International Journal of Advanced Engineering and Management Research

Vol. 11, No. 02; 2026

ISSN: 2456-3676

Ethical Governance and Civil Liberties in AI-supported Behavioral Threat Detection

Dr. Robb Shawe (Lead Author)
Dr. Robert W. Clark (Co-Author)
Capitol Technology University, Department of Sustainability, Department of Cyber-Psychology, 11301 Springfield Road, Laurel, MD 20708, USA

doi.org/10.51505/ijaemr.2026.11217 URL: http://dx.doi.org/10.51505/ijaemr.2026.11217

Received: Mar 09, 2026 Accepted: Mar 17, 2026 Online Published: Mar 24, 2026

Abstract
Advances in artificial intelligence and machine learning have expanded the analytical capabilities available to behavioral threat assessment professionals. These technologies may assist analysts in identifying patterns of grievance expression, identity reinforcement, and behavioral leakage within large volumes of digital communication data. However, the use of AI-supported behavioral analytics also raises significant ethical, legal, and civil liberties considerations. This article examines the governance frameworks necessary to ensure that AI-enabled threat-detection systems operate responsibly within democratic societies. Drawing upon interdisciplinary literature from cybersecurity governance, behavioral threat assessment, data ethics, and public policy, the article proposes a governance framework for responsible AI-supported behavioral threat detection. The analysis emphasizes transparency, accountability, human oversight, and proportionality as essential principles for safeguarding civil liberties while enhancing public safety. The article concludes by outlining policy recommendations and future research directions for developing ethically grounded AI-supported threat assessment systems.

Keywords: behavioral threat assessment, artificial intelligence governance, civil liberties, cybersecurity governance, digital ethics, targeted violence prevention

1. Introduction

Digital communication platforms have become central arenas for the expression of grievances, identities, and ideological narratives. As prior research has demonstrated, individuals who engage in targeted violence frequently exhibit behavioral indicators prior to an attack, including the communication of grievances, threats, or symbolic references to violence (Meloy et al., 2012; Silver et al., 2018). Increasingly, these behavioral signals emerge within online environments where individuals interact with digital communities and publicly express their beliefs and frustrations.

Vol. 11, No. 02; 2026

ISSN: 2456-3676

Advances in artificial intelligence and machine learning have introduced new capabilities for analyzing large volumes of digital communication data. These technologies may assist analysts in identifying patterns of sentiment, linguistic indicators, and behavioral trajectories associated with emerging threats (Jordan & Mitchell, 2015; Ferrara et al., 2016). When integrated with behavioral threat assessment frameworks, AI-supported analytics may enhance analysts' ability to identify concerning behavioral signals within large-scale digital environments.

However, the deployment of AI-enabled threat detection systems also raises significant ethical and governance concerns. Large-scale analysis of digital communication data may create risks related to privacy, algorithmic bias, and potential infringement on civil liberties (Boyd & Crawford, 2012). Without appropriate governance frameworks, the use of computational behavioral analytics may produce unintended consequences that undermine public trust and democratic norms.

This article builds on prior conceptual research examining cyberpsychological indicators of targeted violence, pathways of behavioral leakage, social media intelligence frameworks, and AI-enabled behavioral analytics. These foundational studies explored critical aspects of digital behavioral threat detection, including cyberpsychological markers linked to grievance formation, behavioral leakage patterns indicative of escalating risk, social media intelligence techniques for identifying relevant digital signals, and artificial intelligence–enabled analytical tools for processing large volumes of digital communication data. Collectively, these contributions establish a multidisciplinary analytical base for understanding how digitally mediated behavioral indicators can inform early threat detection. The current article further advances this research by analyzing governance and civil liberties considerations essential for ensuring that AI-supported behavioral threat detection systems function responsibly within democratic societies. This body of work promotes an interdisciplinary approach to recognizing emerging threats within digital environments. Specifically, this article extends the investigation into governance frameworks necessary to facilitate the responsible operation of AI-supported behavioral threat detection systems in democratic contexts.

