Erasmus School of Social and Behavioural Sciences

Authority for the prevention of online Terrorist Content and Child Sexual Abuse Material

From Online Terrorist Content Exposure to Real-World Mobilisation: Understanding Susceptibility, Assessing Risk, and Informing Interventions

Stavroula Spernovasili, Dominika Grudzień, Neelav Novo Parashar

MSc Forensic and Legal Psychology, Erasmus University Rotterdam.

Prepared for the Authority for the prevention of Online Terrorist Content and Child Sexual Abuse Material

Executive Summary

The Authority for the prevention of Online Terrorist Content and Child Sexual Abuse Material (ATKM) assesses whether online material meets the legal criteria for terrorist content under Regulation (EU) 2021/784 and may therefore require removal. Within this mandate, the present report examines how behavioural and psychological evidence can provide context for identifying characteristics of terrorist content that may carry a greater potential to facilitate radicalisation or movement towards real-world mobilisation. The report does not seek to assess or profile the individuals sharing such material. Instead, it considers how different forms and functions of content, including ideological expression, violent imagery, recruitment material, gamified participation, and operational or tactical information, may be interpreted and prioritised.

The report first describes how the contemporary online environment differs from the centralised propaganda systems traditionally examined in radicalisation research. Extremist content is now frequently distributed through decentralised and participatory spaces in which humour, memes, gaming references, competition, social recognition, and peer interaction can contribute to engagement. These dynamics are also visible in loosely organised networks that reward escalating acts of violence or self-harm. Children and adolescents may be particularly exposed to this environment because violent material can be encountered through social media, video games, streaming platforms, and algorithmic recommendation systems, including without being actively sought.

The report then reviews the psychological, behavioural, and neurological evidence concerning exposure to violent digital content. Repeated exposure may contribute to desensitisation, reduced empathic responding, moral disengagement, aggressive priming, and more permissive attitudes towards violence. Some studies also associate violent-media

exposure with increased aggression and other adverse outcomes. However, these effects are generally modest, findings are not fully consistent, and much of the evidence is correlational or based on laboratory measures that do not reflect serious real-world violence. The evidence therefore does not support treating exposure to, or sharing of, violent content as a direct predictor of mobilisation. Rather, exposure should be understood as one probabilistic influence whose effects depend on the interaction between individual vulnerabilities, social circumstances, and protective factors.

The report subsequently examines factors that may increase susceptibility to gamified extremist environments. These include psychological strain, social isolation, unmet needs for belonging and recognition, identity uncertainty, sensation-seeking, and limited family or community support. Gamified structures may reinforce engagement by rewarding participation, status, competition, and behavioural escalation. Conversely, parental involvement, supportive relationships, digital literacy, community belonging, and access to alternative sources of identity and meaning may reduce susceptibility. These factors are presented as contextual explanations rather than as a checklist or a basis for ATKM to infer an individual's psychological characteristics from content.

The operational section considers how content and, where lawfully available, associated behavioural signals may be interpreted more systematically. It discusses four heuristic types of content-sharing behaviour: ideological-operational content, status-driven network content, performative shock content, and incidentally redistributed content. Building on this distinction, the report proposes a provisional risk-tiered framework organised around four dimensions: escalation, network, ideology, and capability and operational behaviour. Operational indicators, such as tactical learning, logistical planning, weapon acquisition, target selection, reconnaissance, or attack preparation, are weighted most heavily because

they may distinguish movement towards mobilisation more reliably than ideological expression alone. The framework assigns content to minimal, lower, medium, or high-risk tiers to support prioritisation, but it remains conceptual rather than diagnostic and requires piloting, inter-rater reliability testing, legal safeguards, and calibration against relevant case outcomes.

Finally, the report presents recommendations across three levels of prevention. Primary prevention focuses on parental awareness, media literacy, digital resilience, and education for young people. Secondary prevention includes search-based redirection, credible-messenger interventions, counter-narratives, active-inoculation games, and referral to psychosocial support by stakeholders with the appropriate mandate and access to user-level information. Tertiary and operational measures include prioritising content containing operational or tactical indicators, strengthening trained human review of coded or ambiguous material, introducing technical friction into platform design, and supporting cross-platform hash-sharing for known extremist content. Together, these recommendations aim to strengthen prevention and content prioritisation without treating ideological or provocative expression alone as evidence of an individual's progression towards violence.

Introduction: The Behavioural Decision Problem

The ATKM operates under a defined regulatory mandate. Under Regulation (EU) 2021/784 (2021), it is empowered to issue removal orders which require hosting service providers to take down terrorist content within one hour. ATKM analysts assess whether content meets the criteria for terrorist content set out in the Regulation and therefore requires removal. Given the volume and diversity of content encountered online, a further consideration is how to prioritise attention towards content that may carry a greater risk of facilitating radicalisation or mobilisation. Crucially, the Authority's remit is confined to content assessment. ATKM analysts assess

terrorist content itself, but they do not profile the person who shares it. This distinction is particularly important when considering behavioural and psychological research, as evidence about an individual's vulnerabilities or trajectory cannot be directly inferred from ATKM's content assessment alone.

Within this remit, behavioural research can nevertheless provide useful context for understanding why certain types of terrorist content may be more consequential than others. Online extremist and violent content ecosystems contain material that may serve different functions, ranging from ideological expression and identity performance to recruitment, tactical learning, or the encouragement of violence. Understanding these different functions can help identify characteristics of content that warrant greater attention when considering its potential to contribute to radicalisation or mobilisation. The behavioural decision problem addressed in this report is therefore not whether an individual poses a threat, but how evidence on online behaviour can inform the prioritisation and interpretation of terrorist content. In particular, the report considers which characteristics of content may be associated with greater potential for radicalisation of movement towards mobilisation, while recognising that content-level evidence does not provide a direct assessment of an individual's risk trajectory.

The distinction also helps avoid conflating symbolic or performative forms of online participation with evidence or mobilisation. The report therefore considers how psychological vulnerabilities, social dynamics, and features of gamified digital environments may shape engagement with extremist content, while distinguishing these broader behavioural processes from the content-based assessment undertaken by ATKM. This requires a careful interpretation of behavioural evidence without extending content assessment into individual profiling. Against this background, the report addresses three questions.

The first concerns how the online content ecosystem has evolved, what the current evidence shows about the effects of exposure to violent content, and which existing counterterrorism policy responses are effective in reducing risk or may produce unintended consequences.

The second asks what psychological, neurological, and behavioural effects exposure to violent content, and especially violent video games, has on psychological development, brain functioning, and behaviour of children and young people in the context of today's highly immersive and algorithm-driven digital world, and what protective and risk factors influence the extent and nature of these effects.

The third explores how psychological vulnerabilities may make individuals more susceptible to engaging with gamified extremist environments. It then examines observable indicators that may signal a transition towards mobilisation and, finally, considers the protective factors that may reduce the likelihood of individuals progressing towards real-world violence.

This report concludes with practical recommendations addressing each of these three questions in turn.

Why Today’s Online Environment is Different

In the past, research on radicalisation focused on centralised propaganda. Within this traditional model, extremist groups disseminated their ideology to what researchers called ‘passive audiences’ (LaCroix & Berkovits, 2026). Originally, ideological commitment, group membership, or direct contact with recruiters were considered as indicators of escalation risk (LaCroix & Berkovits, 2026). However, a shift has occurred. Modern extremist spaces tend to be more decentralised and shaped by user participation. Here, ideas are spread through peer-to-peer interaction instead of formal organisational structures (LaCroix & Berkovits, 2026; Schlegel, 2021). Extremist ideology may also be communicated through

humour, memes, and aesthetic imitation rather than through direct ideological instruction (Schlegel, 2021). Hence, engagement in these spaces is therefore driven as much by social dynamics, identity formation, and recognition within online communities as by ideology itself.

These participatory dynamics are becoming more structured through gamification. Specifically, gamification refers to the application of game design elements such as competition, scoring systems, and performance rewards in non-game contexts to increase engagement (Lankford et al., 2024). When violence is experienced as an interactive, competitive, and entertaining activity, it may cause individuals to become desensitised and disengaged (Bandura, 1999; Schlegel, 2021). The 2019 Christchurch attack, live-streamed from a first-person perspective resembling a first-person shooter video game, illustrated how violent acts can be turned into immersive, real-time experiences. Online communities subsequently reinforced this pattern by circulating modified gaming environments in which the attacker appeared as a playable character (Lankford et al., 2024). Extremist groups have also adopted gaming terminology to derive attacks, comparing 'high scores' or 'leaderboards' based on casualty counts (Lakhani & Wiedlitzka, 2023), while some attacker manifestos take the form of tactical 'walkthroughs' providing step-by-step instructions on weapon selection and strategy (Lankford et al., 2024).

The same escalatory, gamified logic now extends beyond ideologically defined extremism into loosely organised networks built around self-harm and violence for their own sake. In February 2026, the Global Network on Extremism and Technology (GNET) analysed online groups such as 764, Maniac Murder Cult, and No Lives Matter, in which young people climb rank according to whether they have documented and demonstrated escalating acts of self-harm and violence against themselves or others, as opposed to according to the sincerity of their beliefs. Within these networks, a young

person filming themselves cutting symbols into their own skin functions as a first step rather than an outlier behaviour (Global Network on Extremism and Technology [GNET], 2026). Altogether, these developments point to a broader shift toward treating violence, including violence directed at the self, as something to be optimised, ranked, and measured, further blurring the boundary between passive consumption of violent content and active behavioural escalation.

Children and adolescents are especially exposed to this evolving digital landscape. They are the first generation to grow up with constant and largely unfiltered access to the internet and social media, engaging with platforms such as Instagram, Snapchat and Facebook at increasingly younger ages than previous generations (Rideout, 2022; ITU, 2024). At the same time, studies consistently show that children are frequently exposed to age-inappropriate and violent content, including depictions of real-world violence such as assaults, armed conflict, and terrorist events, even when platforms implement age restrictions or moderation systems (Eltaher et al., 2025; Xue et al., 2025). Such safeguards are often ineffective in practice, since underage users can bypass age verification easily and, as a result, are exposed to comparable levels of harmful content as adults. Alarmingly, exposure to harmful digital content is not limited to intentional searches. Algorithm-driven recommendation systems on YouTube can rapidly direct children toward violent content, even when they initially engage with benign material (Papadamou et al., 2020). Similarly, on Twitch, there is content targeting specifically young audiences that contains inappropriate material, such as violence, substance abuse, and highly sexualized content (Gonçalves et al., 2026; Silva et al., 2025). This phenomenon is going unnoticed by parents; approximately 70% of the parents were unaware of their child's exposure to violence on the mobile device (Liu et al., 2021). International organizations and law enforcement authorities have raised growing concerns

about children's exposure to violent online content. For example, UNICEF (2025) has warned that the graphic and violent material online may negatively affect children's well-being and safety, and that repeated exposure can contribute to the normalization of such behaviour. Similarly, the UK police have developed the "Let's guide, not guard, our online spaces" campaign to address the issue of violent and distressing content being shared online (Staffordshire Police, n.d.).

Video games play a very big role in exposing children to violent content. As of 2023, the global gaming community is estimated at 2.7 billion individuals, approximately 25% of the world's population (Statista, 2023). Almost 80% of those younger than 18 years old are gamers, spending 30% of their entertainment time gaming (Bain & Company, 2024). Considering the large number of underage people who are using video games, data highlight the alarming fact that up to 85 % of video games contain some type of violence, and on top of that, 70% of US teenagers who identify themselves as "gamers" report violent content in at least some of the video games they play (American Psychological Association [APA], 2020; Gottfried & Sidoti, 2024). In addition to the violent content present in video games, another relevant issue is that over the past decades, video games have evolved from simple pixelated graphics into highly immersive, lifelike experiences. Especially during the COVID lockdowns, the gamers' preference for immersive video games increased and has been increasing ever since. This development reflects more than simply having additional time available during the pandemic: immersive gaming environments increasingly function as social and entertainment spaces in which players can interact with others and participate in content creation. (Bain & Company, 2024).

These converging developments, the gamification and decentralisation of extremist content, the escalatory logic of self-harm and violence

networks, and children's pervasive and often unsupervised exposure to violent material, are the landscapes ATKM now must read.

What Violent Content Does: Mechanisms and the Developing Brain

Psychological Effects

Repeated exposure to violent digital content may influence how children and adolescents emotionally and cognitively process violence. The main mechanisms identified in the literature are desensitisation, reduced empathy, moral disengagement, and aggressive priming.

Desensitisation refers to reduced emotional and physiological responsiveness to violence. Experimental studies show that violent-game exposure can reduce heart-rate and skin-conductance responses to later real-life violence (Carnagey et al., 2007), while frequent violent-game players have shown weaker neural responses to violent images and reduced activation in regions involved in emotional processing (Gentile et al., 2016; Jabr et al., 2018). Repeated exposure may therefore make violence feel less disturbing and increasingly normal.

A related effect concerns empathy. Greater exposure to violent media has been associated with lower empathic and prosocial responding, with some longitudinal evidence suggesting decreases over time (Brockmyer, 2022; Funk, 2005; Mößle et al., 2014; Teng et al., 2019). Guo et al. (2013), for example, found that short-term exposure to violent videos reduced neural responses to others' pain and produced an approximately 29% reduction in perceived pain ratings. These findings suggest attenuation rather than a complete or permanent loss of empathy.

