AICIVIA
Journal of Artificial Intelligence, Technology, and Social Transformation

Aicivia: Journal of Artificial Intelligence, Technology, and Social Transformation
Vol. 1, No. 1, 2026
E-ISSN: XXXX-XXXX
punggawajournal.com/aicivia

Can Social Media Tell Us Who Will Kill? Digital Warning Signs and the Problem of Predicting Rare Violence

Zul Khaidir Kadir[1]
[1] Fakultas Hukum, Universitas Muslim Indonesia, Indonesia
Correspondence: zulkhaidir.kadir@umi.ac.id

Article Info

Article History:
Received July 15th, 2026
Revised July 25th, 2026
Accepted July 29th, 2026

Keyword:
Digital prediction paradox; Digital warning signs; Rare violence; Social media.

ABSTRACT

Digital traces often look persuasive after lethal attacks because the outcome gives earlier posts a meaning they did not necessarily have when first encountered. Threats, grievance, fixation, admiration for earlier attackers, and planning language can matter, but their presence does not solve the harder task of identifying a future offender before violence occurs. This article examines that problem through a qualitative integrative review. Evidence from mass public shootings, targeted violence, threat assessment, computational language analysis, and adjacent risk research shows that digital warning behavior is most informative when it develops as a changing trajectory and converges with preparation outside the platform. Yet retrospective offender studies select cases after the outcome is known, while prospective systems confront a low base rate in which many people display fragments of the same pattern and never commit lethal violence. This tension is conceptualized as the digital prediction paradox. Social media is therefore more defensible as a source of behavioral threat information and triage than as an instrument for assigning individual probabilities of homicide, with important implications for automated detection, policing, privacy, fairness, and proportionate prevention.

INTRODUCTION

Online material rarely arrives with its later significance attached. A photograph with a weapon may be bravado, hobby-related content, intimidation, or preparation. A threat can be performative, impulsive, coercive, or part of a developing plan. After a killing, those ambiguities shrink because investigators know which person acted and which fragments preceded the event. Before an attack, police, schools, platforms, relatives, and threat assessment teams face a much larger population of people whose posts are disturbing but whose future behavior remains uncertain. Research on mass public shootings makes the dilemma difficult to ignore. Peterson et al. (2021) found that 46.5% of perpetrators in their sample communicated an intent to harm before their attacks, while Peterson et al. (2023) showed that social media could be used to express grievance, seek validation, engage with earlier mass violence, or shape a desired legacy. Allwinn et al. (2025) similarly identified recognizable preattack warning behaviors in the digital history of a fame-seeking rampage shooter. These studies show that useful information can be present before lethal violence. They do not show that the same information identifies, with acceptable accuracy, which person will eventually kill.

Mass public shootings are particularly visible in this literature because perpetrators often leave searchable records and because the events generate intensive retrospective investigation. They are also a poor shortcut for homicide as a whole. Motives, preparation, target relationships, and the role of publicity vary markedly across mass shootings, and epidemiological work continues to describe substantial heterogeneity rather than a stable offender type (Peterson et al., 2024). Comparisons with homicide offenders, active shooters, and people who die by suicide reveal overlap alongside important differences (Lankford et al., 2021). A digital sign that has value in a planned public attack may contribute

little to understanding an impulsive interpersonal killing or an instrumental homicide in which public recognition has no role. This matters methodologically because spectacular attacks are also the cases most likely to generate abundant online evidence, so the available literature may overrepresent offenders for whom communication and audience were already important. The present article therefore uses rare targeted lethal violence as the main evidentiary domain and treats broader homicide prediction as an implication that requires qualification.

Prediction requires a comparison that offender-only research cannot provide. If a feature appears frequently among attackers but also appears among thousands of non-offenders, it may have explanatory value while offering little prospective specificity. Threatening language, suicidal expression, fascination with previous killers, extremist browsing, and intense grievance all occur outside completed homicide. Their prevalence among people who do not offend determines how many false alarms a preventive system will generate. That denominator is easily forgotten when an inquiry begins with a known attacker and works backward.

Kaati et al. (2026) make the gap visible by showing differences in linguistic markers across violent offenders, high-risk persons of concern, and a broad online comparison population. Planning-related and operational language helped distinguish groups, but the study did not claim that a model could prospectively identify which individual would later commit lethal violence. Trood et al. (2026), working in the different setting of near-lethal and lethal family and intimate partner violence, reached the statistical problem from another direction. Severe outcomes remained uncommon even where familiar risk indicators were available, producing substantial overprediction. The two studies should not be treated as evidence about the same offender population. Read together, however, they expose a recurring difficulty: group-level behavioral differences do not remove individual uncertainty when the outcome is rare.