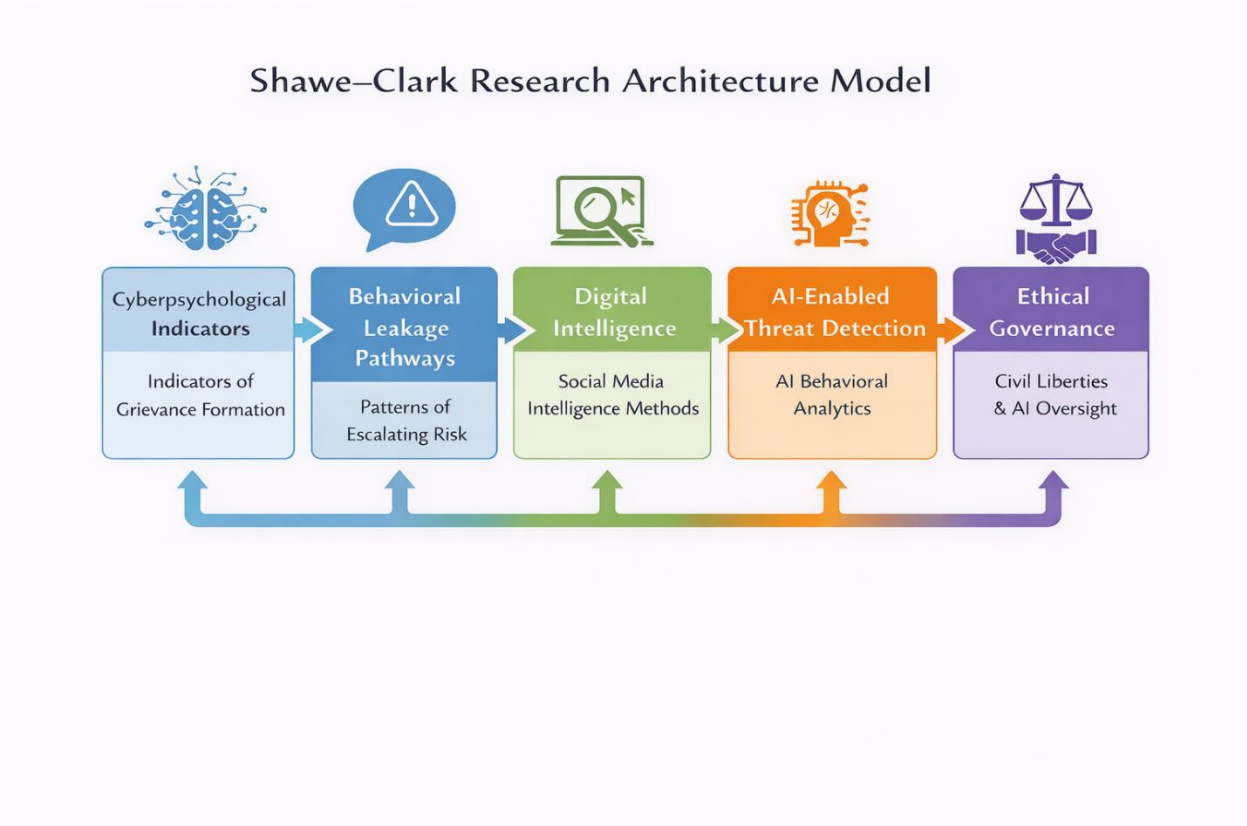
To synthesize the preceding theoretical, operational, and technological dimensions of cyberpsychological threat detection and to illustrate their integration within a unified governance-oriented framework, the following conceptual model presents the relationship between digital behavioral indicators, analytic interpretation, and prevention-oriented decision-making. To illustrate the intellectual structure of this interdisciplinary research program, Figure 1 presents the **Shawe–Clark Research Architecture Model**, which conceptualizes the progression from cyberpsychological indicators and behavioral leakage patterns to digital intelligence methodologies, AI-enabled threat detection, and ethical governance.