Violent content may also promote moral disengagement, whereby harmful behaviour is justified or reframed as acceptable (Hartmann et al., 2014). Games often portray aggression as necessary, rewarding, or morally justified while minimising consequences for victims. Longitudinal evidence indicates that violent-game exposure can predict increased moral

disengagement, which is in turn associated with higher aggression (Teng et al., 2019).

Finally, violent content may temporarily increase physiological arousal and make aggressive thoughts and interpretations more accessible (Anderson et al., 2004; Gentile et al., 2017). This can occur among both players and observers (Zhang et al., 2021) and may contribute to more permissive attitudes towards violence (Sánchez-García & Usán Supervía, 2025). These mechanisms suggest that exposure may shape how violence is emotionally experienced, cognitively interpreted, and morally evaluated, without implying an inevitable progression towards violent behaviour.

Behavioural Effects

Behavioural studies similarly associate violent-media exposure with aggression, although findings are not uniform. Zhang et al. (2021) found higher aggressive behaviour among children who actively played violent games than among those who merely observed them, although observers of violent gameplay also behaved more aggressively than those exposed to non-violent material. The way violence is framed also matters: aggression may increase when violent acts are presented as justified or necessary within a game narrative (Chen et al., 2023).

Survey studies report associations between violent gaming and physical aggression, anger, and antisocial behaviour (Medeiros et al., 2020). Longitudinal evidence also suggests that some effects may persist. Prescott et al. (2018), in a meta-analysis involving more than 17,000 participants, found a modest association between violent-game exposure and later physical aggression. Similarly, Ybarra et al. (2022) found that childhood exposure to violent media was associated with a greater likelihood of serious violence during adolescence or young adulthood, even after controlling for other risk factors.

However, negative effects are not consistently found. Kühn et al. (2019), for example, observed no significant changes in aggression, empathy, or impulsivity after two months of daily violent-game play. Other research has linked violent gaming to broader outcomes including social contagion of aggression (Delhove & Greitemeyer, 2023), substance-use problems (Fahs et al., 2022), self-harm and poorer mental health (Cosquer et al., 2025), and poorer sleep, academic engagement, and family interaction (Charmaraman et al., 2020; Felemban et al., 2021). Many of these findings are correlational, meaning that causal conclusions cannot be drawn. Overall, violent content may be associated with increased aggression and other maladaptive outcomes, but these effects are generally modest and differ substantially across individuals and contexts.

Magnitude and Limitations of the Evidence

The magnitude and real-world relevance of violent-media effects remain contested. Anderson and Bushman (2001) reported an association of approximately $r = .19$ between violent-media exposure and aggression, whereas Ferguson (2015) reported a much smaller association of approximately $r = .06$ among community youth after controlling for relevant confounders.

An important limitation is how aggression is measured. Experimental studies often use laboratory proxies such as allowing participants to administer an unpleasant noise blast to another person. Such measures can capture short-term aggressive responding but have limited ecological validity: choosing a louder noise blast is clearly not equivalent to committing serious violence or preparing a terrorist attack. Many experiments also examine only immediate responses following brief exposure and therefore cannot demonstrate that effects persist over time.

Correlational research presents the opposite problem: associations between violent-media exposure and aggression cannot determine causality. Individuals who are already more aggressive, impulsive, or sensation-seeking may also be more likely to seek violent content. Other influences, including competition, frustration, excessive gaming, sleep disruption, and social reinforcement, are also difficult to separate from the effects of violence itself.

Finally, much of this literature concerns fictional violent games and entertainment media rather than terrorist propaganda, extremist communities, or real-world atrocity footage. The evidence therefore supports a cautious conclusion: violent content can produce measurable psychological and behavioural effects, but exposure alone has limited value for predicting serious real-world violence or mobilisation.

Risk and Protective Factors

Importantly, violent-content exposure does not affect all individuals in the same way. A risk and resilience perspective views exposure as one influence among several rather than as an independent cause of aggression (Gentile & Bushman, 2012). Its effects may become stronger when other vulnerabilities are present and weaker when protective factors buffer them.

The family environment is particularly relevant. Family conflict can strengthen the relationship between violent-media exposure and aggression, producing a "double-dose" effect in adolescents exposed to both high conflict and violent media (Fikkers et al., 2013). More generally, adverse environmental circumstances can amplify the association between violent content and aggression (Addo et al., 2021). Conversely, parental involvement can be protective. Parental restriction of violent-game content, particularly when combined with warmth and supervision, has been associated with reductions in later fighting behaviour (Cote et al., 2020).

Supportive parental mediation may similarly reduce exposure and aggression, whereas inconsistent mediation appears less effective (Fikkers et al., 2017).

Peer norms also matter. Violent-media exposure appears more strongly related to aggression when adolescents perceive aggression as common or acceptable among their peers, whereas the relationship can be substantially weaker in predominantly prosocial environments (Fikkers et al., 2016). At the individual level, impulsivity, anger, emotional reactivity, and low frustration tolerance may increase susceptibility to aggressive responses following exposure (Devilly et al., 2021). Patterns of media use are also relevant, as excessive gaming or gaming that interferes with everyday functioning has been associated with more favourable attitudes towards violence and greater behavioural risk (Sánchez-García & Usán Supervía, 2025). These findings reinforce that exposure is best understood as a probabilistic risk factor whose effects depend on the wider individual and social context, rather than as a direct pathway to aggression.

The Developing Adolescent Brain

Children and adolescents may be particularly susceptible because their brains are still undergoing substantial structural and functional development (Blakemore, 2012). Importantly, this development is uneven. Systems involved in emotion, motivation, and reward mature relatively early, while the prefrontal cortex, which supports planning, impulse control, and decision-making, develops more gradually into early adulthood (Casey et al., 2008; Menon & D'Esposito, 2022).

This developmental imbalance can make adolescents particularly sensitive to emotionally intense experiences, immediate rewards, novelty, social approval, and peer influence. At the same time, the frontal regions of the brain continue to mature. Around ages 11–12, grey matter, which

contains many of the brain cells involved in processing information, reaches a peak and is then gradually refined as less-used connections are removed. Meanwhile, white matter, which helps different brain regions communicate with one another, continues to develop (Hummer, 2015). These processes reflect heightened neuroplasticity, meaning that adolescent neural systems remain particularly responsive to environmental experiences. The period of greatest sensitivity appears to fall broadly between approximately 10 and 17 years, although prefrontal development continues into the mid-twenties. Consistent with this, the association between violent-game exposure and aggression appears strongest during early adolescence (Burkhardt & Lenhard, 2022). Neuroimaging studies suggest that violent gameplay can temporarily affect cognitive-control and emotional-processing systems. For example, adolescents have shown reduced activity in a frontal brain region involved in controlling impulses and suppressing automatic responses after playing violent video games (Hummer et al., 2010), while other studies have reported weaker emotional and empathic neural responses.

However, evidence for permanent neural effects remains inconsistent. Longitudinal and experimental studies have not consistently demonstrated lasting reductions in empathy-related neural activity (Goodson et al., 2021; Kühn et al., 2019). Repeated exposure may instead influence more specific processes, such as how violence is interpreted or judged as justified (Adebimpe et al., 2019).

For ATKM, the key conclusion is therefore not that exposure predicts violence, but that repeated violent content may lower emotional, cognitive, or moral barriers in some individuals, particularly during adolescence. Exposure or sharing should consequently be treated as one probabilistic signal rather than a stand-alone predictor of mobilisation.

Factors Influencing Susceptibility to Extremist Digital Environments

The following themes explore how individual vulnerabilities interact with gamified extremist digital spaces. It is important to clarify that these factors are not a definitive checklist or a linear progression. Instead, they represent a constellation of conditions that may increase some individuals' receptivity to extremist narratives (Schlegel, 2021). Some individuals may face all these conditions and never move toward violence, while others may be moved to action by a single factor or by a combination of them (LaCroix & Berkovits, 2026). As such, these themes should be viewed as a set of considerations for potential risk, acknowledging that moving from online engagement to real-world behaviour is an individualised and non-deterministic process.

Psychological Vulnerability and Susceptibility to Radicalisation

One vulnerability is pre-existing psychological strain. For example, research by Kenyon et al. (2022a) indicates that individuals who radicalise primarily online are more likely to have a history of mental illness or personality disorder compared to those radicalising through hybrid or offline social networks. Specifically, the online-only group was 6.27 times more likely to have these factors present than the offline group. However, this relationship should be interpreted as correlational rather than causal. Supporting this cautious interpretation, Kenyon et al. (2022b) also found that online-only individuals were often professionally assessed as having the lowest levels of intent and capability for serious physical violence compared to their 'hybrid' or 'offline' counterparts. Hence, psychological vulnerabilities may increase susceptibility to online radicalising narratives but do not necessarily translate to intent or capability for violence.

Independent of psychological strain, psychological openness to radical narratives may also emerge during periods of identity disruption. Neo et al. (2016) argue that when existing beliefs or future expectations are

undermined, individuals may experiment with alternate ideas with lower perceived risk. During such phases, they may seek narratives that offer a sense of shared identity to alleviate personal grievances. Neo et al. (2016) also note that the anonymity of the online environment may lower psychological barriers, which allows individuals to experiment with different identities to fulfil unmet social identification needs. These theoretical patterns are supported by qualitative empirical evidence. For instance, Koehler et al. (2023) found that among children radicalised on gaming platforms, offline stressors such as family conflict or bullying may co-occur with and interact with online experiences to facilitate close social-emotional bonds with extremist peers. Thus, it could be inferred that extremist groups may provide a sense of stability and belonging during periods of uncertainty. Nevertheless, the specific mechanisms of these interactions remain an area requiring further empirical investigation.

Social Isolation as an Amplifier of Online Risk

Social isolation is another factor that may be associated with susceptibility to extremist narratives, as individuals may rely on online communities for social interaction. This is particularly pronounced among individuals who radicalise mainly online, which is a relatively small group that often shows distinct social patterns. A study of individuals in England and Wales convicted of extremist-related offences found that 85% of the 'internet' group were lone actors who lacked connections to offline extremist networks. These individuals were primarily convicted for online-only activities, such as disseminating extremist materials or encouraging terrorist acts, rather than participating in physical violence. This suggests that socially isolated individuals may turn to online forums and gaming communities to fulfil unmet needs for connection in their offline lives. Empirical evidence further demonstrates that, unlike most digital users who

interact with existing offline acquaintances, these individuals often engage predominantly with online contacts (Lankford et al., 2024). In doing so, they may substitute real-world relationships with digital ones. As a result, online spaces may become the primary setting for social interaction.

A related aspect of social isolation is the absence of offline moderating influences. Theoretical frameworks suggest that family, peers, or colleagues can act as protective factors. They do this by providing informal correction when individuals express extreme views (Neo et al., 2016). Research indicates that when these social bonds are weak, individuals may be more likely to join homogeneous online communities where radical interpretations often go unchallenged (Hassan et al., 2018). In these contexts, the need for belonging also may be redirected toward digital spaces that offer immediate validation (Neo et al., 2016). Research synthesised by Hassan et al. (2018) suggests a conceptual link between these processes. The absence of corrective feedback, combined with frequent validation in isolated online environments, may normalise extremist narratives. Although the empirical evidence for this process remains preliminary and requires further longitudinal validation (Hassan et al., 2018), it is theorised that for certain isolated individuals, the digital communities may function as key moral and social reference points. Where offline correction is largely absent, the practical implication is that corrective input has to be introduced into the online environment itself. This is the reasoning behind the counter-messaging, redirection, and credible-messenger recommendations set out later in this report.

The Gamification of Radical Narrative Participation

The psychological needs for recognition, belonging, and connection may be activated and reinforced in gamified environments. Schlegel (2021) argues that these desires are transformed into digital rewards such as

rankings and progress markers. These features utilise positive reinforcement to motivate ongoing engagement (Schlegel, 2021). Schlegel (2021) further proposes that this systematic application of game design elements may shift a user's role from a passive recipient of propaganda to an active participant in a shared digital experience. In this sense, individuals may begin to experience themselves as agents or 'players' within a radical narrative (Lakhani & Wiedlitzka, 2023). According to Lakhani & Wiedlitzka (2023), this shift may be reflected in how extremist events are interpreted by online audiences. In their analysis of the Christchurch attack, users employed gaming metaphors such as 'kill count' and 'leader board' when discussing the assailant's action. This indicates that ideological participation may be reframed as a form of play or achievement, potentially incentivising more sustained engagement with extremist material. Furthermore, Schlegel (2021) theorises that these game elements appeal to specific theoretical user profiles. For instance, 'Achievers' are motivated by leaderboards and the urge to outperform others, whereas 'Socialisers' are drawn to features that facilitate connection with others or foster a sense of community within the ecosystem. Taken together, this suggests that gamified environments may shape engagement. They do so by transforming ideological participation into reward-based systems aligned with psychological needs.

Another mechanism that may also drive engagement in these spaces is the provision of social recognition through gamified rewards. Online spaces reward participation through visible signals such as likes, reposts, and symbolic status markers, which function as social currency within these competitive environments (Lamphere-Englund, 2024). Schlegel (2021) suggests that for certain users, such engagement may evolve into a strategic pursuit of status that may eventually overshadow pure ideological motivation. For this reason, individuals may become incentivised to maintain or improve their position within a perceived digital hierarchy, potentially

leading them to adopt extreme positions to remain relevant or visible. While the current evidence for these dynamics is largely anecdotal and lacks empirical basis, Schlegel's (2021) work offers a preliminary theoretical framework for understanding how the pursuit of digital prestige structures and intensifies radicalisation.

