending minds, not just networks. Cognitive resilience can be understood as a form of immunity: the capacity to recognize, resist, and recover from manipulative attempts. This requires a multi-layered framework. Watermarking, provenance standards, and detection tools are essential at the technical layer. At the educational level, cognitive literacy programs must equip citizens with the skills to recognize manipulation and resist their biases. At the institutional layer, national cognitive security centers should coordinate efforts across intelligence agencies, educational systems, regulators, and civil society. Investment in trust, transparency, and cohesion is essential at the societal layer.

From a comparative perspective, NATO, the EU, and Türkiye illustrate both the challenges and opportunities of building such resilience. NATO has begun to recognize the cognitive domain as an emerging area of competition; however, it has yet to integrate cognitive security into its operational doctrines fully. The European Union has taken steps to regulate platforms through the Digital Services Act and monitor disinformation through the East StratCom Task Force. Still, these efforts remain primarily focused on content. Türkiye, positioned at the crossroads of Europe, the Middle East, and Eurasia, faces a unique combination of external manipulation and internal polarization, underscoring the need for a holistic cognitive security strategy that integrates technical, educational, institutional, and societal measures.

Looking to the future, several challenges loom large. The accelerating sophistication of AI ensures that synthetic influence will become more challenging to detect and more persuasive in its impact. The asymmetry between offense and defense favors adversaries, who can generate manipulative content at scale and low cost. Attribution remains difficult, complicating deterrence and accountability. The ethical risks of extensive intervention are significant: in seeking to protect cognitive security, democracies risk undermining the very freedoms they aim to defend.

Yet the future also presents opportunities. The same technologies that enable manipulation can be harnessed for defense. AI can be used to detect synthetic media, model narrative spread, and generate prebunking interventions. Through organizations such as NATO and the EU, international cooperation can provide collective protection against transnational influence campaigns. Investment in education, media literacy, and civic resilience can rebuild trust and restore the cognitive foundations of democracy.

The research agenda for cognitive security is therefore urgent and interdisciplinary in nature. It requires collaboration across various fields, including psychology, neuroscience, computer science, law, and political science. It demands rigorous testing of resilience measures, including the long-term effectiveness of prebunking and inoculation strategies. It calls for critical engagement with the ethical dilemmas of cognitive security, ensuring that protective measures do not undermine human rights and democratic principles. It also requires comparative research across regions, recognizing that cognitive vulnerabilities and strategies vary depending on cultural, political, and institutional contexts.

In many ways, cognitive security can be seen as the logical culmination of the trajectory that has been traced since the early twentieth century. Where once security was conceived in material terms—borders, armies, and weapons—it has now expanded to encompass the infrastructures of information and cognition.

This expansion reflects the realities of the Fourth Industrial Revolution, in which the lines between physical, digital, and cognitive domains are increasingly blurred. Cognitive security is thus not an optional add-on but an essential pillar of resilience in the contemporary world.

For democracies, the imperative is particularly acute. Democratic governance depends on the capacity of citizens to make informed decisions, deliberate collectively, and trust in institutions. If these cognitive foundations are undermined, democracy itself is at risk. The age of synthetic influence poses a direct challenge to these foundations. Deepfakes, voice cloning, algorithmic persuasion, and synthetic narratives are not simply nuisances; they are strategic weapons designed to erode trust, sow confusion, and destabilize societies.

The call to action is therefore clear. Democracies must invest in cognitive security as urgently as they invest in cyber defense or physical protection. This requires developing and funding national cognitive security strategies, embedding cognitive literacy into education, establishing institutional mechanisms for coordination, and fostering societal resilience. It requires international cooperation, particularly within NATO and the EU, to address transnational threats and share best practices. It also requires a commitment to ethical principles, ensuring that efforts to protect cognition do not come at the cost of freedom and autonomy.

If these steps are not taken, the consequences could be severe. Societies may enter an era of synthetic influence in which truth becomes indistinguishable from falsehood, trust in institutions collapses, and democratic processes are hollowed out. In such an environment, adversaries can achieve their objectives without firing a shot, while democracies are paralyzed by uncertainty and division. The erosion of cognitive security is thus not a peripheral concern but a central threat to the survival of democratic self-governance in the digital age.