The review asks what kinds of digital warning behavior appear before rare targeted lethal violence and why those behaviors remain insufficient for reliable individual prediction. Its central argument is that two processes create an illusion of greater predictability than the evidence supports. Retrospective research selects and interprets digital traces with knowledge of the outcome, whereas prospective prevention operates under a low base rate and must evaluate many people who will never commit the event of concern. Social media can make behavioral pathways more visible, sometimes much earlier than conventional institutions would have seen them. Greater visibility, however, is not the same as greater certainty about the endpoint.

RESEARCH METHODS

A qualitative integrative literature review was used because the research question cuts across evidence that is not methodologically uniform. The direct literature includes studies of mass public shootings, foiled attacks, leakage, offender-generated texts, social networks, and threat assessment. Other studies examine violence-risk prediction, computational language markers, online radicalization, platform recommendation, or automated decision support. Combining these as though they estimated one common effect would be misleading. The review instead uses them to answer different parts of the same problem: what can be observed before severe violence, how that information has been interpreted, and what happens to predictive performance once comparison groups and rare outcomes are considered.

The formal corpus contains 31 peer-reviewed journal articles published between 2021 and 2026, all with DOIs. Searches were undertaken in July and August 2026 using Google Scholar for broad discovery, PubMed for health and violence-risk studies when indexed there, Crossref and publisher search pages for bibliographic verification, and backward reference chaining from recent reviews and empirical articles. Search strings combined terms including “mass shooting,” “targeted violence,” “leakage,” “warning behavior,” “social media,” “manifesto,” “violence risk,” “false positive,” “rare violence,” “linguistic markers,” and “threat assessment.” Searches were repeated with narrower pairings when broad combinations returned literature about general aggression rather than rare severe violence. Priority was given to studies that examined preattack behavior, completed or foiled targeted attacks, offender communication, prospective severe-violence prediction, or the interpretation of digital traces. Adjacent studies were retained only when they clarified a mechanism that the direct literature could not answer, especially self-selection, algorithmic exposure, structured risk assessment, or the difference between classification and prediction. The aim was not exhaustive coverage of every violence-risk

publication but a bounded contemporary corpus capable of testing the inferential jump from warning behavior to individual prediction.

Sources were not treated as interchangeable evidence about homicide. Findings from violent extremism, institutional aggression, online hate, and intimate partner violence were used only for the narrower inference they could support. A prison violence screener, for example, can show how predictive performance changes when an outcome is more common, but it cannot establish that the same tool predicts murder. Online radicalization research can clarify self-selection and reinforcement without proving progression to targeted killing. This boundary was applied throughout the synthesis so that adjacent evidence did not silently become homicide evidence.

Each article was first classified by outcome, study design, temporal orientation, and comparison structure. Particular attention was paid to whether warning behavior was identified retrospectively after a known attack, prospectively before an outcome, or through a comparison between offenders and non-offenders. A second reading traced how studies interpreted leakage, fixation, violent identification, operational language, escalation, and offline preparation. These categories were not treated as a checklist. They were compared according to timing, convergence, and the evidentiary claim each study could actually sustain. The recurring distinction was between behavioral relevance and predictive specificity, and that distinction became the organizing principle of the review.

RESULTS AND DISCUSSION

1. Digital Traces Before Lethal Violence

Communication provides the clearest digital evidence because it is observable without requiring an inference about a hidden psychological state. Peterson et al. (2021) found communicated intent in nearly half of the mass public shooters they examined. Rose and Morrison (2023) likewise found substantial variation in leakage among lone-actor terrorists, including differences in content, recipients, and channels. Public posts matter, but some of the most specific disclosures still occur in private or interpersonal communication. Platform monitoring alone therefore cannot solve the detection problem. A public feed may reveal grievance, while a friend, relative, teacher, or colleague holds the more specific statement about timing or intended harm. A visible post is not necessarily the strongest signal simply because it is easy to collect, and a detection system built around public content may privilege accessibility over behavioral significance.

Manifestos and other self-authored texts provide dense material, although their apparent richness can be misleading for prevention. Slemaker (2023) identified established warning behaviors in mass-shooter writings, and LaMothe (2025) found that school-shooter manifestos could reveal intention alongside alienation, rejection, and attempts to justify action. These documents are valuable because they show how offenders narrate a path toward violence in their own words. Many are also selected artifacts, produced close to an attack or released because the perpetrator wanted an audience. What becomes analytically clear after publication may have been unavailable, incomplete, or ambiguous while intervention was still possible. That timing sharply limits their prospective usefulness.