Vol. 11, No. 02; 2026

ISSN: 2456-3676

Figure 1

Shawe–Clark Digital Behavioral Threat Detection Architecture



Note. Author created. The Shawe–Clark Research Architecture Model illustrates the five-part interdisciplinary research architecture developed across this article series, linking cyberpsychological indicators, behavioral leakage pathways, digital intelligence methodologies, AI-enabled behavioral threat detection, and ethical governance frameworks.

Taken together, the model demonstrates that effective prevention of targeted violence in digitally mediated environments depends on integrating cyberpsychological insights, structured threat-assessment methodologies, and technologically augmented analytical capabilities. As illustrated in Figure 1, the **Shawe–Clark Research Architecture Model** conceptualizes this progression from identifying digital behavioral signals associated with targeted violence to developing analytical intelligence methods and governance frameworks that support responsible interpretation, proportional intervention, and the protection of civil liberties.

Moreover, this article forms part of a broader program of research examining digitally mediated behavioral threats and the evolving analytical frameworks required to identify, interpret, and responsibly govern them. The research program investigates the progression of online behavioral indicators, the mechanisms through which grievance expression may escalate into behavioral

Vol. 11, No. 02; 2026

ISSN: 2456-3676

leakage, and the role of digital intelligence methods and artificial intelligence–enabled analytics in supporting early threat detection. Complementing these analytical dimensions, the program also explores the ethical and governance considerations necessary to ensure that emerging detection capabilities are implemented responsibly and in alignment with democratic norms and civil liberties. Collectively, this body of work seeks to advance an integrated interdisciplinary framework for understanding and managing digitally mediated threat environments.

Furthermore, this article advances the existing body of literature by proposing an interdisciplinary governance framework that integrates principles of behavioral threat assessment, artificial intelligence ethics, and public policy considerations to promote responsible AI-enabled behavioral threat detection. This publication is part of an ongoing research program examining digitally mediated behavioral threats, including cyberpsychological indicators, behavioral leakage pathways, AI-enabled threat detection, and ethical governance frameworks.

2. Literature Review

Behavioral Threat Assessment and Public Safety

Behavioral threat assessment frameworks have become central tools for preventing targeted violence. Rather than relying on demographic profiling, contemporary threat assessment models emphasize behavioral indicators and contextual analysis to identify individuals who may pose a risk of violence (Borum et al., 1999; Meloy & O'Toole, 2011). Multidisciplinary threat assessment teams often integrate expertise from law enforcement, psychology, and public safety organizations to evaluate concerning behaviors and determine appropriate interventions.

Artificial Intelligence and Digital Behavioral Analysis

Artificial intelligence technologies have significantly expanded the analytical capabilities available to researchers and practitioners analyzing digital communication patterns. Machine learning techniques such as natural language processing and topic modeling allow analysts to detect emerging patterns across large datasets that would be difficult to analyze manually (Blei, 2012; Ferrara et al., 2016). These technologies may assist analysts in identifying linguistic signals associated with grievance expression, hostility, and ideological narratives.

Data Ethics and Algorithmic Governance

The expansion of large-scale data analytics has also raised significant ethical concerns. Scholars examining big data governance have emphasized that algorithmic systems may produce unintended biases or reinforce existing social inequalities if not carefully governed (Boyd & Crawford, 2012). Concerns regarding transparency, accountability, and data privacy have therefore become central issues in the governance of AI-enabled analytical systems.

The present study builds upon earlier work in this research program that examined cyberpsychological indicators and behavioral leakage pathways associated with digitally mediated threat escalation (Shawe & Clark, 2026).

