E.ISSN: 2661-7404 Academic Journal of Legal and Political Researchs P.ISSN:2571-9971
Year: 2026 Vol : 10 No : 02 Page : 150 - 168



# From Digital Soft War to Knowledge Crisis: Frameworks for Digital Governance in the Age of A.I

# من الحرب الناعمة الرقمية إلى أزمة المعرفة
# أطر الحوكمة الرقمية في عصر الذكاء الاصطناعي

**Abdelghani Hadjab***
Mohamed Boudiaf University of M'sila (Algeria)
abdelghani.hadjab@univ-msila.dz

**Slimane Melouki**
Mohamed Boudiaf University of M'sila (Algeria)
slimane.melouki@univ-msila.dz

**Received: 23- 06- 2026** **Accepted: 08- 08- 2026** **Published: 22- 09- 2026**

**Abstract:**

The contemporary security environment is undergoing a fundamental transformation as conflict transitions from the physical domain to the digital information ecosystem. Automated digital agents have emerged as primary instruments in the framework of digital soft war, designed to mimic human behavior, manipulate public sentiment, and distort the perceived consensus on critical issues. This research explores the multifaceted impact of these agents on societal security, emphasizing the mechanisms of agenda-building, the amplification of polarization, and the deployment of synthetic media. Analysis of empirical data suggests that automated networks effectively target cognitive vulnerabilities, leading to the erosion of institutional trust and the fragmentation of national identity. The study identifies the emergence of the "liar's dividend" and the decay of epistemic security as critical threats to democratic stability. Findings indicate that while technical detection remains a challenge, enhancing societal resilience through digital literacy and integrated policy frameworks is the most viable path for safeguarding the digital public sphere.

**Keywords:** Automated Accounts; Societal Security; Public Opinion; Information Integrity; Digital Conflict.

**ملخص**

يشهد المشهد الأمني المعاصر تحولا جذريا مع انتقال الصراع من المجال المادي إلى النظام البيئي للمعلومات الرقمية. وقد برزت الوكلاء الرقميون الآليون كأدوات رئيسية في إطار الحرب الناعمة الرقمية، مصممة لمحاكاة السلوك البشري، والتأثير على الرأي العام، وتشويه الإجماع السائد حول القضايا الحاسمة. يستكشف هذا البحث التأثير متعدد الأوجه للوكلاء على الأمن المجتمعي، مع التركيز على آليات تضخيم الاستقطاب ونشر الوسائط الاصطناعية. ويشير تحليل البيانات التجريبية إلى أن الشبكات الآلية تستهدف بفعالية نقاط الضعف المعرفية، مما يؤدي إلى تآكل الثقة المؤسسية وتفتيت الهوية الوطنية. وتحدد الدراسة تدهور الأمن المعرفي كتهديدات خطيرة للاستقرار الديمقراطي. تشير النتائج إلى أنه على الرغم من أن

* **Corresponding author**

E.ISSN: 2661-7404 Academic Journal of Legal and Political Researchs P.ISSN:2571-9971
Year: 2026 Vol : 10 No : 02 Page : 150 - 168



الكشف التقني لا يزال يمثل تحديا، فإن تعزيز قدرة المجتمع على الصمود من خلال الثقافة الرقمية وأطر السياسات المتكاملة هو المسار الأمثل لحماية الفضاء الرقمي العام.

**الكلمات المفتاحية:** الحسابات الآلية، الأمن المجتمعي، الرأي العام، سلامة المعلومات، الصراع الرقمي.

## Introduction:

The evolution of digital technology has fundamentally altered the architecture of human interaction, shifting the nexus of social and political discourse into the virtual realm. This transition has birthed new forms of conflict where the objective is not the physical destruction of an adversary, but the subversion of their cognitive and social foundations. Within this digital soft war, the deployment of automated digital agents represents a sophisticated strategy to influence the collective consciousness of nations. These software-driven entities operate with a scale and speed that bypass traditional informational gatekeepers, allowing for the systematic engineering of disinformation and the manipulation of public opinion.

## Research Significance

The significance of this inquiry is rooted in the critical nature of Societal Security, which serves as the bedrock of national stability. When the informational environment is polluted by artificial narratives, the ability of a society to maintain its essential characteristics and identity is compromised. The primary goal of this research is to analyze the structural and behavioral patterns of automated agents, evaluate their impact on public discourse, and identify effective mechanisms for enhancing societal resilience.

## The Central Problem

The central problem addressed in this research focuses on how automated digital agents and the tools of digital soft war affect the societal security of states and what mechanisms can be developed to protect communities from these pervasive threats. This inquiry is guided by several questions:

- How do the interaction structures of automated agents differ from human users?
- In what ways do these agents influence the public agenda and Social Polarization?
- What are the specific threats posed by the integration of Artificial Intelligence and Deepfake Technology into these campaigns?

## Theoretical Framework

The research is built upon two primary hypotheses:

- **First:** The strategic use of automated agents leads to a significant increase in negative sentiment and social polarization by manipulating the saliency of

controversial attributes in public debate.

- **Second:** The proliferation of synthetic media within automated networks creates a state of "Epistemic Decay," where the public loses the capacity to distinguish authentic information from fabricated narratives, thereby undermining the legitimacy of democratic institutions.