Within this performance-oriented environment, irony and transgressive humour may serve as identity signals. Lakhani (2024) emphasises that 'shitposting' and memes are used as tools to identify 'internet insiders' and declare group membership. These features may allow users to signal alignment with extremist subcultures without the immediate social or legal risks of a formal commitment. Lakhani (2024) further theorises that this deliberate blending of extremist narratives with gaming aesthetics and internet pop culture may serve to neutralise moral guilt. By framing hate speech as 'edgy' entertainment, violent narratives may appear more palatable and relatable to younger audiences who have grown up with digital media. Additionally, Schlegel (2021) suggests that gamified engagement may, in theory, contribute to the normalisation of explicit hostility by rewarding 'trolling' as a form of social status. Although empirical evidence for these cognitive shifts remains limited, it is hypothesised that these performative jokes could evolve into shared cultural scripts (Lakhani, 2024; Schlegel, 2021). These scripts may blur the boundary between irony and endorsement, potentially lowering the psychological barriers to more sincere expressions of extremist support (Lakhani, 2024; Schlegel, 2021).

Identity Fusion and Potential Trajectories Toward Action

As identification with online communities intensifies, it may develop into what researchers term 'identity fusion'. This theoretical state occurs when personal and group identities become psychologically inseparable, fostering a sense of 'shared fate' and familial-like bonds with online peers (Swann et al., 2012; Wolfowicz et al., 2021). Unlike simple group membership, identity

fusion is characterised by a 'visceral' feeling of oneness, in which the psychological boundaries between the individual and the collective blur (Swann et al., 2012; Wolfowicz et al., 2021). It is theorised that gamified environments may facilitate this process by transforming abstract social bonds into visible and quantifiable reward systems. Features such as rankings, badges, and competitive recognition validate an individual's standing within the group. This embeds personal identity within broader narratives of competition where digital prestige is tied to demonstrating group commitment (Lamphere-Englund, 2024; Schlegel, 2021). For some, these digital ecosystems may become meaningful social environments rather than just virtual spaces (Lamphere-Englund, 2024). This suggests that gamified incentives may encourage individuals to adopt more extreme positions to preserve status, particularly under perceived existential threat. However, the evidence base remains limited and relies heavily on self-reported data (Wolfowicz et al., 2021). Therefore, identity fusion may be a risk factor that lowers psychological barriers to commitment. This may be especially true when gamified incentives reinforce group identity and status.

Similarly, these dynamics may be further intensified by the emergence of 'onlife' environments. These are hybrid environments that seamlessly incorporate and integrate elements from an individual's online and offline experiences (Valentini et al., 2020). Within such environments, the social-emotional bonds formed through shared gaming or chat experiences may pressure individuals to align their behaviour with group expectations (Koehler et al., 2023). Koehler et al. (2023) demonstrate that the transition from digital interaction to physical action may, in some cases, be catalysed by influential peers acting as 'mentors'. These individuals may guide others from symbolic online participation toward offline criminal acts. In this context, performative extremist displays may acquire real-world significance. They may function as a means of maintaining status within a valued social

environment that is experienced as cohesive and stable. Generally, ‘onlife’ environments may enable the transition from online involvement to offline action. Still, these links between online and offline factors are very individual and require further empirical validation.

Sustained exposure to the gamified framing of violence may contribute to a gradual erosion of moral restraint. Research into extremist online subcultures demonstrates that the reduction of individuals to ‘targets’ or ‘objectives’ may be reinforced by ‘spectators’ who treat violence as a competitive sport, explicitly encouraging perpetrators to surpass previous ‘high scores’ (Lakhani & Wiedlitzka, 2023). Within this framework, Lankford et al. (2024) suggest that when extremist acts are interpreted through a task-oriented lens, attention may shift away from the reality of human harm toward the technical aspects of execution. Drawing on these observations, it may be argued that prioritising technical performance over human consequences could lower the psychological threshold for violence among susceptible individuals. This may, therefore, facilitate a transition from online engagement to real-world acts. However, this process is not a deterministic pathway. It represents one potential risk factor that remains highly individual-dependent and necessitates further study.

Overall, psychological factors may increase receptivity to extremist narratives. Such vulnerabilities may create psychological openings that gamified digital environments leverage through reward-based systems and online social bonds, which foster a sense of shared identity and group commitment. The shift from online engagement to physical action is variable and depends on individual traits intersecting with specific digital and offline triggers. These findings demonstrate the importance for stakeholders across the wider counterterrorism and counter-extremism ecosystem of interpreting online engagement within its broader behavioural and social context. Participation in gamified extremist environments may fulfil psychological and

social functions that are distinct from genuine movement toward violent mobilisation. Consequently, indicators of individual vulnerability or symbolic online participation should not automatically elevate threat assessments in the absence of explicit indicators of action-oriented behaviour.

Observable Indicators of Mobilisation

While the internal processes of radicalisation may be difficult to observe directly, changes in online engagement may be reflected in observable behavioural patterns. These indicators may represent a move from symbolic ideological expression to functional readiness. The following themes outline the potential digital traces associated with increasing risk. Here, it is also essential to acknowledge that while these patterns are identifiable, they are non-linear and require careful contextual interpretation.

Linguistic Shifts

The way individuals use language online may provide insight into possible violent intentions. Theoretical frameworks on 'leakage' suggest that individuals who are moving toward violent action may, in some cases, leave digital traces that reflect ideological fixation or preoccupation with violence (Brown et al., 2024). However, Brown et al. (2024) demonstrate empirically that the mere expression of ideological grievances or hateful rhetoric is not a reliable predictor of actual mobilisation. Instead, their data-driven analysis shows that a key indicator may be discussions related to operational planning, logistics, and the acquisition of technical 'know-how.' These behavioural shifts may be associated with paralinguistic changes. These include increases in word count and higher punctuation frequency, which may mirror higher emotional intensity in communication (Brown et al., 2024). Thus, these findings suggest that ideological expression alone is not a key marker of risk. Certain linguistic features, such as operational

language and paralinguistic changes, may be more informative of differing levels of engagement.

Operational Readiness and Tactical Rehearsal

This pursuit of practical skills and tactical information may also serve as an indicator of intent toward action. This may be reflected by 'tactical learning', which involves the gathering of information on weapons, targets, and methods of attack (Schumann et al., 2025). Schumann et al. (2025) argue that such behaviours may be more indicative of mobilisation than ideological commitment alone, given that levels of extremist belief do not differ between those who mobilise and those who do not. Related to this, Kenyon et al. (2022a) introduce the concept of 'virtual rehearsal' in which individuals use online environments as spaces to simulate or practise aspects of intended behaviour. Kenyon et al. (2022a) also observed that for individuals on an online-only pathway, the internet is the primary operational environment, where the boundary between digital preparation and the real world may become blurred. This may represent a shift to active forms of engagement characterised by individuals 'practising' their extremist identity or potential methods within digital spaces. However, Kenyon et al. (2022a) note that even if the virtual world provides a low-risk environment for rehearsal, it often fails to translate into real-world lethality. Nonetheless, given the rapid evolution of digital environments and recent technological developments, it remains unclear whether this pattern may be changing. Kenyon et al. (2022b) further state that individuals who rehearse their intent in virtual spaces are more likely to leave digital traces or signal intent online compared to those influenced by offline networks. Yet, they remain the group least likely to progress beyond the planning stage. Hence, although the internet provides access to tactical information and supports forms of virtual rehearsal, high levels of online signalling do not necessarily equate to successful implementation of plans.

Changes in how individuals engage with others online may also signal deeper involvement in extremist spaces. Individuals may begin to engage with recruiters and participate in gaming-adjacent ecosystems such as Discord servers or niche forums that reinforce shared identity, extremist narratives, and refine tactical understanding (Lamphere-Englund, 2024). This reflects a broader shift in communication patterns, where activity may move from open platforms to encrypted applications and private channels that facilitate more secure interaction and coordination (Kenyon et al., 2022a). Nevertheless, such relocation may indicate increasing commitment and a desire for operational security, as opposed to indicating capability itself (Kenyon et al., 2022a). Thus, movement towards encrypted communication for logistical planning may be a worthy indicator of deepening involvement, but not necessarily actual capability.

Technical Utility and Performative Signalling

Indicators of potential mobilisation may also be found in the functional intent and interactive behaviours surrounding digital content, rather than its mere aesthetic resemblance to video games. While some researchers suggest that first-person propaganda is designed to imitate first-person shooter games, Andrews (2024) argues that the use of helmet-mounted cameras is primarily a tactical choice. This technology provides an unobstructed 'boots-on-the-ground' perspective that ensures a perpetrator's actions are captured with maximum sensory immersion for online audiences. This may function as an indicator of an intent to distribute the attack to online audiences. In place of mimicking games, Andrews (2024) suggests such footage aligns more closely with cinematic techniques like the 'killer's perspective', which emphasises the power of the perpetrator and vulnerability of the victim. Furthermore, potential mobilisation may also be reflected in 'bottom-up' gamification, where audiences interact with violent

footage by assigning scores, referencing leaderboards, and using metaphors such as ‘kill counts’ (Andrews, 2024). These interactive behaviours may signal a shift to the framing of violence as a competitive social performance. Here, a risk indicator may be the deliberate technical choices made to broadcast an attack and the participatory frameworks that reward real-world action.

In some cases, indicators of potential mobilisation may include the production of materials that formalise intent and address a broader audience. This may involve disseminating manifestos and strategic advice to provide a ‘blueprint’ for future actors (Lankford et al., 2024). Lankford et al. (2024) specifically highlight the creation of ‘legacy tokens’ such as livestreams or references to ‘high-score’ fatality benchmarks, which may be intended to inspire imitation through competitive structures. As such, these artefacts may function as a final psychological commitment for the perpetrator and a transnational signal to aligned audiences (Koehler et al., 2023). While these ‘final act’ materials may indicate elevated risk, researchers caution that these signals are difficult to identify in real time. Thus, these materials may act as late-stage indicators of mobilisation, reflecting the consolidation of intent and attempts to engage with wider audiences. Nonetheless, they remain ambiguous and challenging to evaluate reliably in real time.

In essence, observable patterns of online behaviour may reflect a shift from symbolic engagement toward forms of functional involvement, particularly when activity centres on operational, tactical, or performative aspects of violence. However, these patterns remain inherently insufficient on their own and do not constitute direct indicators of intent or capability. Their relevance depends on how they are interpreted in relation to underlying vulnerabilities and the broader situational context. For stakeholders with access to broader behavioural and investigative

information, developments such as increasingly operational language, growing tactical interest, movement toward private communication channels, or the production of action-oriented materials may provide useful insight into potential mobilisation when assessed collectively and over time. For organisations involved in the assessment of terrorist content, these features may support the identification and prioritisation of content requiring further review, referral, or other appropriate action.

Protective Factors Against Mobilisation

Beyond indicators of escalation, the literature identifies a set of protective factors that may reduce the likelihood that individuals transition from extremist engagement to real-world violence. These protective factors operate across multiple levels. These include interpersonal relationships, cognitive processes, patterns of online engagement, and structural constraints that interrupt pathways to mobilisation.

Interpersonal and Social Integration Factors

At the interpersonal level, stable offline relationships are conceptualised as protective factors that mitigate proclivity to radicalisation and mobilisation. Regular, high-quality interaction with family, peers, and coworkers can provide alternative perspectives that may help counterbalance extremist narratives. This is supported by empirical observations of Lankford et al. (2024), who found that perpetrators of gamified violence operate in a state of profound social isolation in the physical world. In such cases, weak relationships may push individuals toward virtual relationships in extremist online subcultures. However, this substitute may fail to meet underlying social needs. Based on the investigation into the 2019 Christchurch attack, Lankford et al. (2024) note that the absence of a structured social environment can remove daily

interactions that might 'soften the hard edges' of extremist thinking. Thus, these findings imply that the strength of interpersonal relationships may be an essential factor in whether an individual at risk remains embedded in their physical community or drifts toward gamified online environments that may facilitate violence.

Beyond individual relationships, social integration into formal institutions is theorised to reduce susceptibility to radicalisation. These institutions may range from school-based extracurricular activities to broader civic organisations such as sports teams and community groups. Empirical evidence from a systematic review by Wolfowicz et al. (2021) identifies school bonding and institutional trust as protective factors against radical attitudes. The authors also state that although these factors are statistically significant, the meta-analysis reveals that observed effect sizes are small to very small. Conceptually, this suggests that these forms of integration provide a sense of efficacy and a shared identity that counters the 'them-and-us' binary characteristic of extremist ideologies. Such institutional involvement may therefore promote stabilising social ties by embedding individuals within structured social environments that provide alternative sources of belonging, identity, and purpose.

Cognitive Resilience

At the cognitive level, internal resilience may function as a psychological filter that moderates the persuasive impact of extremist content. Empirical research by Gallacher et al. (2021) supports this mechanism. The authors found that online interactions characterised by higher levels of integrative complexity are associated with a lower likelihood of subsequent radical behaviours. Integrative complexity is the ability to acknowledge and integrate multiple perspectives, which may reduce susceptibility to simplistic extremist narratives (Gallacher et al., 2021). This

suggests that cognitive flexibility allows individuals to resist the 'black-and-white' binaries inherent in radical narratives by maintaining more nuanced social identities. Supporting this, Hassan et al. (2018) argue that the relationship between exposure to extremist content and radicalisation outcomes is mediated by individual and social factors that shape interpretations. This implies that cognitive flexibility may enable individuals to engage critically with extremist messaging, rather than internalise it. This provides a theoretical basis for interventions such as media literacy, which aim to strengthen the ability to evaluate and resist manipulative content. On the other hand, Gallacher et al. (2021) caution that online engagement with opposing views may not be inherently protective. A high volume of interactions with low complexity may increase risk. Therefore, this suggests that cognitive flexibility is protective when engagement involves meaningful interpretive depth rather than repetitive or superficial exposure.