Vol. 11, No. 02; 2026

ISSN: 2456-3676

3. Conceptual Framework

This article proposes a governance framework for AI-supported behavioral threat detection based on four foundational principles:

1. **Transparency** – Analytical systems must operate within clearly defined governance structures and oversight mechanisms.
2. **Accountability** – Human analysts and institutions must remain responsible for decision-making processes informed by AI-generated insights.
3. **Proportionality** – Analytical systems should balance public safety objectives with the protection of civil liberties and privacy rights.
4. **Human Oversight** – AI-enabled systems should function as analytical support tools rather than autonomous decision-making systems.

Together, these principles provide a conceptual foundation for the responsible governance of AI-supported behavioral threat detection systems.

4. Methodological Orientation

This article adopts a conceptual and interdisciplinary methodological approach. Rather than presenting empirical data, the study synthesizes insights from behavioral threat assessment research, artificial intelligence governance literature, and public policy analysis. Conceptual synthesis enables researchers to integrate perspectives from multiple disciplines to develop governance frameworks capable of addressing complex sociotechnical challenges (Rocque & Duwe, 2018).

5. Analysis and Discussion

AI-supported behavioral threat detection systems have the potential to enhance public safety by assisting analysts in identifying patterns of concern within large volumes of digital communication data. However, these technologies must be implemented carefully to avoid unintended consequences. Algorithmic systems may produce false positives, amplify biases present in training data, or create surveillance concerns if deployed without appropriate oversight.

Effective governance frameworks must therefore emphasize transparency, accountability, and ethical oversight. Multidisciplinary review boards, independent auditing mechanisms, and clear data governance policies may help ensure that AI-enabled analytical systems operate responsibly.

6. Policy and Governance Implications

Governments and public safety institutions considering the adoption of AI-supported threat detection technologies should establish clear governance frameworks before deploying such

Vol. 11, No. 02; 2026

ISSN: 2456-3676

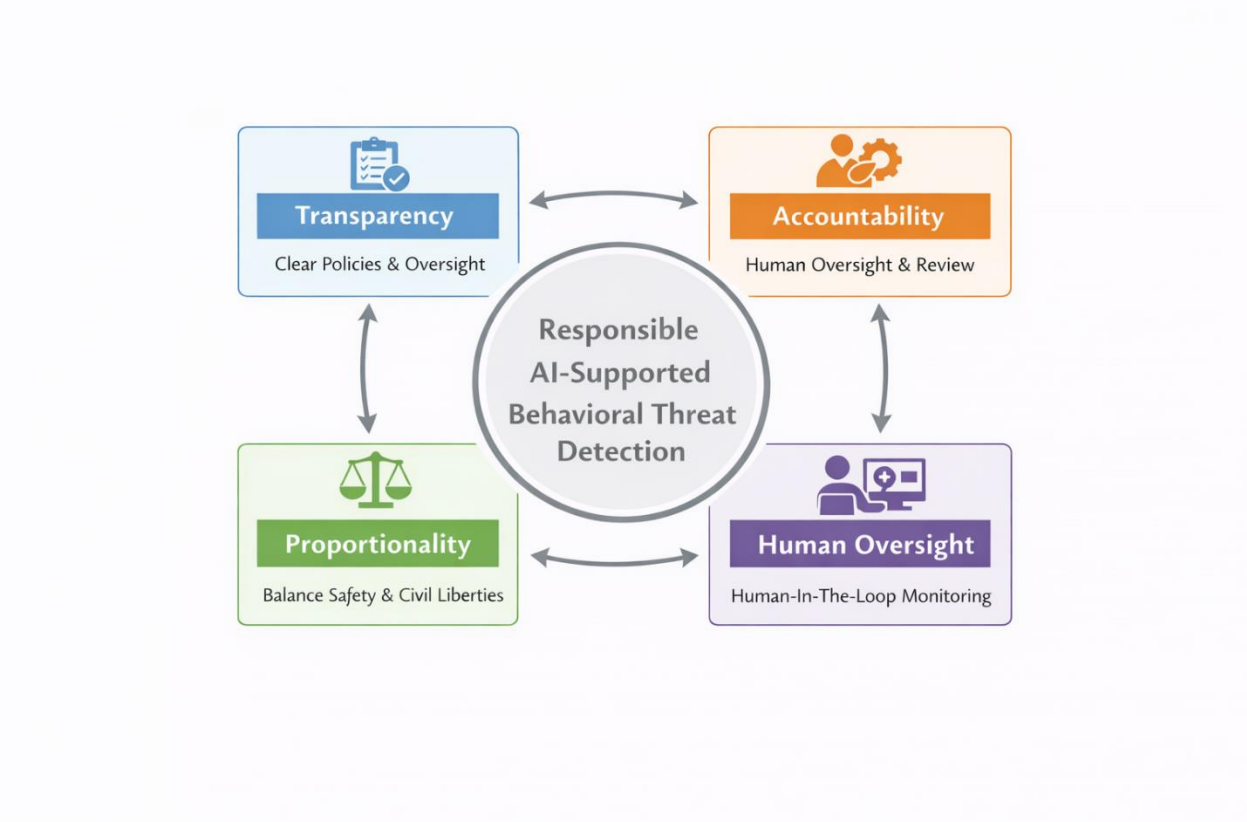
systems. These frameworks should include transparency requirements, algorithmic auditing procedures, and clear protocols for human oversight.