The longitudinal record examined by Allwinn et al. (2025) shows why timing alters the value of an otherwise ambiguous digital sign. Posts, videos, forum activity, diary material, and other self-testimony formed a pattern rather than one alarming statement. A general grievance carries a different implication once the same person begins to narrow a target, acquire means, rehearse, or disclose operational details. Digital evidence becomes more useful when it can be placed inside a developing behavioral trajectory rather than counted as isolated red flags.

Silva and Greene-Colozzi (2024) compared leakage in foiled and completed mass shooting plots and showed that disclosure becomes useful only when somebody recognizes it, reports it, and an institution can assess the concern. Greene-Colozzi et al. (2025), reconstructing the social network around the Marjory Stoneman Douglas High School shooter, found that concerning information was distributed across different people rather than held by one observer. A weapon concern, an online threat, and a change in behavior can each look incomplete in isolation. The completed picture that seems obvious after an attack may never have existed in one place beforehand. Institutional capacity therefore becomes part of the prevention chain.

Platforms can reduce some of that fragmentation because histories are searchable and communications can be connected across time. They also add interpretive noise. A person may maintain several accounts, delete posts, communicate through private channels, or participate in online

subcultures where violent language has meanings that outsiders do not readily recognize. Searchability helps reconstruct chronology, but it does not decide which fragments belong to a meaningful escalation and which are ordinary background behavior.

2. Leakage, Fixation, and Violent Identification

Threat assessment frameworks are useful precisely because they treat warning behavior as dynamic rather than as a fixed profile. Clesle et al. (2024), in a systematic review of instruments for violent extremism, found approaches built around combinations of factors and structured professional judgment rather than one diagnostic sign. The Christchurch case illustrates both the appeal and the limitation of this reasoning. Applying the Path to Intended Violence model and TRAP-18 retrospectively, Allely et al. (2024) identified several proximal warning behaviors in the week before the attack, including pathway behavior, fixation, identification, an energy burst, and last resort. The convergence is informative, especially because the behaviors clustered close to the event rather than appearing as remote background traits. Yet the analysis begins with an offender already known to have attacked. It cannot show how many other persons of concern would have displayed a partly similar configuration during the same period without progressing to violence. This is the point at which threat assessment and prediction must be kept analytically separate.

Fixation and violent identification are particularly difficult to interpret online because repetition is built into platform use. People return to the same grievance, follow communities, reuse images, and repeatedly discuss public figures or earlier attackers for reasons that range from curiosity to identity formation (Kadir, 2025c). Peterson et al. (2023) found that some mass public shooters engaged online with earlier mass violence while expressing grievance or seeking validation. Lankford et al. (2024) showed that gaming-related presentation and online cultural forms could become intertwined with Christchurch-inspired copycat violence. The presence of a borrowed aesthetic or a repeated reference cannot carry the whole inference. It becomes more concerning when it is connected to a narrowing personal trajectory and conduct that reduces the distance between fantasy and action.

Online radicalization research helps explain why immersion should not be confused with mobilization. Frissen (2021) found an association between seeking extremist material and cognitive radicalization, but seeking already reflects selection by the user. Binder and Kenyon (2022), Kenyon et al. (2023), and Mølmen and Ravndal (2023) similarly describe internet use as one part of a pathway shaped by prior attitudes and social interaction. Recommendation systems may intensify what a user encounters. Shin and Jitkajornwanich (2024) found evidence of narrowing exposure on TikTok, while Shaw (2023) cautions that claims of automatic algorithmic radicalization often outrun what current evidence can establish. For prediction, the unresolved boundary is not simply between moderate and extreme content. It is between increasingly extreme attention and actual movement toward targeted violence.

Planning and operational language become more informative when they describe conduct that is closer to implementation than generalized anger or ideological identification. Kaati et al. (2026) found that markers related to planning and operational detail were useful when violent offenders were compared with high-risk persons of concern who had not committed a known attack. The result is more informative than generalized anger because it concerns behavior closer to implementation. Even then, the evidence supports prioritization rather than a forecast. The meaning of a message depends on whether the person has access to the target, whether means are being acquired, whether timing is becoming specific, and whether the communication is part of a wider change in conduct. Social media can supply some of that context, but rarely all of it.