**Methodology and Structural Organization of the Study**

To address these hypotheses, the study adopts a multidisciplinary methodology:

- **Descriptive Approach:** To categorize the various personas and behavioral signatures of automated agents.
- **Comparative Method:** To distinguish between human and bot-led interaction networks.
- **Case-Study Approach:** To examine specific instances of digital interference in regional conflicts and health crises.

The structure of this study is organized into several chapters, beginning with the conceptual framework of Digital Soft War and Societal Security, followed by an examination of the mechanisms of automated influence, the impact on public opinion, the threats posed by synthetic media, and strategies for building Societal Resilience.

## 1. THEORETICAL AND CONCEPTUAL FRAMEWORKS

### 1.1 Conceptual Foundations of Digital Soft War and Societal Security

The concept of power has been fundamentally reshaped by the digital revolution. In the post-Cold War era, the primary currencies of power -culture, political ideals, and policies- are increasingly mediated through social networks. This transition has led to the emergence of "digital soft war," where the openness of information societies is exploited to sow internal discord and weaken national cohesion. Societal security, a concept pioneered by the Copenhagen School, refers to the ability of a society to preserve its essential characteristics under changing conditions and threats[1]. Unlike traditional national security, which focuses on the territorial integrity of the state, societal security centers on the protection of identity, trust, and social bonds. In the digital age, these elements are under constant assault from automated agents that manipulate the information landscape.

The architecture of these threats is increasingly complex. Automated agents are not merely isolated scripts; they are cyber-sociotechnical agents that interact with human psychology and platform algorithms to generate perceived realities. These agents are often deployed as part of computational propaganda, an assemblage of social media platforms, autonomous agents, and big data tasked

with the manipulation of public opinion[2]. The vulnerability of modern societies to these threats is highlighted by the high level of concern reported by citizens. Empirical studies indicate that over 82% of participants in certain demographic assessments identify substantial security risks associated with digital platform interactions. These risks extend beyond simple misinformation to include the erosion of cultural integrity and the introduction of undesirable behaviors within traditional family and social structures[3].

**Table 1.** Levels of Participant Concern and Primary Threat Vectors Impacting Societal Security

| **Dimension of Impact** | **Reported Participant Concern (%)** | **Primary Threat Vector** |
|---|---|---|
| General Security | 82.8% | Cybercrimes and platform-facilitated threats |
| Ethical and Behavioral | 84.4% | Spread of immoral content and value erosion |
| Social Connections | 81.8% | Weakening of real-world bonds and family structures |
| Intellectual Security | Variable | Spread of extremist ideas and religious hostility |

**Source**: Alnaqbi, 2025, Article 1508542.

The table above illustrates the depth of the challenge as perceived by employees in a high-digitalization environment, reflecting a consensus that digital interactions directly threaten the core pillars of societal stability. Security is a cornerstone of stability, and societies plagued by widespread digital crime or misinformation may face deterioration, underdevelopment, and long-term instability[4].

### 1.2 Sociotechnical Systems and the Future of Digital Society

Societal security in digital societies must be understood from a sociotechnical perspective, where physical and social artifacts are mutually dependent and co-evolving. The digital transformation of society can be virtuous if managed correctly, but it also increases vulnerability and concerns about trust and security. A coherent framework is needed to understand the relationship between these concepts and to ensure stability in digital societies[5].

The resilience of society depends on how well organizations and individuals can adapt to evolving circumstances. This requires a measured approach to

adjustment, rather than reactive changes, to fortify resistance against disinformation and influence campaigns. Secure digitalization calls for empirical research on the risks associated with emerging technologies and the establishment of ethical guidelines for their usage[6].

As general-purpose AI continues to evolve, the goal should not be to eliminate synthetic media, but to ensure that communities can absorb rapid shocks without losing trust in shared truth. Resilience in the digital age is an essential property of strong epistemic foundations and social order. The path forward depends on widespread awareness of AI capabilities and risks, and the proactive policymaking that leverages new technologies for the social good.

## 2. TECHNICAL MECHANISMS AND AI-DRIVEN INFLUENCE

### 2.1 Mechanisms of Automated Influence and Behavioral Signatures

The identification of automated threats requires a deep analysis of their structural and behavioral characteristics. Research involving large-scale datasets has revealed fundamental differences in how humans and bots communicate. The "star" interaction structure typical of bots allows for the rapid dissemination of information from a single source to many receivers, effectively gaming platform algorithms to create the illusion of popularity or consensus[7].

This is often utilized in "astroturfing" campaigns, where top-down propaganda is disguised as a grassroots movement.

**Table 2.** Comparative Analysis of Behavioral Signatures and Interaction Structures: Human Users vs. Automated Accounts (Bots)

| Feature | Human Users | Automated Accounts (Bots) |
|---|---|---|
| Interaction Structure | Hierarchical and Complex | Star-shaped and Centralized |
| Communication Cues | Dialogue Understanding (Replies) | Linguistic Automation (Hashtags) |
| Identity Alignment | Variable and Inconsistent | Rigidly Matched to Target Persona |
| Content Pattern | Diverse and Contextual | Repetitive and Template-based |
| Sentiment Profile | Nuanced | Extreme Negativity or Positivity |

**Sources**: Luo,. 2023, Article 102051.
Ng, L.H.X, 2025.

The evolution of these agents has led to the development of sophisticated personas. Advanced generative models now allow for the creation of accounts that produce highly consistent, human-like content[8]. These personas are categorized into content-based and behavior-based types. Content-based personas focus on the message itself, often acting as announcers or amplifiers that push specific themes like health information or political slogans. Behavior-based personas mimic human social interactions, such as liking and following, to build a credible social graph before launching an influence operation[9].