International standards organizations and public policy institutions may also play a role in developing ethical guidelines for AI-supported behavioral analytics. By establishing governance structures that emphasize accountability and protections for civil liberties, policymakers can help ensure that technological capabilities are deployed responsibly.

To conceptualize these governance principles, Figure 2 illustrates an ethical governance framework for AI-supported behavioral threat detection that integrates transparency, accountability, proportionality, and human oversight as core safeguards within AI-enabled analytical systems.

Figure 2

Ethical Governance Framework for AI-Supported Behavioral Threat Detection



Note. Author created. The framework illustrates four governance principles that guide the responsible use of AI-supported behavioral threat detection systems: transparency, accountability, proportionality, and human oversight.

Vol. 11, No. 02; 2026

ISSN: 2456-3676

As illustrated in Figure 2, responsible implementation of AI-supported behavioral threat detection requires governance structures that balance technological capabilities with legal accountability, ethical oversight, and the protection of civil liberties.

7. Limitations and Future Research

This article proposes a conceptual governance framework rather than presenting empirical analysis. Future research should examine how governance principles can be operationalized within real-world AI-supported threat assessment systems. Empirical studies may also explore how governance frameworks influence public trust in AI-enabled public safety technologies.

8. Conclusion

Advances in artificial intelligence have created new opportunities to identify behavioral indicators associated with targeted violence in digital environments. However, these technological capabilities must be balanced with ethical governance frameworks that safeguard civil liberties and democratic principles.

The governance framework proposed in this article emphasizes transparency, accountability, proportionality, and human oversight as essential principles for responsible AI-supported behavioral threat detection. Collectively, the research program developed across this article series integrates cyberpsychology, behavioral threat assessment, social media intelligence, and artificial intelligence analytics to advance interdisciplinary approaches for identifying emerging risks within digitally mediated environments.

To situate the present study within the broader research program from which it emerges, Appendix A presents a conceptual overview of the integrated Shawe–Clark research architecture linking cyberpsychological indicators, behavioral leakage pathways, digital intelligence analysis, artificial intelligence–enabled threat detection, and ethical governance frameworks.

Collectively, the five articles in this research series establish a comprehensive interdisciplinary framework that integrates cyberpsychological dynamics, behavioral leakage patterns, social media intelligence methodologies, artificial intelligence–enabled behavioral analytics, and ethical governance principles to advance the early detection and responsible prevention of targeted violence within digitally mediated environments. Together, these studies establish a foundational interdisciplinary framework for understanding and governing digitally mediated threat environments and provide a basis for future empirical and analytical research in cyberpsychology, digital intelligence, and responsible artificial intelligence.