3. The Prediction Paradox

Retrospective studies can make warning signs appear more predictive than they are. Once researchers begin with a known attacker, the outcome changes what earlier information attracts attention. A post about death is read differently when the author later kills. A weapon image acquires a new significance. A vague threat is placed beside later planning and may become part of a coherent narrative. Retrospective research is indispensable for discovering recurrent behaviors, but it cannot by itself establish how often the same fragments appear among people who never offend. This produces a form of retrospective selection: investigators search an already identified offender's history for signals that fit the known endpoint, while the huge pool of people who produced similar fragments without violence remains largely invisible. The problem is apparent in K-12 shooting threats. Peterson et al.

(2024) found that people making threats overlap in some respects with known perpetrators while most threats do not become shootings. Prevention therefore requires decisions among cases that share some warning features but end very differently.

Trood et al. (2026) found substantial overprediction in prospective assessment of near-lethal and lethal family and intimate partner violence because the severe outcome was uncommon. The population differs from targeted public attackers, but the base-rate logic does not. When very few people experience or commit the endpoint of interest, a system can identify many apparently high-risk cases and still be wrong for most of them. Better risk factors improve discrimination, but they do not abolish the arithmetic.

Consider a hypothetical system monitoring one million accounts over a defined period. If ten users later commit a targeted lethal attack and the system identifies nine of them, its sensitivity is 90%. If the same system also flags ten thousand users who never attack, fewer than one in a thousand flagged accounts is a true future attacker. The figures are illustrative, not an estimate of prevalence. They show why a low false-positive rate can still produce a large false-positive population when the event itself is extremely rare.

Measures of discrimination can hide this practical problem when they are read outside the setting in which they were validated. Smeekens et al. (2024) reported predictive validity for a violence screener in prison, where violent and aggressive incidents are far more common than homicide in the general population. Viljoen et al. (2025) found that structured risk instruments generally outperform unstructured judgment across violence and offending outcomes. Those findings support disciplined assessment. They do not mean that an instrument with good group-level performance can identify a future killer with comparable certainty when deployed across a huge low-risk population. Positive predictive value depends not only on model quality but also on how frequently the outcome occurs in the population being screened. Moving a tool from a high-risk institutional setting to platform-scale surveillance changes that denominator dramatically.

The digital prediction paradox follows from the interaction of these two problems. Platforms produce more observable behavior and may reveal escalation that would previously have remained private or dispersed. Yet the same scale supplies an enormous comparison population containing anger, morbid interests, violent language, and other weak markers without a lethal outcome. More behavioral visibility can improve triage while predictive certainty remains low. Hogan and Corăbian (2025) make the related point that generative AI may increase the speed and volume of violence-risk work without resolving validity, explainability, bias, or the authority given to an automated score. Automation can scale useful detection and overclassification at the same time.

4. Social Media Detection and the Cost of Being Wrong

Platforms hold enough behavioral data to make automated screening attractive. Text, images, timestamps, interactions, and changes in account activity can be processed at a scale no human team could review manually. Kaati et al. (2026) show that computational language analysis can distinguish groups using markers linked to grievance, leakage, identification, and operational content. The crucial limit is prospective use. A model trained to separate known violent offenders from a reference group may help triage while still performing poorly when asked to forecast an extremely rare future act among ordinary users.

Specificity becomes harder to maintain in online environments where hostile or violent language is common enough to function as a local communication norm. Maarouf et al. (2024) found that hateful content follows distinctive diffusion patterns and can be shaped by social signals, while Walther (2022) emphasizes the role of interpersonal approval and self-presentation in online hate. In a setting where violent language is common, a keyword or sentiment score may say more about the local communication norm than about an individual's movement toward violence. Context helps, but it can also make weak indicators look normal until more specific behavior appears.

The online-offline relationship is strong enough to justify concern without turning correlation into a personal forecast. Arcila Calderón et al. (2024) found that inflammatory online language could precede some offline hate crime and avoided a causal claim (Kadir, 2025b). Wolfowicz et al. (2022) likewise found heterogeneous media effects on radicalization. These findings support attention to escalation, not a claim that a high content score identifies a future killer.