The strategic deployment of these personas is evident in their role as agenda-builders. Automated accounts have been shown to significantly transfer object and attribute salience to public discourse[10]. In electoral contexts, these agents do not merely spread information; they influence the attribute saliency of issues, driving negative sentiment toward political adversaries and polarizing the electorate. The effectiveness of these mechanisms is enhanced by psychological vulnerabilities. Individuals often struggle to differentiate between authentic and automated accounts, with identification accuracy frequently no better than random guessing[11].

A third-person perception bias often leads users to believe they are less susceptible to influence than others, which increases vulnerability by reducing cognitive vigilance.

### 2.2 Role of Artificial Intelligence in Scaling Disinformation

The scaling of disinformation through artificial intelligence represents a paradigm shift in digital conflict. Large language models allow for the high-quality, autonomous production of text that mimics official announcements or journalistic styles, increasing the psychological and social impact of the disinformation.

These systems can produce politically relevant false news that humans often cannot distinguish from real news. AI systems can now mimic real individuals with uncanny realism, imitating the style of politicians or public figures with greater perceived authenticity than the persons' actual statements. This lowers the barrier for malicious actors to generate and disseminate false content on a massive scale. The rise of AI-driven disinformation necessitates robust countermeasures, ranging from better detection technologies to updated regulations on AI use[12].

The integration of engagement-optimization algorithms further exacerbates the problem. These systems contribute to the rapid dissemination of disinformation by prioritizing content that triggers high emotional responses, regardless of its accuracy[13].

The resulting information disorder is a systemic problem that requires intervention at multiple levels-platform governance, user behavior, and algorithmic transparency.

### 2.3 Technical Challenges in Bot Detection and Classification

The detection of automated agents is a complex and evolving field within social cybersecurity. Early detection methods relied on simple heuristics and rule-based systems, such as analyzing account activity frequency and metadata.

Modern bots are increasingly sophisticated, mimicking human behavior and adjusting their activity patterns to circumvent these methods. Advanced detection techniques now incorporate machine learning, natural language processing, and deep learning algorithms[14].

These models extract a series of features from user and post data to classify the likelihood of an account being automated. Graph-based methods are also used, inferring the probability of automation based on a user's connections and network structure.

Despite these advancements, ethical considerations remain a challenge. Training datasets for detection algorithms often lack diversity, leading to disproportionate inaccuracies based on language or social media platform. Furthermore, errors in classification, such as false positives that label legitimate users as malicious, can lead to the de-platforming of innocent accounts. Ongoing research is needed to develop robust heuristics and classification systems that can adapt to the evolving nature of automated behavior[15].

## 3. PSYCHOLOGICAL IMPACT AND SOCIAL DISRUPTION

### 3.1 Automated Agents and the Engineering of Social Polarization

The systematic engineering of social polarization is a primary objective within digital soft war. Digital platforms provide environments where actors use automated agents to create spaces for social distancing, where hate speech and incivility flourish. The objective of such campaigns is often to extreme positions and increase negativity in the public sphere, creating an environment of social emergency[16].

Mechanisms of polarization through automated agents include several distinct processes. Selective exposure and echo chambers are reinforced by delivery of content that aligns with a user's previous interactions, a process managed by algorithms designed for maximum engagement[17]. Negativisation and denigration involve the use of political bots to disseminate messages containing negative emotions and to attack opponents. The spiral of silence is triggered through coordinated cyber-troops that create a manufactured consensus, making ordinary

users reluctant to express opposing views for fear of harassment. Emotional contagion is achieved by exploiting human biases, using content that triggers strong emotions like anger and fear, which spreads faster than factual information[18].

The impact of these strategies on societal security is multi-dimensional. On an intellectual level, the spread of extremist and sectarian ideas compromises the unity of the national fabric[19]. On a behavioral level, the increase in toxicity and polarization can spill over into the physical world, resulting in protests, riots, and targeted violence.

Case studies of geopolitical crises illustrate these dynamics. In the US-Iran conflict, AI-generated content and automated accounts were used to produce and disseminate deepfake videos of protesters and misleading narratives of military movements. These campaigns were designed to distort global perceptions and erode confidence in reliable information sources, creating a state of digital deception.

### 3.2 Synthetic Media and the Erosion of Epistemic Security

The integration of generative artificial intelligence into automated campaigns has introduced a significant escalation in the ability to manipulate reality through synthetic media. Deepfakes utilize sophisticated algorithms to create hyper-realistic video and audio that are nearly indistinguishable from authentic material[20]. The threat to societal security posed by these technologies is not merely the spread of individual lies, but the cumulative harm to truth and trust. This is the decay of epistemic security-the collective ability of a society to maintain shared conceptions of truth and reality[21].