Authorship Statement

Dr. Robb Shawe and Dr. Robert W. Clark jointly conceptualized the research framework and analytical orientation of this manuscript. Dr. Clark contributed practitioner expertise derived from federal law enforcement leadership and public safety governance experience. Dr. Shawe

Vol. 11, No. 02; 2026

ISSN: 2456-3676

contributed a scholarly synthesis across cyberpsychology, critical infrastructure protection, and public safety governance, including the integration of literature and manuscript preparation. Both authors reviewed and approved the final manuscript.

Author Note and Copyright Statement

Dr. Robb Shawe, PhD, is a scholar-practitioner, U.S. Navy veteran, and Chief Executive Officer of New York Security Consulting Professionals L.L.C. His research focuses on critical infrastructure protection, cybersecurity governance, and resilience systems. Dr. Shawe served as the lead author and was primarily responsible for conceptualization, development of the analytical framework, and manuscript preparation.

Dr. Robert W. Clark, PhD, currently serves as Deputy Mayor of Public Safety for the City of Los Angeles and previously served as an Assistant Special Agent in Charge with the Federal Bureau of Investigation. His research focuses on mass violence prevention, behavioral threat assessment, and the intersection of cyberpsychology and public safety intelligence. Dr. Clark contributed to the literature synthesis, conceptual refinement, and collaborative development of the research framework. Both authors participated in reviewing and approving the final version of the manuscript.

The authors declare no conflicts of interest related to this research. The manuscript represents original scholarly work and is not currently under consideration by another publication outlet. Copyright will remain with the authors unless transferred to a publisher upon acceptance for publication.

Copyright Notice

Vol. 11, No. 02; 2026

ISSN: 2456-3676

References

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Borum, R., Fein, R., Vossekuil, B., & Berglund, J. (1999). Threat assessment: Defining an approach for evaluating risk of targeted violence. *Behavioral Sciences & the Law, 17*(3), 323–337. https://doi.org/10.1002/(SICI)1099-0798(199907/09)17:3<323::AID-BSL349>3.0.CO;2-G

Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society, 15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., Ó hÉigeartaigh, S., Beard, S., Belfield, H., Farquhar, S., & Amodei, D. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation.* Future of Humanity Institute, University of Oxford.

Clark, R. W. (2025). *Mass shootings in America: A law enforcement response* (Unpublished doctoral dissertation). Capitol Technology University.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM, 59*(7), 96–104. https://doi.org/10.1145/2818717

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media.* Yale University Press.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Linkov, I., Trump, B. D., & Keisler, J. M. (2018). Risk and resilience must be independently managed. *Nature Sustainability, 1*(11), 651–652. https://doi.org/10.1038/s41893-018-0196-1

Meloy, J. R., Hoffmann, J., Guldimann, A., & James, D. (2012). The role of warning behaviors in threat assessment: An exploration and suggested typology. *Behavioral Sciences & the Law, 30*(3), 256–279. https://doi.org/10.1002/bsl.1993

Meloy, J. R., & O'Toole, M. E. (2011). The concept of leakage in threat assessment. *Behavioral Sciences & the Law, 29*(4), 513–527. https://doi.org/10.1002/bsl.986

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0).* U.S. Department of Commerce.

National Institute of Standards and Technology. (2024). *Cybersecurity framework (CSF) 2.0.* U.S. Department of Commerce.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown Publishing.

Vol. 11, No. 02; 2026

ISSN: 2456-3676

Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology, 20*(1), 5–14. https://doi.org/10.1007/s10676-017-9430-8

Raji, I. D., Smart, A., White, R., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAccT' 20)* (pp. 33–44). Association for Computing Machinery.

Reason, J. (1997). *Managing the risks of organizational accidents.* Ashgate.