Outlet Functions and Motivations

Importantly, the majority of online extremist engagement does not appear to indicate an immediate risk of mobilisation. For some individuals, digital spaces may function as social outlets rather than gateways to violence. For instance, Scrivens et al. (2026) found that non-violent extremists often maintain a higher volume of posting activity than those who eventually mobilise. Similarly, Brown et al. (2024) demonstrated that the expression of ideological or hateful content was less common among individuals who later mobilised. Together, these findings suggest that sustained online engagement may allow individuals to perform and reinforce a radical identity, thereby providing a relatively stable outlet for ideological expression without necessitating offline mobilisation. However, this stabilising function may weaken as patterns of online engagement change. Scrivens et al. (2026) also note that individuals who later mobilise may

exhibit reductions in public posting activity prior to offline action. Schumann et al. (2025) suggest that such a decline may reflect a shift from social participation toward more passive or instrumental internet use, such as the acquisition of tactical information. Hence, high levels of online posting may function as a stable outlet for expressing radical views, allowing identity performance without necessarily leading to offline action. Nevertheless, changes in engagement patterns may also suggest a potential shift toward more instrumental forms of internet use associated with mobilisation. This implies that online extremist participation can serve both stabilising and risk-enhancing functions depending on how engagement evolves over time.

A similar pattern appears in digital gaming environments, where different motivational profiles are associated with varying levels of radicalisation risk. For example, Greipl et al. (2025) found that when gameplay is driven primarily by social interaction, there is less acceptance of political violence and lower levels of xenophobia. The author suggests that 'Socialiser' profiles keep diverse social connections, which may help prevent narrow thinking associated with radicalisation. By comparison, profiles such as the 'Absorber' are characterised by strong achievement motives and a pronounced desire for social status, which have a higher risk of accepting extremist ideas and political violence. These findings indicate that risk is not inherent to gaming itself, but rather stems from the motivational profiles and psychological needs behind engagement.

Structural and Platform-Level Constraints

At the structural level, institutional measures may function as protective factors by disrupting the digital ecosystems that enable extremist coordination (Neo et al., 2016). Recent analysis by Lamphere-Englund (2024) emphasises that extremists exploit the interconnectivity between gaming, messaging, and livestreaming platforms. From this perspective, structural interventions may be most effective when they interrupt the social

feedback loops that reward and amplify radical identities. A practical example of such measures can be seen in the platform's Trust and Safety teams. As Lamphere-Englund (2024) outlines, these teams develop behavioural frameworks to distinguish between routine platform use and patterns of engagement indicative of extremist exploitation. Through identifying and disrupting these cross-platform pathways, these mechanisms may reduce opportunities for performative violence and the accumulation of status through extremist visibility. This sustained structural pressure may therefore reduce the incentives that support the maintenance and visibility of radical identities.

Collectively, these findings suggest that protective factors shape how individuals interpret and respond to extremist content. Strong relationships, cognitive resilience, and institutional safeguards can provide alternative sources of identity, meaning, and support, thereby reducing dependence on extremist environments as primary sources of meaning and belonging. Moreover, the motivations underlying digital participation, such as whether engagement is driven by social interaction, identity performance or status seeking, also influence how such environments are experienced. As such, these protective factors may interrupt, redirect, or stabilise trajectories that might otherwise progress toward mobilisation. For stakeholders with access to broader behavioural, social, and psychological information, such factors can be incorporated alongside patterns of extremist engagement to provide a more contextualised assessment of mobilisation risk.

The Operational Response: A Risk-Tiered Framework

Not all terrorist content carries the same potential to draw others further into extremist material. A shock video reposted by a teenager to impress classmates and a recruitment appeal produced by an organised network may both meet the legal threshold, and both fall under the same

one-hour removal clock, but they do not carry the same risk of contributing to radicalisation. ATKM's statutory task is to determine whether content is terrorist content, not to assess the person who posted it. Within that task, however, ATKM still has to decide which referrals to assess first when the volume exceeds the capacity to assess everything at once. The framework set out below is intended to support that prioritisation of content. It is not a means of classifying individuals, and nothing in it should be read as an assessment of a person.

Prioritisation of this kind sits within ATKM's statutory task rather than beside it. Article 2 of the Dutch Implementation Act for the Terrorist Content Online Regulation (Uitvoeringswet verordening terroristische online-inhoud) establishes the Authority and designates it as the competent authority under Article 12(1)(a) to (d) of Regulation (EU) 2021/784. That same article gives the Authority two tasks in relation to terrorist content: to identify such content and to promote and where necessary enforce its inaccessibility, and, separately, to conduct research into and provide information about the presence of online terrorist material in order to limit its distribution among the public, where possible in cooperation with public and private parties (Uitvoeringswet verordening terroristische online-inhoud, 2023, art. 2). The framework proposed here is offered in service of that second task. It is set out as a step-by-step decision sequence in Appendix A, which should be read alongside this section.

Four Types of Content-Sharing Behaviour

Content-sharing behaviour visible in referred material tends to fall into four recurring patterns. These describe what the content itself shows and how it is presented. They are not conclusions about the person behind an account, which ATKM has neither the mandate nor the data to reach, and they are heuristics for prioritisation rather than diagnostic categories. Material from the same source may fit different patterns at different times.

Two features of the material separate the four, and both can be read from the content itself. The first is whether a coherent extremist ideology is expressed. The second is whether the content shows operational capability, meaning tactical detail, logistical language, or evidence of preparation. Type 1 shows both. Type 2 shows capability and status-seeking violence without a consistent ideology. Types 3 and 4 show neither, and are separated from one another by whether the material was shared deliberately for effect or passed on without apparent intent.

Type 1. Ideological-operational content. Coherent extremist framing, repeated rather than one-off posting, and explicit alignment with violent perpetrators or movements, together with at least some operational or tactical material. Binder and Kenyon (2022), reporting data from Kenyon et al. (2022a), found that 48% of individuals on a mixed online-offline radicalisation pathway were rated high on intent, compared with 36% of those radicalised offline and 15% of those radicalised online only. Engagement followed the same ordering, at 67%, 50% and 32% respectively. Capability did not: the highest level was reached by 41% of the offline group, 22% of the mixed group and 5% of the online-only group. Intent and capability therefore do not track each other across pathways, which is part of why capability is treated here as a separate dimension rather than folded into intent.

Type 2. Status-driven network content. Draws on several ideologies at once without committing to any of them. Com and 764 are the networks most often cited here. They are loosely organised online groups in which participants combine accelerationist, neo-Nazi, satanist, and nihilistic references rather than following a single doctrine, and in which standing within the group is earned through the extremity of the violence or self-harm displayed rather than through ideological consistency. This is what the term "salad bar" ideology describes: adherents select elements from different

extremist doctrines in the way a diner selects items from a buffet, rather than following one creed (The Soufan Center, 2021). Content of this kind is better recognised by violence-as-status mechanics, coded in-group language, and accelerationist rhetoric than by any single coherent doctrine, and Hemmila and Perliger (2025) tested the salad bar and hybridisation hypotheses directly, on the online discourse of American violent far-right groups, and found some initial overlap between those groups while concluding that it is too early to assume ideological hybridisation across the far right. The label is therefore used here descriptively, to characterise how this material presents, rather than as a settled empirical finding.

Type 3. Performative shock content. Shock material circulated within ordinary friend networks rather than extremist ones, shared deliberately for social effect but with no escalation over time and no original material produced. Where the legal threshold is met the content is still removed, but it carries a weaker signal of onward radicalisation risk.

Type 4. Incidentally redistributed content. Material passed on through algorithmic exposure, quotation, or misunderstanding, with no indication of intent or awareness in the content itself.

Types 3 and 4 rest on a thinner evidence base than Types 1 and 2, and the absence of visible escalation within a single referral says little about the material that surrounds it. Assigning content to either of them should therefore be treated as a provisional prioritisation decision, open to revision if further material from the same source is referred, rather than as a finding that a case is low risk.

Distal versus Proximal Factors, and the Limits of What ATKM Can See

Distal risk factors, such as mental illness, family conflict, or social isolation, help explain why some individuals become receptive to extremist narratives, but they belong to the work of other organisations. They rest on personal and psychological information that ATKM does not hold and has no

basis to obtain, and assessing them is the responsibility of health, education, welfare, and law-enforcement partners rather than of a content authority. What is visible to ATKM are proximal signals present in the content itself.

This boundary matters for ATKM's core task because it determines what prioritisation can rest on. Signals such as operational or logistical language, tactical detail, named extremist symbols, or accelerationist framing can be read directly from a single referral, and can therefore be used to decide which items are assessed first without any inference about the person who posted them. Other indicators reported in the literature as informative, such as posting-rate trajectories or migration between platforms, cannot be read from one item, because they describe an account over time rather than a piece of content. Drawing on them would mean building up a picture of an account, and the legal basis for that kind of processing is not something this report is in a position to settle. We therefore recommend that the question be examined separately, with ATKM's legal advisers, before any longitudinal indicator of this kind is considered for operational use. The framework set out below is built first on signals observable within the referred content. Two of its four dimensions also draw on information about an account rather than about a single item, and those two are usable only where ATKM holds the relevant data access and a legal basis for it. Appendix A sets out the level of data access each indicator requires.

A Four-Dimension Risk Classification

The section that follows sets out how these content-observable signals can be organised for practical use. It begins from an existing structured risk-assessment framework, explains why that framework cannot be transferred directly to content assessment, describes the four dimensions this report proposes in its place and why one of them carries more weight than the others, and ends with the limits that apply to the resulting thresholds. Two

of the four dimensions rest on information about an account rather than about a single item of content, and are therefore available only where the corresponding data access and legal basis exist. Appendix A states which level of access each indicator needs.

Structured professional-judgement frameworks for extremism risk, including the ERG22+ dimensions of engagement, intent, and capability (Lloyd and Dean, 2015; revised as the ERG-R in Kenyon et al., 2025, which retains those three dimensions and adds a fourth covering protective and risk-mitigating factors), inform but cannot be directly imported into a content-only assessment. This report proposes a four-dimension indicator count for content assessment. It retains capability from the ERG22+ and weights it highest, and replaces engagement and intent, which describe a person rather than an item of content, with dimensions that can be read from referred material. The fourth dimension it adds is not the protective-factor dimension of the ERG-R, which serves a different purpose in offender assessment, but capability and operational behaviour as a content-observable signal. The four are: escalation, meaning rising frequency or intensity of posting and a shift from discussion toward direct threats; network, meaning contact with accounts identified as extremist, cross-posting into named enclaves, and movement toward private or encrypted channels; ideology, meaning coherent extremist framing including accelerationist rhetoric, necessary context but not on its own a reliable predictor of mobilisation; and capability and operational behaviour, meaning operational or logistical language, tactical learning, virtual rehearsal, talk of weapons or targets, manifestos and ”legacy tokens”, and deliberate broadcasting choices such as helmet cameras or livestreams.

Because operational and tactical signals separate those who act from those who do not more reliably than ideological expression alone (Brown et al., 2024; Schumann et al., 2025), the capability dimension carries the most

weight in this classification; ideology is no longer treated as one of three equal indicators, as it was in the ERG22+ as originally conceived. When three or more of the four dimensions are present, the case is classed HIGH risk; two, MEDIUM; one, LOWER; none, MINIMAL. These thresholds are starting points, not validated cut-offs: they have not been piloted against ATKM case outcomes, no inter-rater reliability data exist for their application, and at ATKM's processing volumes, even a well-performing indicator set will generate a substantial number of false positives given how rare actual mobilisation is. Calibration against real cases should precede any use of the HIGH-risk tier to trigger law-enforcement referral at scale.

Proportionate Responses by Tier

The responses set out below describe what a prioritisation outcome could lead to. They are put forward for discussion rather than as a procedure ready for use. Each response, and in particular any sharing of information with an external partner, would need to be checked with ATKM's legal advisers and with the team that carries out content assessment before it could be adopted.

HIGH-risk cases warrant removal together with law-enforcement and AIVD coordination. MEDIUM-risk cases warrant removal, a de-radicalisation referral where the legal basis for data-sharing with an external partner has been established, and continued monitoring for escalation. LOWER-risk cases warrant removal where the legal threshold for terrorist content is met, together with logging for pattern development, itself dependent on a legal basis and a retention period that would need to be established with ATKM's legal advisers. MINIMAL cases should be documented and closed, re-opened only if a pattern later emerges.

Minors require a separate procedure. For under-18 sharers, a non-criminal-justice response should be the default, even where content would otherwise meet a HIGH-risk threshold, unless specific indicators, such as

explicit recruitment activity or documented real-world violence, are present. Lewis et al. (2025) review non-criminal justice interventions for children and adolescents and found only one study eligible for their assessment of effectiveness, so this default is grounded in proportionality rather than in demonstrated outcomes. The General Data Protection Regulation does not itself supply a separate processing basis for treating minors' data differently. Article 8 governs only the conditions under which a child can consent to information society services. Recital 38 does state that children merit specific protection in relation to their personal data, and Article 3 of the United Nations Convention on the Rights of the Child requires that the best interests of the child be a primary consideration. Whether those support a differentiated procedure in ATKM's hands is a question for legal advice rather than one this report can settle.