**Table 3.** Impact of Synthetic Media on Epistemic Security and Social Cohesion Mechanisms

| Impact Category | Specific Threat to Security | Mechanism of Action |
|---|---|---|
| Empowered Inclusion | Silencing Critical Voices | Deepfakes targeting journalists/dissidents |
| Democratic Deliberation | Erosion of Epistemic Quality | Infusion of falsehoods into public debate |
| Institutional Trust | The Liar's Dividend | Allowing denial of authentic evidence |
| Social Cohesion | Heightened Polarization | Content that demonizes identity groups |

**Sources**: Doherty, C. 2024.

Kharvi, P. 2024, 115–122.
Pawelec, M. 2022, Article 19.

The liar's dividend is a particularly insidious effect. As the public becomes increasingly aware of the existence of deepfakes, the mere possibility of fabrication allows political representatives to evade accountability by dismissing uncomfortable truths as AI-generated. This phenomenon leads to post-fact society where empirical evidence no longer serves as a basis for collective decision-making[22].

Deepfakes pose a direct threat to national security by enabling electronic espionage and the fabrication of international crises. For example, a deepfake image depicting an explosion near the Pentagon caused a measurable dip in stock markets, illustrating how even crude synthetic media can trigger real-world consequences by exploiting the speed of digital ecosystems.

### 3.3 Psychological Dimensions and Human-Bot Interaction

The effectiveness of digital soft war is rooted in its ability to exploit human psychology. Automated agents are increasingly designed to provide social support and build relationships with users, a technique utilized by both benign and malicious actors[23]. Malicious bots use human mimicry to gain trust before introducing manipulative narratives. Research into human-bot interactions reveals several critical insights into this dynamic.

The failure of human identification is a primary concern, as individuals are largely unable to differentiate between sophisticated bots and real users. Behavioral studies show that users often overestimate the prevalence of bots when they encounter opposing views, a form of confirmation bias that labels adversaries as non-human to delegitimize their arguments[24]. Exposure to convincing synthetic media can induce cognitive dissonance, where individuals struggle to reconcile new, fabricated information with their existing beliefs[25]. This struggle often results in a further entrenchment of partisan views as users seek comfort in familiar narratives.

The illusion of consensus is another psychological lever. By manipulating engagement metrics like likes and shares, automated accounts create a social pressure that encourages human users to conform to a manufactured majority[26].

These psychological manipulations contribute to a broader societal trust decay. Since trust is a prerequisite for political organization and social cooperation, its erosion diminishes the capacity of a society to solve collective problems.

## 4. GEOPOLITICAL DYNAMICS AND GLOBAL IMPACT

### 4.1 Impact on National Identity and Cultural Security

The long-term deployment of automated agents in soft war targets the cultural integrity of societies. By systematically distorting historical facts and questioning national institutions, these agents aim to reconfigure the national identity of targeted populations. This is often achieved through the amplification of sectarian and regional divisions, utilizing bots to push narratives that highlight internal fissures[27].

Empirical evidence from the UAE and other high-connectivity regions indicates that digital interactions can lead to the introduction of undesirable behaviors and language that challenge traditional family authority and widen generational gaps. The destruction of societal values and customs is a noted outcome of sustained exposure to hostile digital content. Concurrently, the securitization of societal security has become a reaction to these new generation warfare tactics, as states recognize that social media platforms are being used as instruments of power to destabilize identity and coherence[28].

Digital resilience in this context involves not just protecting infrastructure, but preserving the shared values and customs that define a community. This requires a whole-of-society approach where the government, private sector, and civil society collaborate to maintain a healthy social cyber health[29].

The ability of a society to remain resilient in the face of these perceived and actual threats is the ultimate measure of its societal security.

### 4.2 Electoral Integrity and the Distortion of Democratic Processes

Automated agents pose a significant threat to the integrity of democratic processes by manipulating public opinion during election periods. Research indicates that approximately 67% of studies on this topic report attempts to manipulate opinion during elections through disinformation[30]. These agents are used to amplify specific narratives, manipulate engagement metrics, and distort perceptions of popularity or consensus.

In the US 2020 elections, bots were found to augment human users in strategic communications, fabricating narratives to create a polarized society. The collective social pressures exerted by these agents can flip opinions on contentious political or health issues. These strategies undermine the foundation of democratic integrity and erode public trust in electoral outcomes[31].

The pervasive fear of deepfake manipulation in elections further undermines trust in the democratic process. Even when the actual impact on election results is debated, the cumulative harm to democratic stability is significant and growing[32].

This underscores the need for proactive policymaking and the development of strategies to ensure that communities can absorb these shocks without losing trust in shared conceptions of truth.

### 4.3 Global Dynamics of Computational Propaganda

Computational propaganda is now a global problem, utilized to manipulate discourse and target political opponents in at least 81 countries. This new mode of influencing communication is on the rise around the globe, becoming more sophisticated and harder to track at an alarming rate[33].

The assemblage of platforms, autonomous agents, and big data has created a complex web of causality that structures modern life. Several technology diffusion trends explain the emergence of these threats, including the rapid spread of device networks and the shifting demography of users. The majority of digital communications are no longer between people, but involve autonomous agents designed to produce content in interaction with people. These automated programs make up nearly 50% of all online traffic, threatening to undermine the political potential of social media systems[34].

The global nature of the problem necessitates international cooperation and the harmonization of regulations. Establishing a global governance architecture is essential to prevent disinformation actors from exploiting cross-border regulatory gaps. This includes regional regulatory frameworks and cross-border information-sharing mechanisms to foster a healthier online ecosystem.