Missed warnings become especially consequential after a lethal attack because institutions are judged with knowledge of the outcome. Pressure then rises to widen detection so that the next signal is not missed. The resulting harms are less visible because they are spread across people who never become offenders (Kadir, 2026c, 2026b, 2026a). A false positive can lead to questioning, school discipline, police contact, platform restrictions, or prolonged surveillance. Uneven surveillance can also shape the data on which later models are trained, creating a risk that recorded exposure to institutional scrutiny is mistaken for underlying dangerousness. Turner et al. (2022), examining police risk assessment in domestic abuse, show how fairness and predictive performance can pull in different directions as thresholds change. The trade-off is unavoidable: lowering a threshold may reduce missed cases while increasing the number of people subjected to intervention, and the acceptable balance cannot be decided by accuracy statistics alone. The problem is even harder where the predicted event is rarer and the consequences of a dangerousness label are severe.

Privacy concerns increase as systems move from direct threats toward the accumulation of weak indicators. A public threat to a named target differs from private browsing history or an inference drawn from emotional language. Broader data collection can increase coverage while also expanding institutional power over people who have committed no violent act. For that reason, automated detection is easier to defend as triage than as adjudication. It can surface material for human review, connect posts across time, or identify a sudden change. The resulting concern still needs to be tested against context, corroborating information, access to means, target behavior, and evidence of preparation.

5. From Prediction to Behavioral Threat Assessment

Behavioral threat assessment does not require a claim that a practitioner can know who will kill. It asks whether observable conduct suggests movement toward violence and what can be done to manage that movement. This is a better fit for digital evidence than fixed profiling because online traces are often dynamic and incomplete. The difference is visible in leakage research. Silva and Greene-Colozzi (2024) show that disclosure becomes preventive only through what happens after the information reaches another person. Greene-Colozzi et al. (2025) similarly demonstrate how warning information can remain dispersed across a social network. Prevention often fails as an organizational problem before it fails as a prediction problem. Someone has to recognize that separate observations may belong together, know where to report them, and trust that a report will lead to assessment rather than indiscriminate punishment (Kadir, 2026d).

Digital records can strengthen assessment when used as collateral behavioral evidence. A post can establish chronology, repeated references to a target may alter an earlier threat, and planning language can suggest movement from fantasy toward preparation. Kaati et al. (2026) found that operational markers helped distinguish violent offenders from high-risk persons of concern. The signal still belongs inside a broader assessment rather than standing in for one.

Structured approaches are useful because prevention decisions are made under uncertainty rather than after uncertainty has disappeared. Schools, police, platforms, and clinicians may all need to decide whether a concern justifies further assessment, contact, or another form of intervention. The meta-analysis by Viljoen et al. (2025) supports structured assessment over unaided judgment across several violence outcomes. Gill et al. (2021), however, found enough variation across lone-actor terrorists and public mass murderers to caution against a single pathway. Digital signs become more credible when interpreted alongside what the person is doing offline, especially changes that increase capability or bring the person closer to a target. This also explains why a threat assessment process can be useful even when individual prediction remains weak: the immediate question is often whether a situation is becoming more concerning and manageable, not whether the assessor can correctly forecast the eventual legal outcome.

Intervention intensity should rise with specificity, corroboration, and observable movement toward action rather than with the mere accumulation of weak digital markers. A nonspecific hostile post and a direct threat accompanied by target research and preparation should not produce the same response. Treating every flagged user as a future killer would overwhelm assessment systems, magnify false positives, and make reporting less credible. The aim is better judgment under uncertainty, not a technological promise to remove uncertainty.

The evidence does not support the strong version of the question posed in the title. Social media cannot reliably tell practitioners who will kill (Kadir, 2024, 2025a). It can, however, reveal behavior

that deserves attention before some forms of rare targeted violence. Leakage is common enough among mass public shooters to matter, and offender-generated digital material can show grievance, identification, planning, or a changing relationship to earlier attackers (Peterson et al., 2021; Peterson et al., 2023; Allwinn et al., 2025). What matters most is not the presence of one recognizable sign but the way information changes over time and converges with behavior outside the platform. This is why operational detail has a different evidentiary position from generalized anger. It is closer to implementation, although still not equivalent to intent or an inevitable outcome. Treating all digital markers as interchangeable would recreate the profile logic that behavioral threat assessment was developed to avoid. That limit matters for both theory and prevention.

The digital prediction paradox explains why more data do not automatically solve the problem. After an attack, investigators can select the posts that fit the known outcome and reconstruct a pathway with unusual clarity. Before an attack, the same institution sees many people who reproduce fragments of that pattern. Kaati et al. (2026) demonstrate that language can improve differentiation between groups, while Trood et al. (2026) show how a low base rate can still produce substantial overprediction in a prospective setting. The paradox is not that digital evidence is useless. It is that platforms expand both the observable pathway and the comparison population. They improve the amount of information available for triage at the same time that the scale of weak signals limits individual predictive certainty.