Taken together, the data-protection questions raised above are preconditions for any longitudinal or cross-platform indicator rather than afterthoughts, and we recommend they be settled with ATKM's legal advisers before indicators of that kind are used. Once they are, the framework offers ATKM a proportionate triage tool it can pilot, rather than a rule based on content severity alone.

Policy Recommendations

Primary Prevention: Population-Level Prevention and Resilience

Primary prevention should aim to strengthen resilience before exposure to violent or extremist content develops into sustained engagement. Existing platform safeguards and age restrictions cannot fully prevent young people from encountering harmful material, particularly when recommendation systems expose users to content they did not actively seek. At the same time, parents often underestimate the extent of their children's exposure (Liu et al., 2021; eSafety Commissioner, 2023, 2024). This

suggests that prevention should not rely only on restricting access, but should also strengthen young people's ability to navigate and critically interpret digital environments.

A first recommendation is therefore an early parent- and caregiver-focused intervention. Caregivers should receive clear information about the prevalence of violent online content and the limitations of existing safeguards, together with practical guidance on parental controls, content filters, age-appropriate settings, and other safety tools. Importantly, parental involvement should extend beyond technical restrictions. Parents should also be encouraged to discuss disturbing content openly, help children distinguish between real, staged, or manipulated material, and use guided discussion to place what they see online into context. This is particularly relevant because supportive parental monitoring and communication can function as protective factors, whereas family conflict may increase vulnerability to harmful media effects.

Because the effectiveness of parental supervision depends on caregivers' time, resources, and digital literacy, a second recommendation is universal school-based media-literacy and digital-safety education. Governments and educational authorities could integrate age-appropriate digital-safety education into school curricula or provide recurring seminars and interactive sessions. These programmes should develop practical skills such as recognising harmful or manipulative content, understanding how recommendation systems shape exposure, using reporting and safety functions, and critically evaluating the credibility and purpose of online material. School-based delivery would ensure that these protective skills are available to all children rather than depending solely on the resources available within the home.

In addition to strengthening media literacy and digital safety, prevention should also build cognitive resilience by strengthening individual

cognitive filters and disrupting the initial narrowing of social reality. Here, integrative complexity training offers a feasible and non-stigmatising approach to building psychological resilience (Savage et al., 2023). This approach focuses on how individuals process information, rather than directly targeting their specific ideologies. It develops socio-cognitive skills that enable individuals to recognise multiple perspectives and resist the simplistic, binary thinking characteristic of gamified propaganda (Savage et al., 2023). Empirical evidence from pilot studies suggests that such interventions can produce significant gains in cognitive complexity (Savage et al., 2023). While longitudinal research is ongoing, current data suggest that these gains may facilitate social reintegration and reduce susceptibility to extremist narratives by providing a psychological buffer (Savage et al., 2023). This approach may also reduce the risks associated with traditional targeted interventions, such as perceived stigmatisation or community distrust, thereby improving uptake and long-term engagement (Savage et al., 2023).

Complementing these cognitive filters, the use of digital 'attitudinal inoculation' provides a proactive tool for disrupting the transition from passive exposure to active engagement. This method functions as a form of 'mental vaccination', whereby users are exposed to weakened versions of extremist arguments, followed by immediate refutations to build resistance against persuasive messaging. Research by Braddock (2022) demonstrates that this triggers psychological reactance, which is a combination of cognitive counter-arguing and affective anger. This reaction may motivate individuals to challenge extremist claims, which can lead to a reduction in the perceived credibility of radical narratives. To translate these findings into practice, hosting service providers could integrate such interventions into digital environments through targeted advertisements, landing pages, or pop-up warnings alongside existing moderation efforts. Such delivery

mechanisms would ensure that evidence-based counter-messaging is situated at the point of digital engagement, potentially reducing the likelihood that radical propaganda will influence user attitudes.

Alongside cognitive and content-based interventions, early-stage prevention also depends on informal social environments that form an individual's everyday exposure to extremist influence. Ellefsen and Sandberg (2024) indicate that effective interruptions of radicalisation occur within private social circles such as family and peer groups. These interventions are assumed to come from trusted relationships rather than state authority, allowing everyday interactions to function as natural 'off-ramps' from extremist commitment (Ellefsen & Sandberg, 2025; Jensen et al., 2023). Conversely, the presence of radicalised family members can act as a significant barrier to disengagement, which underscores the role of family dynamics either facilitating or hindering exit pathways (Jensen et al., 2023). This implies that effective early-stage prevention may benefit from strengthening informal support systems through non-securitised, community-based resources that enhance the ability of trusted individuals to provide alternative perspectives. At the same time, overly formalised or securitised interventions may risk undermining the trust upon which these informal mechanisms depend (Ellefsen & Sandberg, 2024). Altogether, this indicates the importance of strengthening informal social supports as protective mechanisms while preserving the relational trust required for their effectiveness

Secondary Prevention: Supporting Those at Risk

The implementation of interactive soft interventions can provide off-ramps by engaging users at the point of digital search activity. For instance, the Redirect Method is a strategy that uses search engine optimisation to steer individuals seeking extremist content toward counter-narratives and educational resources (Bilazarian, 2019). Empirical evidence indicates that

such campaigns achieve click-through rates comparable to commercial standards, suggesting that at-risk audiences may be receptive to content that challenges their world-view (Bilazarian, 2019). This is supported by Saltman et al. (2021), who evaluated both proactive counterspeech and search-based redirection among users who had engaged with Islamist extremist or white supremacist content. While the counterspeech intervention did not produce a statistically significant effect on violations at the aggregate level, qualitative analysis suggested more positive behavioural shifts among higher-risk subgroups, including reduced engagement with violent extremist content. The Redirect Method, meanwhile, achieved a click-through rate of around 4% above typical search advertising benchmarks and was associated with a substantial increase in referrals to disengagement services (Saltman et al., 2021). This suggests that redirecting users at an early and potentially vulnerable stage of engagement may create an opportunity for more active contact with disengagement practitioners.

However, for these strategies to achieve long-term impact, the literature suggests they may benefit from prioritising messenger credibility. Interventions that employ former extremists or peers, alongside direct interpersonal engagement, may be more effective in fostering a sense of belonging outside extremist networks (Bilazarian, 2019). Evidence indicates that former extremists may be effective because their lived experience provides a level of authenticity and authority that government or police may lack, potentially allowing them to navigate extremist rhetoric more effectively (Bilazarian, 2019). These messengers could engage individuals through direct messaging to provide counter-arguments, emotional support, and models for disengagement (Bilazarian, 2019). Such approaches should nevertheless be matched to the individual's level of risk, as direct counter-messaging can also trigger psychological reactance when individuals perceive an intervention as targeted, patronising, or ideologically imposed

(OECD, 2022; Saleh et al., 2021)**.** These findings suggest that interventions should prioritise credible messengers and interactive delivery formats to enhance engagement and foster alternative pathways of belonging.

Where online engagement indicates a need for more sustained support, referral to psychosocial prevention programmes may provide an additional off-ramp. Jugl et al. (2020), drawing on approximately 14,000 individual reports, found modest but statistically significant positive effects of psychosocial interventions on behavioural and mental-state outcomes (Cohen's d = 0.50). These programmes included approaches such as social-skills training, mentalisation-based treatment, and multi-component community interventions. However, the substantial variation between programmes means that this pooled effect should be interpreted as evidence that psychosocial prevention is plausible, rather than as evidence that any particular intervention is effective. Jugl et al. (2020) also found stronger results where high-risk individuals were supported within mixed-group programmes rather than being isolated into exclusively high-risk interventions. For stakeholders using ATKM findings to inform further assessment or intervention, this supports referral pathways through which individuals who may benefit from psychosocial support can be directed towards appropriate services without necessarily being labelled or treated as extremists. The evidence base for such referrals remains limited, however. When Lewis et al. (2025) searched for studies of programmes that steer people away from the criminal justice system and into support instead, only one study was rigorous enough to be included, which shows how little good evidence there is that such programmes work. This does not demonstrate that psychosocial or diversionary approaches are ineffective, but it does mean that referral should not be assumed to result in disengagement. Any referral pathway should therefore be accompanied by independent outcome

evaluation, particularly where interventions are delivered through external partners (Lewis et al., 2025; Sydes et al., 2023).

In addition to redirecting engagement and leveraging credible messengers, further interventions could build on the gamified nature of extremist environments by using interactive simulations to reframe user decision-making. For example, serious games such as *DECOUNT* are designed to help prevent radicalisation by safely simulating online spaces where extremist recruitment and propaganda are often spread (Lippe et al., 2022). It is a browser-based game that shows how individual choices can influence a person's path toward or away from radicalisation. Players enter a simulated social media environment with elements such as news feeds and messaging features. They move through a story by making choices that shape subsequent outcomes. According to Lippe et al. (2022), empirical evaluation of the project shows that after playing, participants agreed less with extremist narratives, such as the legitimisation of violence or conspiracy theories. This effect was even stronger when the game was used as part of a guided workshop, which also reduced authoritarian attitudes. Thus, the combination of immersive and interactive counter-narrative environments with structured reflection may support sustained cognitive engagement with alternative perspectives.

A related approach is active inoculation, which aims to strengthen resistance before or during exposure by teaching users to recognise the manipulative techniques used in extremist recruitment. Saleh et al. (2021) developed *Radicalise*, a computer game that exposes users to these persuasion techniques and found that participants who played the game were significantly better able to identify extremist persuasion attempts online than those who played a non-inoculating comparison game. However, evidence also suggests that engagement with extremist content is partly a sorting mechanism: individuals who actively and repeatedly engage may

already display characteristics associated with radicalisation, including stronger preferences for hierarchy, in-group identification, out-group hostility, and radicalism intentions (Reeve, 2021). For stakeholders with access to user-level or behavioural information, this distinction between incidental exposure and sustained engagement may provide important context when assessing potential risk. Escalating interaction with extremist material, repeated engagement, or attempts to establish connections with extremist accounts may therefore provide stronger indicators of risk and justify more intensive referral or intervention.

A further recommendation concerns content that falls outside the threshold for legally prohibited terrorist content but may nevertheless contribute to the normalisation or diffusion of extremist ideas. Humorous, ironic, or coded material can be difficult to assess consistently because its meaning often depends on cultural and contextual knowledge, while automated systems may struggle to distinguish satire from implicit extremist messaging (Lakhani, 2024; Schlegel, 2021). Van Ginkel et al. (2025), writing for the International Centre for Counter-Terrorism, therefore recommend clearer operational definitions of implicit extremist content and greater reliance on hybrid moderation, whereby automated detection systems identify potentially problematic content for review by trained human moderators. It also emphasises the importance of ongoing cultural, ideological, and contextual training for moderators, particularly when assessing content whose meaning is ambiguous or context dependent. Within its existing statutory mandate, ATKM can contribute to this broader approach by identifying recurring forms and characteristics of implicit or borderline content encountered during content assessments and sharing relevant trend-level insights with stakeholders responsible for moderation and prevention. For hosting service providers, this could support more consistent escalation of ambiguous cases to trained human moderators,

supported by clear guidance on indicators such as coded language, extremist symbols, ideological references, and the deliberate use of humour or irony to obscure harmful messaging (Van Ginkel et al., 2025). Such an approach would allow stakeholders to address potentially harmful content without treating all ambiguous or lawful expression as equivalent to terrorist material, thereby reducing the risks of both under-intervention and over-removal.

Borderline material of this kind also raises a question that sits above ATKM's mandate, namely what should be done about content that is harmful but lawful. The Digital Services Act already offers one route that does not require widening the definition of terrorist content. Article 34(1) of Regulation (EU) 2022/2065 requires providers of very large online platforms and very large online search engines to identify and assess the systemic risks arising from the design or functioning of their services, and the categories listed there reach beyond illegal content to include negative effects on civic discourse and public security, and on the protection of minors and on physical and mental well-being. Article 35 then requires those providers to put mitigation measures in place. Because these duties attach to risk rather than to illegality, they are the more appropriate instrument for the coded, ironic, and borderline material described above.

Our recommendation on this point is therefore addressed to policymakers, to the national Digital Services Coordinator, and to law enforcement rather than to ATKM. The recurring forms and characteristics of borderline extremist content that ATKM encounters during content assessment are precisely the evidence that systemic risk assessment under Article 34 needs and currently lacks. Establishing a route by which trend-level findings of that kind can be shared with the bodies responsible for Digital Services Act oversight would let this material be acted upon without

expanding what counts as terrorist content, and without asking ATKM to take on a role it does not have.

Tertiary & Operational Work: Targeted Intervention and Operational Response

To support disruption at later stages of engagement, interventions may prioritise the identification of behavioural signals associated with potential mobilisation. Brown et al. (2024) demonstrate that discussions related to operational planning, logistics, and weapon acquisition may be more closely associated with escalation than ideological expression alone. To support this, research suggests that machine learning models could assist in identifying patterns of operational language. According to Brown et al. (2024), such approaches may help differentiate between user groups by analysing linguistic and behavioural features such as topic distributions, post length, and punctuation patterns. Specifically, such systems could prioritise vocabulary related to physical logistics, including surveillance, targeting discussions, or tactical preparation, which may assist in identifying potentially higher risk engagement patterns. In this sense, this approach may help analysts differentiate between individuals who use extremist spaces as a performative outlet and those whose digital activity may reflect operational planning considerations.