Rocque, M., & Duwe, G. (2018). Rampage shootings: An historical, empirical, and theoretical overview. *Current Opinion in Psychology, 19*, 28–33. https://doi.org/10.1016/j.copsyc.2017.03.005

Shawe, R., & Clark, R. (2026). Behavioral leakage pathways and digital threat escalation: A cyberpsychological analysis. *Manuscript submitted for publication.*

Shawe, R. (2025). *Exploring smart grid technologies' impact on sustainable energy management in New York State* (Publication No. 32284053) [Doctoral dissertation, Capitol Technology University]. ProQuest Dissertations & Theses Global.

Silver, J., Simons, A., & Craun, S. (2018). *A study of the pre-attack behaviors of active shooters in the United States between 2000 and 2013*. Federal Bureau of Investigation, U.S. Department of Justice.

Trump, B. D., Cegan, J. C., Wells, E., & Linkov, I. (2020). Social and ethical implications of autonomous systems. *Environment Systems and Decisions, 40*(4), 532–538. https://doi.org/10.1007/s10669-020-09764-9

Weick, K. E., & Sutcliffe, K. M. (2015). *Managing the unexpected: Sustained performance in a complex world* (3rd ed.). Wiley.

Vol. 11, No. 02; 2026

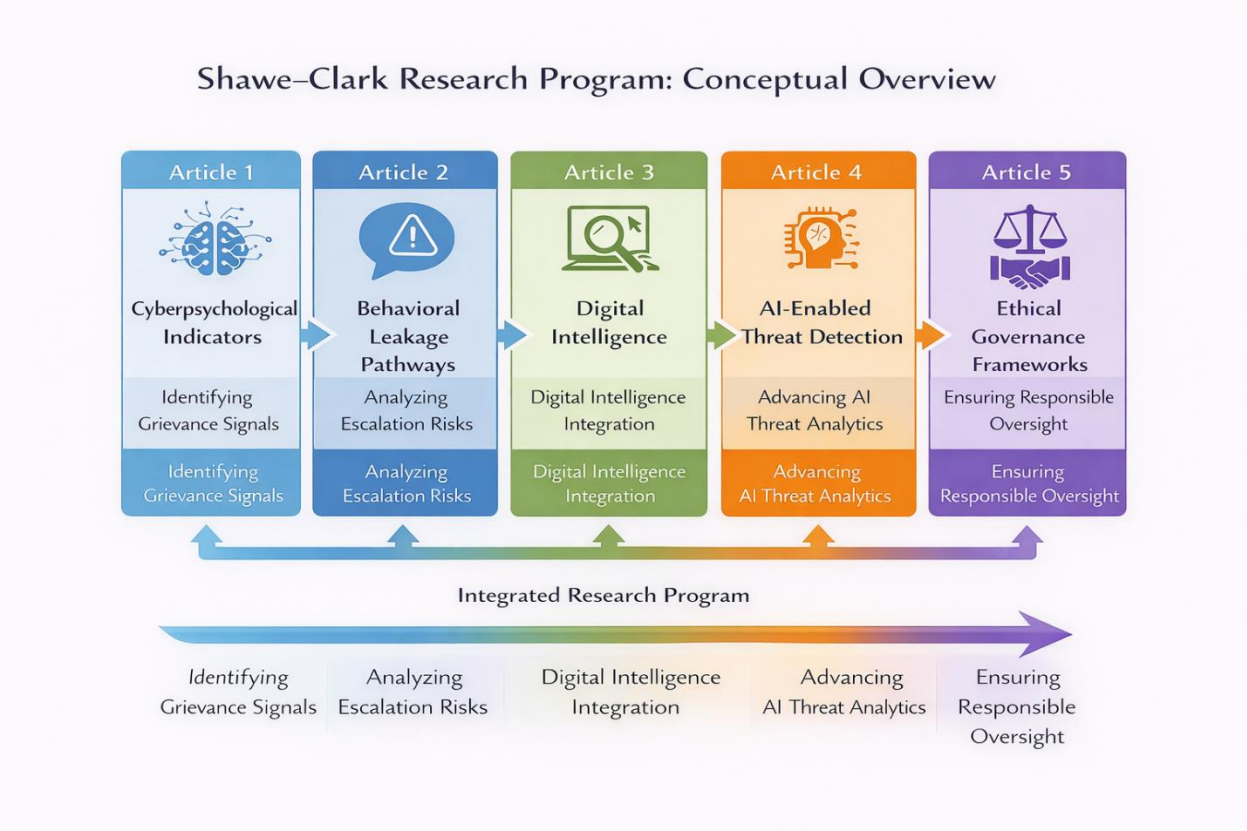
ISSN: 2456-3676

Appendix A

Conceptual Overview of the Shawe–Clark Research Program

Figure A1

Conceptual Overview of the Shawe–Clark Five-Article Research Program



Note. Author created. The figure illustrates the integrated conceptual progression of the five-article research program, beginning with cyberpsychological indicators and behavioral leakage pathways, followed by digital intelligence integration and artificial intelligence–enabled threat detection, and culminating in ethical governance frameworks for responsible oversight of digitally mediated threat environments.

To provide a structured overview of the integrated research architecture, Table A1 summarizes the core focus, guiding research questions, analytical perspectives, and principal scholarly contributions of each article within the five-paper research program.

Vol. 11, No. 02; 2026

ISSN: 2456-3676

Table A1

Summary of the Shawe–Clark Five-Article Research Program

Article	Core Focus	Primary Research Question	Analytical Lens	Key Contribution to the Literature
Article 1	Cyberpsychological Indicators	What early cyberpsychological indicators signal the emergence of digitally mediated grievance formation?	Cyberpsychology: behavioral threat assessment	Introduces a conceptual framework identifying online behavioral indicators associated with grievance formation and potential threat emergence.