Fixation, grievance, violent identification, and online immersion should not be treated as interchangeable steps in a universal sequence. They become more informative as they are connected to conduct that reduces practical barriers to violence. Allely et al. (2024) identified multiple proximal behaviors in a retrospective TRAP-18 analysis of the Christchurch attacker, but the value of that finding lies in convergence and temporal proximity rather than in any single indicator. Clesle et al. (2024) similarly show that contemporary threat assessment instruments rely on structured combinations rather than one sign. Digital criminology therefore gains little from searching for a new online profile of the future killer. The more defensible task is to understand how digital behavior enters an already developing pathway and when it changes the assessment of capability, target focus, or preparation. The theoretical unit should be the changing relationship among observable behaviors, not a static inventory of suspicious content or labels.

Automated systems can assist that task in a limited role. They can retrieve material, connect events across time, detect changes in language, or prioritize accounts for human review. They should not be treated as machines for producing a definitive probability of homicide. Hogan and Corăbian (2025) warn that generative AI will introduce familiar problems of validity and bias into violence-risk work while adding new questions about explainability and professional authority. The governance issue is therefore not simply whether an algorithm may be used. It is whether the purpose of the system, its threshold, the data it relies on, and the consequences of being flagged are proportionate to what the evidence can support. A triage tool that triggers contextual review is different from a score that directly drives police action, account removal, or another coercive response.

The richest digital evidence still comes from spectacular targeted attacks whose perpetrators leave public traces, manifestos, threats, or identifiable networks. Less performative homicide produces fewer analyzable records and may involve different causal structures, creating a visibility bias in the literature itself. Prospective research that follows persons of concern and includes non-offender comparison groups would do more to clarify predictive specificity than additional post hoc profiling of famous attackers. Researchers also need to report how often proposed markers occur outside offender samples. Digital traces may help practitioners decide where closer assessment is warranted, but they do not remove the uncertainty between disturbing communication and future killing.

CONCLUSION

Social media can reveal warning behavior before rare targeted lethal violence, but its value remains conditional. Communication of intent and operational detail become more informative when they change over time and converge with preparation outside the platform. Retrospective studies identify these patterns, yet they benefit from knowing the outcome in advance. Prospective prevention faces the opposite problem: many people display fragments of the same behavior and never commit a lethal act. Low base rates therefore remain a major barrier to reliable individual prediction even when group-level markers are meaningful.

The strongest use of digital information is consequently behavioral threat assessment and triage rather than murder prediction. Automated systems may help locate material or detect escalation, but final interpretation requires context, corroboration, and a response proportionate to what is actually known. Most direct evidence reviewed here concerns mass public shootings and other forms of targeted violence, so the conclusions should not be generalized without qualification to domestic, gang-related, robbery-related, or spontaneous interpersonal homicide. The evidence base is also still heavily retrospective. More prospective work is needed before the digital prediction paradox can be treated as a stable empirical model rather than a synthesis of recurring limitations. Greater digital visibility can create earlier opportunities for prevention. It does not provide a dependable way to identify the future killer in advance.

REFERENCES

Allely, C. S., Wicks, S. J., & McLaren, S. A. (2024). The application of the Path to Intended Violence model and the TRAP-18 in the case of the Christchurch Mosque Shooter. Journal of Threat Assessment and Management, 11(1), 32–47. https://doi.org/10.1037/tam0000211

Allwinn, M., King, S., Tultschinetski, S., & Görgen, T. (2025). Preattack warning behaviors in the digital space: A case study of a fame-seeking rampage shooter. Journal of Threat Assessment and Management, 12(3), 178–196. https://doi.org/10.1037/tam0000240

Arcila Calderón, C., Sánchez Holgado, P., Gómez, J., Barbosa, M., Qi, H., Matilla, A., Amado, P., Guzmán, A., López-Matías, D., & Fernández-Villazala, T. (2024). From online hate speech to offline hate crime: The role of inflammatory language in forecasting violence against migrant and LGBT communities. Humanities and Social Sciences Communications, 11, 1369. https://doi.org/10.1057/s41599-024-03899-1

Binder, J. F., & Kenyon, J. (2022). Terrorism and the internet: How dangerous is online radicalization? Frontiers in Psychology, 13, 997390. https://doi.org/10.3389/fpsyg.2022.997390

Clesle, A., Knäble, J., & Rettenberger, M. (2024). Risk and threat assessment instruments for violent extremism: A systematic review. Journal of Threat Assessment and Management. Advance online publication. https://doi.org/10.1037/tam0000223