Beyond user-focused interventions, reducing the spread of extremist content may also be effective at this stage. This would, however, demand possible changes to the technical architectures of digital platforms. A key component of this architecture is the recommender system, which is an automated algorithm designed to suggest content based on previous user interactions (Kunst et al., 2026). These systems are designed to optimise engagement and can function as a ‘supernormal stimulus’ by presenting exaggerated or emotionally salient content that exploits cognitive biases and reinforces user attention (Kunst et al., 2026). Although the causal impact of

algorithmic curation is still debated, such systems may contribute to what is termed the 'AI architecture of radicalisation' (Kunst et al., 2026). In this process, users may unintentionally train algorithms to push increasingly extreme content. To mitigate these risks, interventions should focus on embedding safety within platform designs. One such approach is the introduction of 'technical friction'. These are design features such as warning labels, sharing limits, or interaction delays that slow content dissemination and disrupt rapid sharing of potentially harmful material. Another possible approach is cross-platform hash-sharing coordinated through organisations such as the Global Internet Forum of Counter Terrorism (GIFCT). This involves generating unique digital signatures (hashes) for known extremist content, including manifestos or tactical guides (GIFCT; Heller, 2019; Saltman & Karhili, 2024). Once they are identified, this material can be flagged and removed across participating platforms before it spreads further (Heller, 2019; Saltman & Karhili, 2024). This approach enables coordinated prevention without requiring the exchange of sensitive user data. Together, such measures aim to shift platform incentives away from engagement maximisation toward safety-oriented design, thereby reducing opportunities for extremist content to circulate and accumulate visibility. Such measures could help disrupt the technical feedback loops that steer vulnerable users toward extreme content, and in some cases, violent action.

Limitations and Evidence Gaps

Several limitations apply to the operational framework proposed here. The four-type typology is conceptual rather than diagnostic. It is intended to structure analyst judgement about what a given piece of content signals, not to sort individuals into fixed categories, and it has not yet been validated against case data. No inter-rater reliability testing or pilot has been conducted, so it is not known whether two analysts applying the same indicators to the same material would arrive at the same tier. The decision

sequence that operationalises it is set out in Appendix A, and the limitations described in this section apply to that appendix in full.

The base-rate problem is the most serious constraint. Mobilisation to violence is rare relative to the volume of content ATKM assesses, which means that even indicators with reasonable sensitivity will produce a substantial number of false positives. The thresholds proposed for each tier are therefore starting points rather than settled values. The capability and operational thresholds in particular require calibration against ATKM's own referral outcomes before high-risk classification is relied upon at scale.

The evidence base carries its own limits. Much of the research underpinning the indicator set was published between 2010 and 2020, and the online environment it describes has changed considerably since. Platform architectures, gaming ecosystems, and the routes through which young people encounter extremist material continue to evolve faster than the evaluation literature can track, which means that indicators derived from earlier cohorts may not transfer cleanly to current cases. This argues for treating the framework as a living instrument subject to scheduled review rather than as a fixed protocol.

References

Addo, P. C., Fang, J., Kulbo, N. B., Gumah, B., Dagadu, J. C., & Li, L. (2021). Violent Video Games and Aggression Among Young Adults: The Moderating Effects of Adverse Environmental Factors. *Cyberpsychology, Behavior, and Social Networking*, *24*(1), 17-23. https://doi.org/10.1089/cyber.2020.0018

Adebimpe, A., Bassett, D. S., Jamieson, P. E., & Romer, D. (2019). Intersubject synchronization of late adolescent brain responses to violent movies: A virtue-ethics approach. *Frontiers in Behavioral Neuroscience, 13*, Article 260. https://doi.org/10.3389/fnbeh.2019.00260

American Psychological Association. (2020). *Resolution on violent video games*. https://www.apa.org/about/policy/resolution-violent-video-games.pdf

Anderson, C. A., & Bushman, B. J. (2001). Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior: A meta-analytic review of the scientific literature. *Psychological Science, 12(5), 353-359.* https://doi.org/10.1111/1467-9280.00366

Andrews, S. (2024). The "first person shooter" perspective: A different view on first person shooters, gamification, and first person terrorist propaganda. *Games and Culture*, *19*(1), 55–74. https://doi.org/10.1177/15554120231153789

Autoriteit online Terroristisch en Kinderpornografisch Materiaal. (2026). Online terroristisch materiaal. https://www.atkm.nl

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, *3*(3), 193–209.

Bilazarian, T. (2019). Countering violent extremist narratives online: Lessons from offline countering violent extremism. *Policy & Internet*, *12*(1), 46–65. https://doi.org/10.1002/poi3.204

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

Blakemore, S. J. (2012). Imaging brain development: The adolescent brain. *NeuroImage, 61*(2), 397–406. https://doi.org/10.1016/j.neuroimage.2011.11.080

Böckler, N., Hoffmann, J., & Zick, A. (2015). The Frankfurt Airport attack: A case study on the radicalization of a lone-actor terrorist. *Journal of Threat Assessment and Management*, *2*(3-4), 153-163. https://doi.org/10.1037/tam0000045

Braddock, K. (2022). Vaccinating against hate: Using attitudinal inoculation to confer resistance to persuasion by extremist propaganda. *Terrorism and Political Violence*, *34*(2), 240–262. https://doi.org/10.1080/09546553.2019.1693370

Brown, O., Smith, L., Davidson, B. I., Racek, D., & Joinson, A. (2024). Online signals of extremist mobilization. *Personality and Social Psychology Bulletin*, *52*(1), 70–89. https://doi.org/10.1177/01461672241266866

Burkhardt, J., & Lenhard, W. (2022). A Meta-Analysis on the Longitudinal, Age-Dependent Effects of Violent Video Games on Aggression. *Media Psychology*, *25*(3), 499–512. https://doi.org/10.1080/15213269.2021.1980729

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small

sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365-376. https://doi.org/10.1038/nrn3475

Carnagey, N. L., Anderson, C. A., & Bushman, B. J. (2007). The effect of video game violence on physiological desensitization to real-life violence. *Journal of Experimental Social Psychology*, *43*(3), 489-496. https://doi.org/10.1016/j.jesp.2006.05.003

Casey, B. J., Getz, S., & Galvan, A. (2008). The adolescent brain. *Developmental Review*, *28*(1), 62-77. https://doi.org/10.1016/j.dr.2007.08.003

Casey, B. J., Jones, R. M., & Hare, T. A. (2008). The adolescent brain. *Annals of the New York Academy of Sciences, 1124*(1), 111-126. https://doi.org/10.1196/annals.1440.010

Charmaraman, L., Richer, A. M., & Moreno, M. A. (2020). Social and Behavioral Health Factors Associated with Violent and Mature Gaming in Early Adolescence. *International Journal of Environmental Research and Public Health*, *17*(14), 4996. https://doi.org/10.3390/ijerph17144996

Chen, S., Mao, B., & Liu, Y. (2023). The effect of justified video game violence on aggressive behavior and moderated immersion: An experimental approach. Aggressive Behavior, 49, 68–75. https://doi.org/10.1002/ab.22054

Cohen, M. S., Leong, Y. C., Ruby, K., Pape, R. A., & Decety, J. (2024). Intersubject correlations in reward and mentalizing brain circuits separately predict persuasiveness of two types of ISIS video propaganda. *Scientific Reports, 14*, 13455. https://doi.org/10.1038/s41598-024-62341-3

Conway, M., Khawaja, M., Lakhani, S., Reffin, J., Robertson, A., & Weir, D. (2019). Disrupting Daesh: Measuring takedown of online terrorist material and its impacts. *Studies in Conflict & Terrorism, 42*(1-2), 141-160. https://doi.org/10.1080/1057610X.2018.1513984

Cosquer, M., Finck, C., Jousselme, C., & Lefebvre, A. (2025). Violent video gaming among French adolescents: Impact on mental health by gender. *L'Encephale*, *51*(2), 119-126 https://doi.org/10.1016/j.encep.2024.02.001

Cote, A. C., Coles, S. M., & Dal Cin, S. (2021). The interplay of parenting style and family rules about video games on subsequent fighting behavior. *Aggressive Behavior*, *47*(2), 135-147. https://doi.org/10.1002/ab.21931

Cubitt, T., & Morgan, A. (2024). Exposure to and sharing of fringe or radical content online (Trends & Issues in Crime and Criminal Justice No. 705). Australian Institute of Criminology. https://www.aic.gov.au/publications/tandi/tandi705

Delhove, M., & Greitemeyer, T. (2023). The social contagious effects of violent video game play: retaliatory, displaced, and the mere observation of aggression. *Social Influence*, *18*(1). https://doi.org/10.1080/15534510.2023.2277476

Denmark. European Counter Terrorism Centre. https://www.europol.europa.eu/media-press/newsroom/news/terrorgram-network-dealt-another-blow-member-sentenced-in-denmark

Devilly, G. J., O'Donohue, R. P., & Brown, K. (2023). Personality and frustration predict aggression and anger following violent media. *Psychology, Crime & Law*, *29*(1), 83–119. https://doi.org/10.1080/1068316X.2021.1999949

Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce). (2000). *Official Journal of the European Communities*, *L 178*, 1-16.https://eur-lex.europa.eu/eli/dir/2000/31/oj

Duindam, H. M., Asscher, J. J., & Lösel, F. (2025). Understanding unrest: Conspiracy belief and violent radicalization patterns in young people

during the COVID-19 pandemic in the Netherlands. *European Journal on Criminal Policy and Research.* Advance online publication. https://doi.org/10.1007/s10610-025-09634-z

Duindam, H. M., van Loon, A. W. G., & Lösel, F. (2025). Protective factors against extremism and violent radicalization in youth: A multilevel meta-analysis. *Aggression and Violent Behavior, 85*, 102101. https://doi.org/10.1016/j.avb.2025.102101

Duindam, H. M., Verhoef, R. E. J., van den Akker, A. L., & Asscher, J. J. (2025). Far-right violent radicalization profiles of youth in the Netherlands. *Psychology of Violence, 15*(4), 385-394. https://doi.org/10.1037/vio0000545

Ellefsen, R., & Sandberg, S. (2024). Everyday prevention of radicalization: The impacts of family, peer, and police intervention. *Studies in Conflict & Terrorism* , *47*(10), 1342–1365. https://doi.org/10.1080/1057610x.2022.2037185

Eltaher, F., Gajula, R. K., Miralles-Pechuán, L., Crotty, P., Martínez-Otero, J., Thorpe, C., & McKeever, S. (2025). Protecting young users on social media: Evaluating the effectiveness of content moderation and legal safeguards on video-sharing platforms. *arXiv preprint arXiv:2505.11160*. https://doi.org/10.48550/arXiv.2505.11160

Engelhardt, C. R., Bartholow, B. D., Kerr, G. T., & Bushman, B. J. (2011). This is your brain on violent video games: Neural desensitization to violence predicts increased aggression. *Journal of Experimental Social Psychology*, *47*(5), 1033-1036. https://doi.org/10.1016/j.jesp.2011.03.027

eSafety Commissioner. (2023). *Mind the gap: The disconnect between parents' and children's awareness of online risks*. https://www.esafety.gov.au/research/mind-the-gap

eSafety Commissioner. (2024). *Gore online: How violent content is reaching children and what you can do*. https://www.esafety.gov.au/newsroom/blogs/gore-online-how-violent-content-is-reaching-children-and-what-you-can-do

European Parliament & Council of the European Union. (2021). Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online. *Official Journal of the European Union, L*, *172*, 72–109.

Europol. (2025). *European Union terrorism situation and trend report 2025 (EU TE-SAT)*. Publications Office of the European Union. https://www.europol.europa.eu/publication-events/main-reports/european-union-terrorism-situation-and-trend-report-2025-eu-te-sat

Europol. (2026a). *2024 EU IRU transparency report*. Publications Office of the European Union. https://www.europol.europa.eu/publications-events/publications/eu-internet-referral-unit-transparency-report-2024

Europol. (2026b). *Project Compass: A project dedicated to fighting against The Com.* https://www.europol.europa.eu/how-we-work/operations/project-compass

Europol. (2026c, April 30). "Terrorgram" network dealt another blow as member sentenced in

Fahs, I., Akel, M., Malaeb, D., Haddad, C., Kheir, N., Hallit, S., & Obeid, S. (2022). Association of violent video games with alcohol use disorder and smoking dependence among male adolescents in Lebanon. *Archives de Pediatrie*, *29*(5), 354-358. https://doi.org/10.1016/j.arcped.2022.05.014

Feddes, A. R., Mann, L., & Doosje, B. (2015). Increasing self-esteem and empathy to prevent violent radicalization: A longitudinal quantitative evaluation of a resilience training focused on adolescents with a dual identity. *Journal of Applied Social Psychology*, *45*(7), 400-411. https://doi.org/10.1111/jasp.12307

Felemban, O., Al-Zahrani, A., Alhazmi, A., Alahmadi, R., Alsalhi, A., & Alahmadi, R. (2021). The effect on playing violent video games among school going children. *International Journal of Clinical Skills*, *15*(9), 518-524. https://d1wqtxts1xzle7.cloudfront.net/107680237

Ferguson, C. J. (2015). Do angry birds make for angry children? A meta-analysis of video game influences on children's and adolescents' aggression, mental health, prosocial behavior, and academic performance. *Perspectives on Psychological Science, 10*(5), 646-666. https://doi.org/10.1177/1745691615592234

Fikkers, K. M., Piotrowski, J. T., & Valkenburg, P. M. (2017). A matter of style? Exploring the effects of parental mediation styles on early adolescents' media violence exposure and aggression. *Computers in Human Behavior*, *70*, 407-415. https://doi.org/10.1016/j.chb.2017.01.029

Fikkers, K. M., Piotrowski, J. T., Lugtig, P., & Valkenburg, P. M. (2016). The Role of Perceived Peer Norms in the Relationship Between Media Violence Exposure and Adolescents' Aggression. *Media Psychology*, *19*(1), 4–26.https://doi.org/10.1080/15213269.2015.1037960

Fikkers, K. M., Taylor Piotrowski, J., Weeda, W. D., Vossen, H. G., & Valkenburg, P. M. (2013). Double dose: High family conflict enhances the effect of media violence exposure on adolescents' aggression. *Societies*, *3*(3), 280-292. https://doi.org/10.3390/soc3030280

Funk, J. B. (2005). Children's exposure to violent video games and desensitization to violence. *Child and Adolescent Psychiatric Clinics*, *14*(3), 387-404. https://doi.org/10.1016/j.chc.2005.02.009

Gallacher, J. D., Heerdink, M. W., & Hewstone, M. (2021). Online engagement between opposing political protest groups via social media is linked to physical violence of offline encounters. *Social Media + Society*, *7*(1), 1–16. https://doi.org/10.1177/2056305120984445

Gentile, D. A., Bender, P. K., & Anderson, C. A. (2017). Violent video game effects on salivary cortisol, arousal, and aggressive thoughts in children. *Computers in Human Behavior*, *70*, 39-43. https://doi.org/10.1016/j.chb.2016.12.045

Gentile, D. A., Swing, E. L., Anderson, C. A., Rinker, D., & Thomas, K. M. (2016). Differential neural recruitment during violent video game play in violent- and nonviolent-game players. *Psychology of Popular Media Culture, 5*(1), 39-51. https://doi.org/10.1037/ppm0000009

Global Internet Forum to Counter Terrorism. (2024). *2023 GIFCT annual transparency report*.