Article 2	Behavioral Leakage Pathways	How do grievance-based expressions evolve into observable behavioral leakage across digital environments?	Behavioral threat assessment; online radicalization dynamics	Expands understanding of behavioral leakage pathways and escalation signals within digitally mediated communication spaces.
Article 3	Digital Intelligence Integration	How can digital intelligence methods support the identification and contextual interpretation of emerging behavioral threat signals?	Digital intelligence; OSINT methodologies	Demonstrates how digital intelligence approaches can enhance early threat detection by systematically analyzing online behavioral indicators.
Article 4	AI-Enabled Threat Detection	In what ways can artificial intelligence and machine learning enhance the detection	Artificial intelligence; behavioral analytics	Proposes an AI-enabled analytical framework for identifying complex

Vol. 11, No. 02; 2026

ISSN: 2456-3676

Article	Core Focus	Primary Research Question	Analytical Lens	Key Contribution to the Literature
		and analysis of digitally mediated behavioral threats?		behavioral threat patterns within large-scale digital datasets.
Article 5	Ethical Governance Frameworks	What governance and oversight mechanisms are required to ensure that AI-enabled behavioral threat detection systems are implemented responsibly?	AI governance; socio-technical systems theory	Establishes ethical governance principles and oversight mechanisms to guide the responsible deployment of AI-assisted threat-detection technologies.

Note. The table summarizes the integrated structure of the five-article research program examining cyberpsychological indicators, behavioral leakage pathways, digital intelligence analysis, artificial intelligence–enabled threat detection, and ethical governance frameworks for digitally mediated threat environments.

As illustrated in Table A1, the five articles collectively advance an integrated interdisciplinary research architecture that links cyberpsychological indicators, behavioral leakage dynamics, digital intelligence methodologies, artificial intelligence–enabled threat detection, and ethical governance frameworks to support the responsible analysis of digitally mediated threat environments.