### 4.4 Case Studies in Regional Conflict and Crisis

Case studies provide valuable insights into the practical application of automated agents in soft war. During the Russo-Ukrainian War, the behavior of automated accounts was found to heighten toxicity and influence thematic discourse in multiple languages[35].

In the US-Iran conflict, AI-generated fake news circulated on social media, influencing public perception and polarization during the geopolitical crisis.

Health crises also present opportunities for automated manipulation. During the COVID-19 pandemic, bots were used to spread misinformation and anti-vaccine narratives, applying social pressure to influence human ideology. These campaigns exploited public uncertainty to increase polarization and unrest[36]. The results demonstrate how digital misinformation can put both national stability and social harmony at risk[37].

These cases underscore the need for real-time detection and intervention strategies. Counter-disinformation efforts should center on monitoring spikes in keywords or sentiment shifts that indicate artificial virality. Collaboration with independent fact-checkers and the implementation of accuracy prompts can help reduce the spread of unintended misinformation.

## 5. STRATEGIC RESPONSE AND DIGITAL RESILIENCE

### Mechanisms of Resilience and Strategic Response

Countering the threats of automated agents and digital soft war requires a multidimensional approach that combines technological, educational, and policy strategies. The goal is to build digital resilience-the capacity of a society to maintain its essential functions despite disruptions in the digital information environment. Analysis of various countermeasure strategies reveals distinct effectiveness ratings and associated challenges.

**Table 4.** Effectiveness Ratings and Core Challenges of Strategic Responses for Enhancing Digital Resilience

| Strategy | Perceived Effectiveness (%) | Core Challenges |
|---|---|---|
| Digital Literacy Programs | 78% | Requires long-term sustained education |
| Fact-Checking Efforts | 65% | Struggles with the speed of AI generation |
| Content Regulation | 59% | Ethical and legal conflicts with speech |
| AI Detection Tech | Variable | Constant technological arms race |

**Sources**: Große, C. et al, 2025.
Rice, J. 2024.
Surjatmodjo, D. et al., 2024, Article 418.

Enhancing digital literacy is consistently emphasized as the most effective defense against information manipulation. Literacy programs should focus on teaching citizens about the role of algorithms in content curation and inoculating them against common manipulation techniques-a process known as pre-bunking[38].

Effective policy must also address the regulatory asymmetry between traditional media and digital platforms. Recommended mechanisms include regulatory sandboxes, where new moderation and detection technologies can be

tested before implementation, and sunset clauses that ensure legislation remains relevant as technology evolves.

Transparency mandates are also critical, requiring platforms to disclose why specific content is targeted at users and granting researchers access to data on automated activity. Parallel to this, advancements in content provenance offer a promising technical path. By embedding cryptographic metadata into digital content, a verifiable chain of identification can be established from creation to dissemination[39]. Prosocial bots can also be deployed to safeguard social trust by providing verified information and moderating online toxicity.

## Conclusion

The proliferation of automated digital agents represents a fundamental shift in the nature of contemporary conflict, where the primary objective is the subversion of societal security through the manipulation of information. These agents, functioning as advanced tools of digital soft war, exploit cognitive biases and platform algorithms to engineer polarization, distort public opinion, and erode the epistemic foundations of society. The integration of artificial intelligence and synthetic media has further complicated this landscape, introduced the liar's dividend and threatened the legitimacy of democratic processes.

The analysis of empirical data suggests that the impact of these agents is not uniform; it is most profound among younger demographics and within highly polarized political environments. The star structure of bot interactions and the strategic use of attribute saliency are key drivers of the increased negativity and social distancing observed in digital discourse. However, the study also indicates that the threat is not insurmountable. Digital resilience can be fostered through a combination of enhanced media literacy, transparent platform governance, and robust international cooperation.

## Recommendations

Based on the findings of this research, the following recommendations are proposed to safeguard societal security in the digital age.

1. National education systems should incorporate advanced digital literacy and cognitive resilience training from an early age, focusing on the mechanics of algorithmic manipulation and the identification of synthetic media. This approach fosters a society that is less vulnerable to the initial spread of disinformation.

2. Governments should mandate regular, independent audits of social media algorithms to identify and mitigate bias and the inorganic amplification of harmful content. Platforms should be required to provide clear provenance markers for all AI-generated material to assist users in distinguishing between authentic and synthetic content.
3. Financial and technical support should be provided to independent, non-partisan fact-checking organizations to enhance their capacity to respond to the speed and scale of automated disinformation. Crowdsourcing and AI technologies should be utilized to improve the effectiveness of these efforts.
4. National security frameworks should be expanded to include dedicated units for information environment monitoring and threat foresight. These units should focus on horizon scanning for emergent disinformation vectors and coordinating early warning responses to influence campaigns.
5. International bodies should work toward the creation of a global governance architecture for digital conduct, establishing norms that prohibit the state-sponsored use of automated agents for the subversion of another nation's societal security. This includes the harmonization of regulations to prevent cross-border regulatory arbitrage.
6. Researchers and developers should be encouraged to create and deploy prosocial bots designed to detect toxicity, provide verified information, and bridge the divide in polarized online communities. By leveraging the same technologies used in soft war for the benefit of social cohesion, nations can turn the tools of the digital threat into instruments of societal resilience. Protecting societal security in the 21st century requires a shift from reactive moderation to proactive resilience, ensuring that the shared values and trust that define a community remain secure in the face of evolving digital conflict.