Global Network on Extremism and Technology (2026, February 2). *Escape the void: Responding to youth-led nihilistic violence*. https://gnet-research.org/2026/02/02/escape-the-void-responding-to-youth-led-nihilistic-violence/

Gonçalves de Medeiros, B., Eduardo Pimentel, C., Miranda Sarmet, M., & Evangelista Mariano, T. (2020). " Brutal Kill!" Violent video games as a predictor of aggression. *Psico-USF*, *25*(2). https://doi.org/10.1590/1413-82712020250205

Goodson, S., Turner, K. J., Pearson, S. L., & Carter, P. (2021). Violent Video Games and the P300: No Evidence to Support the Neural Desensitization Hypothesis. *Cyberpsychology, Behavior, and Social Networking*, *24*(1), 48–55. https://doi.org/10.1089/cyber.2020.0029

Greipl, S., Lechner, M., Fischer, J., Schulze, H., Hohner, J., & Rieger, D. (2025). Tracking down the Candy Crush terrorist: The fragile relation between gaming motives and radical attitudes. *Frontiers in Psychology*, *16*. https://doi.org/10.3389/fpsyg.2025.1585576

Guo, X., Zheng, L., Wang, H., Zhu, L., Li, J., Wang, Q., Dienes, Z., & Yang, Z. (2013). Exposure to violence reduces empathetic responses to

other's pain. *Brain and Cognition, 82*(2), 187–191. https://doi.org/10.1016/j.bandc.2013.04.005

Hamid, N., & Ariza, C. (2022). *Offline versus online radicalisation: Which is the bigger threat? Tracing outcomes of 439 jihadist terrorists between 2014-2021 in 8 Western countries*. Global Network on Extremism and Technology. https://gnet-research.org/wp-content/uploads/2022/02/GNET-Report-Offline-Versus-Online-Radicalisation.pdf

Hassan, G., Brouillette-Alarie, S., Alava, S., Frau-Meigs, D., Lavoie, L., Fetiu, A., Varela, W., Borokhovski, E., Venkatesh, V., Rousseau, C., & Sieckelinck, S. (2018). Exposure to extremist online content could lead to violent radicalization: A systematic review of empirical evidence. *International Journal of Developmental Science*, *12*(1-2), 71–88. https://doi.org/10.3233/dev-170233

Heller, B. (2019). *Combating terrorist-related content through AI and information sharing*. (Transatlantic Working Group Papers Series). Annenberg Public Policy Center. https://www.annenbergpublicpolicycenter.org/twg

Hemmila, T., & Perliger, A. (2025). Hybridization or salad bar ideology? Testing ideological convergence within the American violent far right. *Crime & Delinquency, 71*(3), 830-859. https://doi.org/10.1177/00111287241271288

Horgan, J. (2005). *The psychology of terrorism.* Routledge. https://doi.org/10.4324/9780203496961

Hummer, T. A. (2015). Media violence effects on brain development: What neuroimaging has revealed and what lies ahead. *American Behavioral Scientist, 59*(14), 1790-1806. https://doi.org/10.1177/0002764215596553

Hummer, T. A., Wang, Y., Kronenberger, W. G., Mosier, K. M., Kalnin, A. J., Dunn, D. W., & Mathews, V. P. (2010). Short-Term Violent Video

Game Play by Adolescents Alters Prefrontal Activity During Cognitive Inhibition. *Media Psychology*, *13*(2), 136–154. https://doi.org/10.1080/15213261003799854

Jabr, M. M., Denke, G., Rawls, E., & Lamm, C. (2018). The roles of selective attention and desensitization in the association between video gameplay and aggression: An ERP investigation. *Neuropsychologia*, *112*, 50-57. https://doi.org/10.1016/j.neuropsychologia.2018.02.026

Jensen, M., James, P., & Yates, E. (2023). Contextualizing disengagement: How exit barriers shape the pathways out of far-right extremism in the United States. *Studies in Conflict & Terrorism*, *46*(3), 249–277. https://doi.org/10.1080/1057610x.2020.1759182

Jugl, I., Lösel, F., Bender, D., & King, S. (2020). Psychosocial prevention programs against radicalization and extremism: A meta-analysis of outcome evaluations. *European Journal of Psychology Applied to Legal Context, 13*(1), 37-46. https://doi.org/10.5093/ejpalc2021a6

Kenyon, J., Baker-Beall, C., & Binder, J. (2022a). The internet and radicalisation pathways: technological advances, relevance of mental health and role of attackers. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4955993

Kenyon, J., Binder, J. F., & Baker-Beall, C. (2022b). 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

Kenyon, J., Carter, A. J., Watson, S., & Farr, J. (2025). Adapting risk assessments to a changing terrorism landscape: Revising the extremism risk guidance. *Journal of Forensic Sciences*, *70*, 2031-2041. https://doi.org/10.1111/1556-4029.70101

Klompmaker, N. (2019). Censor them at any cost? A social and legal assessment of enhanced action against terrorist content online. *Amsterdam Law Forum, 11*(3), 24-47. https://www.readkong.com/page/censor-them-at-any-cost-a-social-and-legal-assessment-of-2345364

Koehler, D., Fiebig, V., & Jugl, I. (2023). From gaming to hating: Extreme-right ideological indoctrination and mobilization for violence of children on online gaming platforms. *Political Psychology*, *44*(2), 419–434. https://doi.org/10.1111/pops.12855

Kozyreva, A., Smillie, L., & Lewandowsky, S. (2023). Incorporating psychological science into policy making: The case of misinformation. *European Psychologist, 28*(3), 206-224. https://doi.org/10.1027/1016-9040/a000493

Kühn, S., Kugler, D., Schmalen, K. *et al.* Does playing violent video games cause aggression? A longitudinal intervention study. *Mol Psychiatry* **24**, 1220–1234 (2019). https://doi.org/10.1038/s41380-018-0031-7

Kühn, S., Kugler, D., Schmalen, K., Weichenberger, M., Witt, C., & Gallinat, J. (2019). The myth of blunted gamers: No evidence for desensitization in empathy for pain after a violent video game intervention in a longitudinal fMRI study on non-gamers. *Neurosignals*, *26*(1), 22-30. https://doi.org/10.1159/000487217

Kuner, C., Bygrave, L. A., Docksey, C., & Drechsler, L. (Eds.). (2020). *The EU General Data Protection Regulation (GDPR): A commentary*. Oxford University Press. https://doi.org/10.1093/oso/9780198826491.001.0001

Kunst, J. R., Obaidi, M., Gollwitzer, A., Brandtzaeg, P. B., Hinrichs, Y., Saini, N., & Schroeder, D. T. (2026). Intelligent systems, vulnerable Minds: A framework for radicalization to violence in the age of AI. *Personality and Social Psychology Review*, 1–32. https://doi.org/10.1177/10888683261430089

LaCroix, H., & Berkovits, S. (2026). The digital radical: Online extremism, propagation, and terror by design's strategic implications for homeland security. *Journal of Strategic Security*, *19*(1), 121–167. https://doi.org/10.5038/1944-0472.19.1.2547

Lakhani, S. (2024). A is for apple, B is for bullet: The gamification of (violent) extremism. In L. Schlegel & R. Kowert (Eds.), *Gaming and extremism: The radicalization of digital playgrounds* (pp. 148–162). Routledge. https://doi.org/10.4324/9781003388371

Lakhani, S., & Wiedlitzka, S. (2023). "Press f to pay respects": An empirical exploration of the mechanics of gamification in relation to the christchurch attack. *Terrorism and Political Violence*, *35*(7), 1586–1603. https://doi.org/10.1080/09546553.2022.2064746

Lamphere-Englund, G. (2024). Theories of digital games & radicalization. In L. Schlegel & R. Kowert (Eds.), *Gaming and extremism: The radicalization of digital playgrounds* (pp. 32–56). Routledge. https://doi.org/10.4324/9781003388371

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 and Terrorism*, 1–25. https://doi.org/10.1080/1057610x.2024.2413184

Lewis, J., Marsden, S., Hewitt, J., Squires, C., & Stefaniak, A. (2025). Non-criminal justice interventions for countering cognitive and behavioural radicalisation amongst children and adolescents: A systematic review of effectiveness and implementation. *Campbell Systematic Reviews*, *21*(4), Article e70079. https://doi.org/10.1002/cl2.70079

Lippe, F., Walter, R., & Hofinger, V. (2022). Evaluating an online-game intervention to prevent violent extremism. *Journal for Deradicalization*, *32*, 1–34.

Liu, S., Chang, F.-C., Chiu, C.-H., Li, F., Chen, P.-H., Chen, C.-Y., Lin, Y.-P., & Chiang, J.-T. (2021). Parent–Child Discrepancies in Reports of Exposure to Violence/Pornography on Mobile Devices and the Impact on Children's Psychosocial Adjustment. *Cyberpsychology, Behavior, and Social Networking*, *24*(6), 390-398. https://doi.org/10.1089/cyber.2020.0120

Lloyd, M., & Dean, C. (2015). The development of structured guidelines for assessing risk in extremist offenders. *Journal of Threat Assessment and Management*, *2*(1), 40-52. https://doi.org/10.1037/tam0000035

Meulen, P., Koning, N., Assink, M., van Hooren, S., Kolthoff, E., & Heynen, E. (2026). Letting work what works—Effectively preventing juvenile delinquency in the Netherlands: A meta-analysis of the evidence. *International Journal of Offender Therapy and Comparative Criminology, 70*(5), 475-514. https://doi.org/10.1177/0306624X251344620

Mößle, T., Kliem, S., & Rehbein, F. (2014). Longitudinal effects of violent media usage on aggressive behavior—The significance of empathy. *Societies*, *4*(1), 105-124. https://doi.org/10.3390/soc4010105

Neo, L. S., Dillon, L., & Khader, M. (2016). Identifying individuals at risk of being radicalised via the internet. *Security Journal*, *30*(4), 1112–1133. https://doi.org/10.1057/s41284-016-0080-z

Noonan, W. C., & Moyers, T. B. (1997). Motivational interviewing. *Journal of Substance Misuse, 2*(1), 8-16. https://doi.org/10.3109/14659899709084610

OECD (2022), "Misinformation and disinformation: An international effort using behavioural science to tackle the spread of misinformation", *OECD Public Governance Policy Papers*, No. 21, OECD Publishing, Paris, https://doi.org/10.1787/b7709d4f-en.

Papadamou, K., Papasavva, A., Zannettou, S., Blackburn, J., Kourtellis, N., Leontiadis, I., Stringhini, G., & Sirivianos, M. (2020). Disturbed YouTube for Kids: Characterizing and Detecting Inappropriate Videos Targeting Young Children. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*(1), 522-533. https://doi.org/10.1609/icwsm.v14i1.7320

Prescott, A. T., Sargent, J. D., & Hull, J. G. (2018). Metaanalysis of the relationship between violent video game play and physical aggression over time. *Proceedings of the National Academy of Sciences of the United States of America, 115*(40), 9882-9888. https://doi.org/10.1073/pnas.1611617114

Pretus, C., Hamid, N., Sheikh, H., Ginges, J., Tobena, A., Davis, R., Vilarroya, O., & Atran, S. (2018). Neural and behavioral correlates of sacred values and vulnerability to violent extremism. *Frontiers in Psychology*, *9*, Article 2462. https://doi.org/10.3389/fpsyg.2018.02462

Prochaska, J. O., & DiClemente, C. C. (1983). Stages and processes of self-change of smoking: Toward an integrative model of change. *Journal of Consulting and Clinical Psychology, 51*(3), 390–395. https://doi.org/10.1037/0022-006X.51.3.390

Public Prosecution Service of Canada. (2026, March 27). *Canadian Terrorgram member sentenced to 20 years*. https://www.ppsc.gc.ca/eng/nws-nvs/2026/27_03_26.html

Reeve, Z. (2021). Engaging with online extremist material: Experimental evidence. *Terrorism and Political Violence, 33*(8), 1595-1620. https://doi.org/10.1080/09546553.2019.1634559

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). (2016). *Official Journal of the European Union*, *L 119*, 1-88. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online. (2021). *Official Journal of the European Union, L 172*, 79-109. https://eur-lex.europa.eu/eli/reg/2021/784/oj

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act). (2022). Official Journal of the European Union, L 277, 1-102. https://eur-lex.europa.eu/eli/reg/2022/2065/oj

Saleh, N. F., Roozenbeek, J., Makki, F. A., McClanahan, W. P., & van der Linden, S. (2021). Active inoculation boosts attitudinal resistance against extremist persuasion techniques: A novel approach towards the prevention of violent extremism. *Behavioural Public Policy*, 1-24. https://doi.org/10.1017/bpp.2020.60

Saltman, E. M., Kooti, F., & Vockery, K. (2021). New models for deploying counterspeech: Measuring behavioral change and sentiment analysis. *Studies in Conflict & Terrorism*. Advance online publication. https://doi.org/10.1080/1057610X.2021.1888404

Saltman, E., & Karhili, N. E. (2024). Policies, Practices, and Positive Interventions to counter terrorism and violent extremism in gaming spaces. In L. Schlegel & R. Kowert (Eds.), *Gaming and extremism: The radicalization of digital playgrounds* (pp. 163–184). Routledge.