Frissen, T. (2021). Internet, the great radicalizer? Exploring relationships between seeking for online extremist materials and cognitive radicalization in young adults. Computers in Human Behavior, 114, 106549. https://doi.org/10.1016/j.chb.2020.106549

Gill, P., Silver, J., Horgan, J., Corner, E., & Bouhana, N. (2021). Similar crimes, similar behaviors? Comparing lone-actor terrorists and public mass murderers. Journal of Forensic Sciences, 66(5), 1797–1804. https://doi.org/10.1111/1556-4029.14793

Greene-Colozzi, E. A., Schildkraut, J., Arrigo, L., Krebs, L., & Klein, B. R. (2025). Social network analysis as a tool for understanding mass shooting prevention: A case study of the Marjory Stoneman Douglas High School shooting. Journal of Criminal Justice, 101, 102540. https://doi.org/10.1016/j.jcrimjus.2025.102540

Hogan, N. R., & Corăbian, G. (2025). Generative artificial intelligence in violence risk assessment: Emerging technology and the ethics of the inevitable. Behavioral Sciences & the Law, 43(6), 606–615.

Kadir, Z. K. (2024). Menggugat Netralitas Hukum Pidana: Perdebatan Ideologis di Balik Kebijakan Kriminal di Negara-Negara Liberal. *Eksekusi: Jurnal Ilmu Hukum Dan Administrasi Negara*, *2*(4), 380–400.

Kadir, Z. K. (2025a). Kejahatan atau Kebijakan? Membongkar Bias Ideologis dalam Realisme Kanan dan Kiri di Era Neo-Liberalisme. *Jurnal Intelek Dan Cendikiawan Nusantara*, *1*(6), 11146–11161.

Kadir, Z. K. (2025b). Kejahatan Berbasis Identitas Digital: Menggagas Kebijakan Kriminal untuk Dunia Metaverse. *Jurnal Litigasi Amsir*, *12*(2), 124–137.

Kadir, Z. K. (2025c). Membongkar Relasi Tersembunyi: Pola Hubungan Pelaku Dan Korban Kekerasan Seksual Dalam Perspektif Kriminologi. *Jurnal Padamu Negeri*, *2*(2), 20–29.

Kadir, Z. K. (2026a). Beyond Firearms: Mapping Global Patterns of Non-Firearm Homicide in Comparative Criminological Perspective. *Punggawa Global Research: Jurnal Multidisiplin*, *1*(2), 1–10.

Kadir, Z. K. (2026b). Mapping Recorded Homicide in Indonesia: An Ecological Criminological Analysis Using BPS Data. *Punggawa Law Review*, *1*(2), 9–18.

Kadir, Z. K. (2026c). Narrative Closure in Honor killing Cases: How Judgments Stabilise Meaning, Eliminate Ambiguity, and Produce Sentencing Certainty. *Punggawa Law Review*, *1*(1), 1–10.

Kadir, Z. K. (2026d). Preventing Murder Without Expanding Punishment: Rethinking Homicide Policy Beyond Deterrence. *Punggawa Law Review*, *1*(3), 49–58. https://doi.org/10.67707/plr.v1i3.40

Kaati, L., Akrami, N., Gibson, K. A., & Craun, S. W. (2026). Linguistic markers in the writings of violent offenders, high-risk individuals, and online users. SN Social Sciences, 6, 71. https://doi.org/10.1007/s43545-026-01360-5

Kenyon, J., Binder, J. F., & Baker-Beall, C. (2023). Online radicalization: Profile and risk analysis of individuals convicted of extremist offences. Legal and Criminological Psychology, 28(1), 74–90. https://doi.org/10.1111/lcrp.12218

LaMothe, J. (2025). Decoding school mass shooter manifestos to predict behavior: A theory based qualitative study. Professional Psychology: Research and Practice, 56(4), 266–274. https://doi.org/10.1037/pro0000638

Lankford, A., Allely, C. S., & McLaren, S. A. (2024). The gamification of mass violence: Social factors, video game influence, and attack presentation in the Christchurch mass shooting and its copycats. Studies in Conflict & Terrorism, 1–25. https://doi.org/10.1080/1057610X.2024.2413184

Lankford, A., Silver, J., & Cox, J. (2021). An epidemiological analysis of mass public shooters and active shooters: Quantifying key differences between perpetrators and the general population, homicide offenders, and people who die by suicide. Journal of Threat Assessment and Management, 8(4), 125–144. https://doi.org/10.1037/tam0000166