**References:**

Alnaqbi, Huda Hassan, & Ali, Eman Ahmed Mohamed. (2025). Social media impact on societal security. Frontiers in Sociology, 10, Article 1508542. https://doi.org/10.3389/fsoc.2025.1508542

Arceneaux, Phillip, Anderson, Joshua, Lukito, Josephine, Shah, Mansi, & Kiousis, Spiro. (2026). Social bots as agenda-builders: Evaluating the impact of algorithmic amplification on organizational messaging. Journal of Public Relations Research, 1–34. https://doi.org/10.1080/1062726X.2025.2606676

Boediman, Eko Putra. (2025). Exploring the impact of deepfake technology on public trust and media manipulation: A scoping review. Jurnal Komunikasi, 19(2), 313–334. https://doi.org/10.20885/komunikasi.vol19.iss2.art8

Doherty, Camille. (2024). The danger of deepfakes: Information pollution and epistemic insecurity [Senior thesis, Claremont Colleges]. https://scholarship.claremont.edu/cmc_theses/3750

Ferrara, Emilio. (2023). Social bot detection in the age of ChatGPT: Challenges and opportunities. First Monday, 28(6). https://doi.org/10.5210/fm.v28i6.13185

Große, Christine, & Sundberg, Leif. (2025). Generative AI and digital resilience: A research agenda. Journal of Risk Research, 1–26. https://doi.org/10.1080/13669877.2025.2539105

Haddad, Asma. (2022). The social networks as software tool: "An analysis of the societal security's implications". Journal of Aqlam, 1(1), 75–85. https://asjp.cerist.dz/en/downArticle/860/1/1/207388

Ibrahim, Nourhan Tosson, & Attia, Nada Adel. (2025). The impact of disinformation generated by AI on democracy case studies: The US presidential elections in 2016 & 2024. Review of Economics and Political Science. Advance online publication. https://doi.org/10.1108/REPS-12-2024-0104

Kharvi, Prakash L. (2024). Understanding the impact of AI-generated deepfakes on public opinion, political discourse, and personal security in social media. IEEE Security & Privacy, 22(4), 115–122. https://doi.org/10.1109/MSEC.2024.3405963

Kim, Myungsung, Lee, Seonmi, Kim, Sieun, Heo, Jeong-in, Lee, Sangil, Shin, Yu-Bin, Cho, Chul-Hyun, & Jung, Dooyoung. (2025). Therapeutic potential of social chatbots in alleviating loneliness and social anxiety: Quasi-experimental mixed methods study. Journal of Medical Internet Research, 27, Article e65589. https://doi.org/10.2196/65589

Liberg, Øyvind. (2023). Digital propaganda: Analyzing modern influencing operations [Master's thesis, University of Inland Norway]. https://www.researchgate.net/publication/378705060_Digital_propaganda_Analyzing_modern_influencing_operations

Luo, Han, Meng, Xiao, Zhao, Yifei, & Cai, Meng. (2023). Rise of social bots: The impact of social bots on public opinion dynamics in public health emergencies from an information ecology perspective. Telematics and Informatics, 85, Article 102051. https://doi.org/10.1016/j.tele.2023.102051

Mitter, Silvia, Wagner, Claudia, & Strohmaier, Markus. (2014). Understanding the impact of socialbot attacks in online social networks. arXiv. https://doi.org/10.48550/arXiv.1402.6289

Muringani, Jonathan, & Noll, Josef. (2021). Societal security and trust in digital societies: A socio-technical perspective. 2021 14th CMI International Conference - Critical ICT Infrastructures and Platforms (CMI), Copenhagen, Denmark, 1–7. https://doi.org/10.1109/CMI53512.2021.9663764

Mustafa, Hassan, Luczak-Roesch, Markus, & Johnstone, David. (2025). Conceptualizing the evolving nature of computational propaganda: A systematic literature review. Annals of the International Communication Association, 49(1), 45–60. https://doi.org/10.1093/anncom/wlaf001

Nemr, Christina, & Gangware, William. (2019). Weapons of mass distraction: Foreign state-sponsored disinformation in the digital age. Park Advisors. https://www.state.gov/wp-content/uploads/2019/05/Weapons-of-Mass-Distraction-Foreign-State-Sponsored-Disinformation-in-the-Digital-Age.pdf

Ng, Lynnette Hui Xian, & Carley, Kathleen M. (2025). A global comparison of social media bot and human characteristics. Scientific Reports, 15(1), Article 10973. https://doi.org/10.1038/s41598-025-96372-1

Ng, Lynnette Hui Xian, & Carley, Kathleen M. (2025). The dual personas of social media bots. arXiv. https://doi.org/10.48550/arXiv.2504.12498

Ng, Lynnette Hui Xian, Pan, Ethan, Yoder, Michael Miller, & Carley, Kathleen M. (2026). FATe of bots: Ethical considerations of social bot detection. arXiv. https://arxiv.org/html/2602.05200v1

Pawelec, Maria. (2022). Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. Digital Society, 1(2), Article 19. https://doi.org/10.1007/s44206-022-00010-6