At the same time, there is reason for cautious optimism. By recognizing cognition as a strategic domain, societies can build the resilience necessary to navigate the challenges of the Fourth Industrial Revolution. Cognitive security offers a defensive necessity and a positive vision: a vision of societies in which citizens are empowered to navigate complexity, resist manipulation, and participate meaningfully in democratic life. It is a vision in which technology serves human autonomy rather than undermining it, and international cooperation fosters trust rather than eroding it.

In conclusion, the age of artificial intelligence has transformed the security landscape. The decisive terrain of conflict is no longer only the battlefield or the network but the human mind. Cognitive security must be recognized as a central pillar of national and international resilience to meet this challenge. The task is urgent, but it is not insurmountable. By investing in cognitive resilience, democracies can defend their infrastructures and capacity for collective self-governance. The future of democracy depends on our ability to secure the cognitive domain.

BIBLIOGRAPHY

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest,* 20(1), 1–68. https://doi.org/10.1177/1529100619832930

Bone, J. (2017). *Cognitive Hack: The New Battleground in Cybersecurity—the Human Mind.* CRC Press.

Bone, J., & Lee, J. H. (2023). *Cognitive Risk.* CRC Press.

Bufano, P., Rumi, G., & Terlizzi, S. (2023). Digital phenotyping for monitoring mental disorders: A systematic review. *Diagnostics,* 13(21), 3296. https://doi.org/10.3390/diagnostics13213296

Cambridge Social Decision-Making Lab. (2022). *Prebunking interventions in the field: Evidence from randomized experiments.* University of Cambridge.

Card, N. S., Moses, D. A., Chartarifsky-Lewis, R., et al. (2024). An accurate and rapidly calibrating speech neuroprosthesis. *Nature Communications,* 15, 5372. https://doi.org/10.1038/s41467-024-43292-2

Choi, A., Ooi, A., & Lottridge, D. (2024). Digital phenotyping for stress, anxiety, and mild depression: Systematic literature review. *JMIR mHealth and uHealth,* 12, e40689. https://doi.org/10.2196/40689

Damiani, J. (2019, September 3). A voice deepfake was used to scam a CEO out of $243,000. *Forbes.* https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/

DARPA. (2019, May 20). *Six paths to the nonsurgical future of brain–machine interfaces (N3 awards).* https://www.darpa.mil/research/programs/next-generation-nonsurgical-neurotechnology

Dawson, R. (2022, November 8). *Leading Brain-Computer Interface Companies*. RossDawson.com.https://rossdawson.com/futurist/companies-creating-future/leading-brain-computer-interface-companies-bci/

Department of the Army. (2023, November 27). *Army Doctrine Publication 3-13, Information*. Army Publishing Directorate. Retrieved from https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN39736-ADP_3-13-000-WEB-1.pdf

Dubois, D. J., Kolcun, R., Mandalari, A. M., Paracha, M. T., Choffnes, D. R., & Haddadi, H. (2020). When speakers are all ears: Characterizing misactivations of IoT smart speakers. *Proceedings on Privacy Enhancing Technologies,* 2020(4), 255–275. https://doi.org/10.2478/popets-2020-0069

European Commission. (2018). *Action Plan Against Disinformation.* European Union.

European Commission. (2022). *The Digital Services Act.* European Union.

Hao, J. (2018, March 25). *China's "Three Warfares" in Theory and Practice in the South China Sea.* Georgetown Security Studies Review. Retrieved from https://georgetownsecuritystudiesreview.org/2018/03/25/chinas-three-warfares-in-theory-and-practice-in-the-south-china-sea/

Horvath, J. C., Forte, J. D., & Carter, O. (2015). Quantitative review finds no evidence of cognitive effects in healthy populations from single-session tDCS. *Brain Stimulation,* 8(3), 535–550. https://doi.org/10.1016/j.brs.2015.01.400

Huang, L., & Zhu, Q. (2023). *Cognitive Security: A System-Scientific Approach.* Springer.
Kahneman, D., & Tversky, A. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124–1131.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* University of Chicago Press.