Maarouf, A., Pröllochs, N., & Feuerriegel, S. (2024). The virality of hate speech on social media. Proceedings of the ACM on Human-Computer Interaction, 8(CSCW1), Article 186, 1–22. https://doi.org/10.1145/3641025

Mølmen, G. N., & Ravndal, J. A. (2023). Mechanisms of online radicalization: How the internet affects the radicalization of extreme-right lone actor terrorists. Behavioral Sciences of Terrorism and Political Aggression, 15(4), 463–487. https://doi.org/10.1080/19434472.2021.1993302

Peterson, J., Densley, J., Spaulding, J., & Higgins, S. (2023). How mass public shooters use social media: Exploring themes and future directions. Social Media + Society, 9(1). https://doi.org/10.1177/20563051231155101

Peterson, J., Erickson, G., Knapp, K., & Densley, J. (2021). Communication of intent to do harm preceding mass public shootings in the United States, 1966 to 2019. JAMA Network Open, 4(11), e2133073. https://doi.org/10.1001/jamanetworkopen.2021.33073

Peterson, J., Densley, J., Riedman, D., Spaulding, J., & Malicky, H. (2024). An exploration of K–12 school shooting threats in the United States. Journal of Threat Assessment and Management, 11(2), 106–120. https://doi.org/10.1037/tam0000215

Peterson, J. K., Densley, J. A., Hauf, M., & Moldenhauer, J. (2024). Epidemiology of mass shootings in the United States. Annual Review of Clinical Psychology, 20(1), 125–148. https://doi.org/10.1146/annurev-clinpsy-081122-010256

Rose, M., & Morrison, J. (2023). An exploratory analysis of leakage warning behavior in lone-actor terrorists. Behavioral Sciences of Terrorism and Political Aggression, 15(2), 179–214. https://doi.org/10.1080/19434472.2021.1900325

Shaw, A. (2023). Social media, extremism, and radicalization. Science Advances, 9(35), eadk2031. https://doi.org/10.1126/sciadv.adk2031

Shin, D., & Jitkajornwanich, K. (2024). How algorithms promote self-radicalization: Audit of TikTok's algorithm using a reverse engineering method. Social Science Computer Review, 42(4), 1020–1040. https://doi.org/10.1177/08944393231225547

Silva, J. R., & Greene-Colozzi, E. A. (2024). Assessing leakage-based mass shooting prevention: A comparison of foiled and completed attacks. Journal of Threat Assessment and Management, 11(4), 203–217. https://doi.org/10.1037/tam0000205

Slemaker, A. (2023). Studying mass shooters' words: Warning behavior prior to attacks. Journal of Threat Assessment and Management, 10(1), 1–17. https://doi.org/10.1037/tam0000198

Smeekens, M. V., De Vries Robbé, M., Popma, A., & Kempes, M. (2024). The Risk Screener Violence (RS-V): Retrospective prediction of violent and aggressive incidents within the prison setting. Frontiers in Psychology, 15, 1359535. https://doi.org/10.3389/fpsyg.2024.1359535

Trood, M. D., Spivak, B. L., Ogloff, J. R. P., & McEwan, T. E. (2026). The limits of predicting near lethal and lethal family and intimate partner violence. Journal of Family Violence. Advance online publication. https://doi.org/10.1007/s10896-025-01029-2

Turner, E., Brown, G., & Medina-Ariza, J. (2022). Predicting domestic abuse (fairly) and police risk assessment. Psychosocial Intervention, 31(3), 145–157. https://doi.org/10.5093/pi2022a11

Viljoen, J. L., Goossens, I., Monjazeb, S., Cochrane, D. M., Vargen, L. M., Jonnson, M. R., Blanchard, A. J. E., Li, S. M. Y., & Jackson, J. R. (2025). Are risk assessment tools more accurate than unstructured judgments in predicting violent, any, and sexual offending? A meta-analysis of direct comparison studies. Behavioral Sciences & the Law, 43(1), 75–113. https://doi.org/10.1002/bsl.2698

Walther, J. B. (2022). Social media and online hate. Current Opinion in Psychology, 45, 101298. https://doi.org/10.1016/j.copsyc.2021.12.010

Wolfowicz, M., Hasisi, B., & Weisburd, D. (2022). What are the effects of different elements of media on radicalization outcomes? A systematic review. Campbell Systematic Reviews, 18(2), e1244. https://doi.org/10.1002/cl2.1244