Robles, José-Manuel, Guevara, Juan-Antonio, Casas-Mas, Belén, & Gómez, Daniel. (2022). When negativity is the fuel: Bots and political polarization in the COVID-19 debate. Comunicar, 30(71). https://doi.org/10.3916/C71-2022-05

Romanishyn, Anastasiia, Malytska, Olena, & Goncharuk, Valerii. (2025). AI-driven disinformation: Policy recommendations for democratic resilience. Frontiers in Artificial Intelligence, 8, Article 1569115. https://doi.org/10.3389/frai.2025.1569115

Soomro, Siraj Ahmed, Soomro, Fozia, Chandio, Dastar Ali, & Jatoi, Bakhtawar. (2026). Digital deception in geopolitical crises: The role of AI-generated fake news in the US–Iran conflict. Research Journal for Social Affairs, 4(1), 123–128. https://doi.org/10.71317/RJSA.004.01.0684

Surjatmodjo, Dwi, Unde, Andi Alimuddin, Cangara, Hafied, & Sonni, Alem Febri. (2024). Information pandemic: A critical review of disinformation spread on social media and its implications for state resilience. Social Sciences, 13(8), Article 418. https://doi.org/10.3390/socsci13080418

Tomassi, Andrea, Falegnami, Andrea, & Romano, Elpidio. (2024). Mapping automatic social media information disorder: The role of bots and AI in spreading misleading information in society. PLOS ONE, 19(5), Article e0303183. https://doi.org/10.1371/journal.pone.0303183

Woolley, Samuel C., & Howard, Philip N. (2016). Political communication, computational propaganda, and autonomous agents. International Journal of Communication, 10, 4882–4890. https://ijoc.org/index.php/ijoc/article/view/6298

Woolley, Samuel C., & Howard, Philip N. (2018). Computational propaganda: Political parties, politicians, and political manipulation on social media. Oxford University Press. https://mediawell.ssrc.org/citations/computational-propaganda-political-parties-politicians-and-political-manipulation-on-social-media-4/

Yan, Harry Yaojun, Yang, Kai-Cheng, Shanahan, James, & Menczer, Filippo. (2023). Exposure to social bots amplifies perceptual biases and regulation propensity. Scientific Reports, 13(1), Article 20707. https://doi.org/10.1038/s41598-023-46630-x

[1] Haddad, Asma. (2022). The social networks as software tool: "An analysis of the societal security's implications". *Journal of Aqlam*, *1*(1), 75–85. https://asjp.cerist.dz/en/downArticle/860/1/1/207388

[2] Mustafa, Hassan, Luczak-Roesch, Markus, & Johnstone, David. (2025). Conceptualizing the evolving nature of computational propaganda: A systematic literature review. *Annals of the International Communication Association*, *49*(1), 45–60. https://doi.org/10.1093/anncom/wlaf001

[3] Alnaqbi, Huda Hassan, & Ali, Eman Ahmed Mohamed. (2025). Social media impact on societal security. *Frontiers in Sociology*, *10*, Article 1508542. https://doi.org/10.3389/fsoc.2025.1508542

[4] Ibid, Article 1508542.

[5] Muringani, Jonathan, & Noll, Josef. (2021). *Societal security and trust in digital societies: A socio-technical perspective*. 2021 14th CMI International Conference - Critical ICT Infrastructures and Platforms (CMI), Copenhagen, Denmark, 1–7. https://doi.org/10.1109/CMI53512.2021.9663764

[6] Große, Christine, & Sundberg, Leif. (2025). Generative AI and digital resilience: A research agenda. *Journal of Risk Research*, 1–26. https://doi.org/10.1080/13669877.2025.2539105

[7] Ng, Lynnette Hui Xian, & Carley, Kathleen M. (2025). A global comparison of social media bot and human characteristics. *Scientific Reports, 15*(1), Article 10973. https://doi.org/10.1038/s41598-025-96372-1

[8] Luo, Han, Meng, Xiao, Zhao, Yifei, & Cai, Meng. (2023). Rise of social bots: The impact of social bots on public opinion dynamics in public health emergencies from an information ecology perspective. *Telematics and Informatics, 85*, Article 102051. https://doi.org/10.1016/j.tele.2023.102051

[9] Ng, Lynnette Hui Xian, & Carley, Kathleen M. (2025). *The dual personas of social media bots*. arXiv. https://doi.org/10.48550/arXiv.2504.12498

[10] Arceneaux, Phillip, Anderson, Joshua, Lukito, Josephine, Shah, Mansi, & Kiousis, Spiro. (2026). Social bots as agenda-builders: Evaluating the impact of algorithmic amplification on organizational messaging. *Journal of Public Relations Research*, 1–34. https://doi.org/10.1080/1062726X.2025.2606676

[11] Ng, L.H.W, Op.cit.

[12] Romanishyn, Anastasiia, Malytska, Olena, & Goncharuk, Valerii. (2025). AI-driven disinformation: Policy recommendations for democratic resilience. *Frontiers in Artificial Intelligence, 8*, Article 1569115. https://doi.org/10.3389/frai.2025.1569115

[13] Tomassi, Andrea, Falegnami, Andrea, & Romano, Elpidio. (2024). Mapping automatic social media information disorder: The role of bots and AI in spreading misleading information in society. *PLOS ONE, 19*(5), Article e0303183. https://doi.org/10.1371/journal.pone.0303183

[14] Ferrara, Emilio. (2023). Social bot detection in the age of ChatGPT: Challenges and opportunities. *First Monday, 28*(6). https://doi.org/10.5210/fm.v28i6.13185

[15] Ng, L.H.W, Op.cit.