Lee, M., Posard, M. N., Bond, C. A., & Miller, L. L. (2020). *The Internet of Bodies: Opportunities, risks, and governance* (RR-3226). RAND Corporation. https://www.rand.org/pubs/research_reports/RR3226.html

Makransky, G., & Petersen, G. B. (2021). The cognitive and motivational effects of immersive virtual reality in education: A systematic review and meta-analysis. *Computers & Education,* 157, 103. https://doi.org/10.1016/j.compedu.2020.103010

Ministry of Industry and Information Technology, National Development and Reform Commission, (2025). *Implementation opinions of seven departments on promoting the innovative development of the brain-computer interface industry.* Chinese Government Website. Retrieved fromhttps://www.gov.cn/zhengce/zhengceku/202508/content_7035603.htm

Moses, D. A., Leonard, M. K., Makin, J. G., & Chang, E. F. (2021). Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *New England Journal of Medicine,* 385(3), 217–227. https://doi.org/10.1056/NEJMoa2027540

NATO. (2021). *NATO 2030: United for a New Era.* NATO Public Diplomacy Division.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior,* 32(2), 303–330.

OECD. (2019). *Recommendation on Responsible Innovation in Neurotechnology.* Organisation for Economic Co-operation and Development. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0457

OECD. (2025). *Neurotechnology policy toolkit.* Organisation for Economic Co-operation and Development. https://www.oecd.org/sti/emerging-tech/neurotech-toolkit.pdf

Pomerantsev, P. (2019). *This Is Not Propaganda: Adventures in the War Against Reality.* PublicAffairs.

Price, A. R., McAdams, H., Grossman, M., & Hamilton, R. H. (2015). A meta-analysis of tDCS effects on language measures. *Cortex,* 73, 251–261. https://doi.org/10.1016/j.cortex.2015.09.005

RAND Corporation. (2016). *The Russian "firehose of falsehood" propaganda model.* RAND Research Reports.

Rini, R. (2020). *Deepfakes and the epistemic backstop.* Philosophy & Technology, 33(4), 1–21.

Rouvroy, A., & Stiegler, B. (2016). The digital regime of truth: From the algorithmic governmentality to a new rule of law. *La Deleuziana: Online Journal of Philosophy,* 3, 6–29.

Ropitault, T., et al. (2023). IEEE 802.11bf: Enabling the widespread adoption of Wi-Fi sensing. *NIST Publications.* https://www.nist.gov/publications/ieee-80211bf-enabling-widespread-adoption-wi-fi-sensing

Sahoo, A., et al. (2024). Sensing performance of the IEEE 802.11bf protocol and its overhead. *NIST Technical Note.* https://doi.org/10.1109/VTC2024-Fall63153.2024.10757977

Smalley, S. (2022). *Zelenskyy deepfake crude, but still might be a harbinger of dangers ahead.* CyberScoop. https://cyberscoop.com/zelenskyy-deepfake-troubles-experts/

U.S. Federal Trade Commission. (2023, March 20). *Scammers use AI to enhance their family emergency schemes.* https://www.consumer.ftc.gov/consumer-alerts

U.S. Federal Trade Commission. (2024, April 8). *Family emergency scams.* https://www.consumer.ftc.gov/consumer-alerts

U.S. Food and Drug Administration. (2018, August 17). *FDA permits marketing of transcranial magnetic stimulation for treatment of obsessive–compulsive disorder.* https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-transcranial-magnetic-stimulation-treatment-obsessive-compulsive-disorder

Van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges,* 1(2), 1600008.

Van der Linden, S., Roozenbeek, J., & Compton, J. (2020). Inoculating against fake news about COVID-19. *Frontiers in Psychology,* 11, 566790.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science,* 359(6380), 1146–1151.

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making.* Council of Europe.

CREATING SOCIETAL COGNITIVE RESILIENCE AGAINST INFORMATION DISORDERS

resaid

JEAN MONNET CENTRE OF EXCELLENCE

This policy paper has been prepared as part of the Creating Societal Cognitive Resilience Against Information Disorders (RESAID) Project supported by the European Commission Jean Monnet Centres of Excellence.

Co-funded by the European Union