[16] Robles, José-Manuel, Guevara, Juan-Antonio, Casas-Mas, Belén, & Gómez, Daniel. (2022). When negativity is the fuel: Bots and political polarization in the COVID-19 debate. *Comunicar, 30*(71). https://doi.org/10.3916/C71-2022-05

[17] Liberg, Øyvind. (2023). *Digital propaganda: Analyzing modern influencing operations* [Master's thesis, University of Inland Norway]. https://www.researchgate.net/publication/378705060_Digital_propaganda_Analyzing_modern_influencing_operations

[18] Nemr, Christina, & Gangware, William. (2019). *Weapons of mass distraction: Foreign state-sponsored disinformation in the digital age*. Park Advisors. https://www.state.gov/wp-content/uploads/2019/05/Weapons-of-Mass-Distraction-Foreign-State-Sponsored-Disinformation-in-the-Digital-Age.pdf

[19] Alnaqbi, H.H, Op.cit, Article 1508542.

[20] Kharvi, Prakash L. (2024). Understanding the impact of AI-generated deepfakes on public opinion, political discourse, and personal security in social media. *IEEE Security & Privacy, 22*(4), 115–122. https://doi.org/10.1109/MSEC.2024.3405963

[21] Doherty, Camille. (2024). *The danger of deepfakes: Information pollution and epistemic insecurity* [Senior thesis, Claremont Colleges]. https://scholarship.claremont.edu/cmc_theses/3750

[22] Pawelec, Maria. (2022). Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. *Digital Society, 1*(2), Article 19. https://doi.org/10.1007/s44206-022-00010-6

[23] Kim, Myungsung, Lee, Seonmi, Kim, Sieun, Heo, Jeong-in, Lee, Sangil, Shin, Yu-Bin, Cho, Chul-Hyun, & Jung, Dooyoung. (2025). Therapeutic potential of social chatbots in alleviating loneliness and social anxiety: Quasi-experimental mixed methods study. *Journal of Medical Internet Research, 27*, Article e65589. https://doi.org/10.2196/65589

[24] Yan, Harry Yaojun, Yang, Kai-Cheng, Shanahan, James, & Menczer, Filippo. (2023). Exposure to social bots amplifies perceptual biases and regulation propensity. *Scientific Reports, 13*(1), Article 20707. https://doi.org/10.1038/s41598-023-46630-x

[25] Boediman, Eko Putra. (2025). Exploring the impact of deepfake technology on public trust and media manipulation: A scoping review. *Jurnal Komunikasi, 19*(2), 313–334. https://doi.org/10.20885/komunikasi.vol19.iss2.art8

[26] Ng, L.H.W, Op.cit.

[27] Mitter, Silvia, Wagner, Claudia, & Strohmaier, Markus. (2014). *Understanding the impact of socialbot attacks in online social networks*. arXiv. https://doi.org/10.48550/arXiv.1402.6289

[28] Haddad, A. Op.cit, 75–85.

[29] Ng, Lynnette Hui Xian, Pan, Ethan, Yoder, Michael Miller, & Carley, Kathleen M. (2026). *FATe of bots: Ethical considerations of social bot detection. arXiv.* https://arxiv.org/html/2602.05200v1

[30] Surjatmodjo, Dwi, Unde, Andi Alimuddin, Cangara, Hafied, & Sonni, Alem Febri. (2024). Information pandemic: A critical review of disinformation spread on social media and its implications for state resilience. *Social Sciences, 13*(8), Article 418. https://doi.org/10.3390/socsci13080418

[31] Ibid, Article 418.

[32] Doherty, C. Op.cit.

[33] Woolley, Samuel C., & Howard, Philip N. (2018). *Computational propaganda: Political parties, politicians, and political manipulation on social media*. Oxford University Press. https://mediawell.ssrc.org/citations/computational-propaganda-political-parties-politicians-and-political-manipulation-on-social-media-4/

[34] Woolley, Samuel C., & Howard, Philip N. (2016). Political communication, computational propaganda, and autonomous agents. *International Journal of Communication, 10*, 4882–4890. https://ijoc.org/index.php/ijoc/article/view/6298

[35] Luo, H. Op.cit, Article 102051.

[36] Ibrahim, Nourhan Tosson, & Attia, Nada Adel. (2025). The impact of disinformation generated by AI on democracy case studies: The US presidential elections in 2016 & 2024. *Review of Economics and Political Science*. Advance online publication. https://doi.org/10.1108/REPS-12-2024-0104

[37] Soomro, Siraj Ahmed, Soomro, Fozia, Chandio, Dastar Ali, & Jatoi, Bakhtawar. (2026). Digital deception in geopolitical crises: The role of AI-generated fake news in the US–Iran conflict. *Research Journal for Social Affairs, 4*(1), 123–128. https://doi.org/10.71317/RJSA.004.01.0684

[38] Rice, James. (2024). *Digital resilience in the age of synthetic media*. ECPR The Loop. https://theloop.ecpr.eu/digital-resilience-in-the-age-of-synthetic-media/
[39] Ibid.