Sánchez-García, J., & Supervía, P. U. (2025). The role of video game consumption habits in the relationship between gender and attitudes towards violence among adolescents. *Scientific Reports*, *15*(1), 26142. https://doi.org/10.1038/s41598-025-95555-0

Savage, S., Francar, J., & Perry, K. (2023). Promoting cognitive complexity among Yezidi youth impacted by ISIS in Kurdistan, Iraq. *Journal of Strategic Security*, *16*(4), 105–157. https://doi.org/10.2307/48751618

Scarcella, A., Page, R., & Furtado, V. (2016). Terrorism, radicalisation, extremism, authoritarianism and fundamentalism: A systematic review of the quality and psychometric properties of assessments. *PLOS ONE*, *11*(12), e0166947. https://doi.org/10.1371/journal.pone.0166947

Schlegel, L. (2021). Connecting, competing, and trolling: "User Types" in digital gamified radicalization processes. *Perspectives on Terrorism*, *15*(4), 54–64.

Schumann, S., Kenyon, J., & Binder, J. (2025). Identifying distinct types of internet use that predict the likelihood of planning or committing a terrorist attack: Findings from an analysis of individuals convicted on terrorism(-related) charges in England and Wales. *Computers in Human Behavior*, *168*, 1–10. https://doi.org/10.1016/j.chb.2025.108646

Schuurman, B., & Bakker, E. (2016). Reintegrating jihadist extremists: evaluating a Dutch initiative, 2013–2014. *Behavioral Sciences of Terrorism and Political Aggression*, *8*(1), 66–85. https://doi.org/10.1080/19434472.2015.1100648

Scrivens, R., Wojciechowski, T. W., Gaudette, T., & Frank, R. (2026). Examining the persisting and desisting online posting behaviors of violent and non-violent right-wing extremists. *Terrorism and Political Violence*, *38*(3), 246–268. https://doi.org/10.1080/09546553.2024.2416153

Strenziok, M., Krueger, F., Deshpande, G., Lenroot, R. K., van der Meer, E., & Grafman, J. (2011). Fronto-parietal regulation of media violence exposure in adolescents: A multi-method study. *Social Cognitive and Affective Neuroscience*, *6*(5), 537-547. https://doi.org/10.1093/scan/nsq079

Swann, W. B., Jetten, J., Gómez, Á., Whitehouse, H., & Bastian, B. (2012). When group membership gets personal: A theory of identity fusion. *Psychological Review*, *119*(3), 441–456. https://doi.org/10.1037/a0028589

Sydes, M., Hine, L., Higginson, A., McEwan, J., Dugan, L., & Mazerolle, L. (2023). Criminal justice interventions for preventing radicalisation, violent extremism and terrorism: An evidence and gap map. *Campbell Systematic Reviews*, *19*(4), e1366. https://doi.org/10.1002/cl2.1366

Teng, Z., Nie, Q., Guo, C., Zhang, Q., Liu, Y., & Bushman, B. J. (2019). A longitudinal study of link between exposure to violent video games and aggression in Chinese adolescents: The mediating role of moral disengagement. *Developmental Psychology*, *55*(1), 184. https://doi.org/10.1037/dev0000624

The Soufan Center. (2021, March 29). *IntelBrief: The counterterrorism challenge of "salad bar" ideologies*. https://thesoufancenter.org/intelbrief-2021-march-29/

Thorley, T. G., & Saltman, E. (2026). GIFCT Tech Trials: Combining Behavioural Signals to Surface Terrorist and Violent Extremist Content Online. *Studies in Conflict & Terrorism*, *49*(4), 446-471. https://doi.org/10.1080/1057610X.2023.2222901

Uitvoeringswet verordening terroristische online-inhoud. (2023). Stb. 2023, 122. https://wetten.overheid.nl/BWBR0048064

Valentini, D., Lorusso, A. M., & Stephan, A. (2020). Onlife extremism: Dynamic integration of digital and physical spaces in radicalization. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00524

Van Ginkel, B., Mehra, T., Herbach, M., Lanchès, J., & Boerma, Y. (2025). *Blurred boundaries: Legal, ethical, and practical limits in detecting and moderating terrorist, illegal and implicit extremist content online while respecting freedom of expression*. International Centre for Counter-Terrorism. https://icct.nl/publication/blurred-boundaries-legal-ethical-and-practical-limits-detecting-and-moderating

Van der Heide, L., & Schuurman, B. (2018). Reintegrating terrorists in the Netherlands: Evaluating the Dutch approach. *Journal for Deradicalization*, (17), 196-239. https://journals.sfu.ca/jd/index.php/jd/article/view/137

Weimann, G., & Ben Am, A. (2020). Digital dog whistles: The new online language of extremism. *International Journal of Security Studies, 2*(1), Article 4. https://digitalcommons.northgeorgia.edu/ijoss/vol2/iss1/4

Wolbers, H., Dowling, C., Cubitt, T., & Kuhn, C. (2023). *Understanding and preventing internet-facilitated radicalisation*. Australian Institute of Criminology. https://doi.org/10.52922/ti77024

Wolfowicz, M., Litmanovitz, Y., Weisburd, D., & Hasisi, B. (2021). Cognitive and behavioral radicalization: A systematic review of the putative risk and protective factors. *Campbell Systematic Reviews*, *17*(3), 1–90. https://doi.org/10.1002/cl2.1174

Xue, L., Corso, F., Fontana, N., Liu, G., Ceri, S., & Pierri, F. (2025). Towards an Automated Framework to Audit Youth Safety on TikTok. In *Proceedings of the Fourth Workshop on Bridging Human-Computer Interaction and Natural Language Processing (HCI+ NLP)* (pp. 113-119). https://aclanthology.org/2025.hcinlp-1.9/

Ybarra, M. L., Mitchell, K. J., & Oppenheim, J. K. (2022). Violent media in childhood and seriously violent behavior in adolescence and young adulthood. The Journal of Adolescent Health, 71(3), 285–292. https://doi.org/10.1016/j.jadohealth.2022.03.003

Ybarra, M. L., Mitchell, K. J., & Oppenheim, J. K. (2022). Violent media in childhood and seriously violent behavior in adolescence and young adulthood. *Journal of Adolescent Health, 71*(3), 285-292. https://doi.org/10.1016/j.jadohealth.2022.03.003

Youth Endowment Fund. (2024). *Children, violence and vulnerability 2024: What role does social media play in violence affecting young people?* https://youthendowmentfund.org.uk/reports/children-violence-and-vulnerability-2024/social-media/

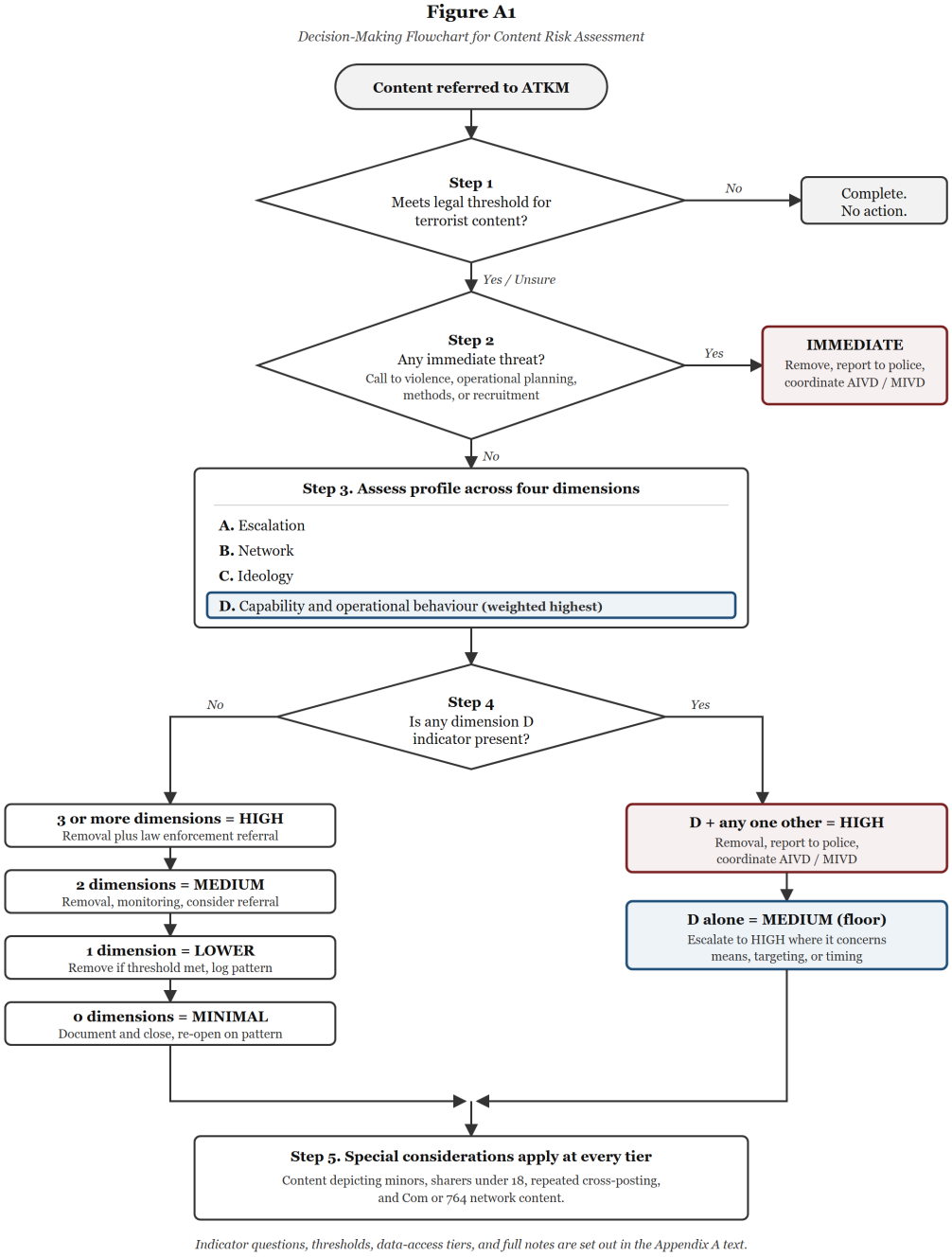
Appendix A. Risk-Tiered Decision Flowchart

Status of this appendix. The flowchart below is a proposal rather than a validated instrument. It has not been tested against ATKM case data, no agreement between analysts has been measured for it, and its thresholds have not been calibrated against known outcomes. We put it forward as a research recommendation: that ATKM pilot it retrospectively on a sample of historical referrals, measure whether analysts applying it independently arrive at the same tier, and compare its tier assignments with the outcomes those cases actually had. Several steps also depend on data that ATKM does not currently receive, and on processing whose legal basis has not been established. Before any use beyond such a pilot, the appendix would need review by ATKM's legal advisers and by the team responsible for content assessment. A study of this kind would suit a student research placement, which would also allow the coding to be done independently of the analysts whose decisions are being compared.

Terminology. "Removal" in this appendix means the issuing of a removal order under Regulation (EU) 2021/784, that is, an instruction to the hosting service provider to remove the content or disable access to it within one hour. It does not refer to any action taken against an account or against a person.

Figure A1

Decision-Making Flowchart for Content Risk Assessment



Note. Risk tier assignments draw on the engagement, intent, and capability dimensions of the Extremism Risk Guidance (ERG22+; Lloyd and Dean,

2015, revised as the ERG-R in Kenyon et al., 2025), the empirical differentiation between online-only and mixed-pathway trajectories (Binder and Kenyon, 2022; Hamid and Ariza, 2022), and operational intelligence on Com network profile mechanics (Europol, 2026b; GNET, 2026). Tier assignment sets a starting floor. The indicator assessment in Step 3 may elevate any profile to a higher tier where the evidence supports it.

Data access prerequisites. The indicators in Step 3 are observable at different levels of data access. Analysts should record which tier is available to them and apply only those indicators for which they hold data access.

- **Tier 1, content-observable.** Available from TCO Regulation referrals: content language, explicit ideological markers, named extremist symbols.
- **Tier 2, platform-accessible.** Requires a cooperation agreement: posting frequency over time, account creation date, follower lists.
- **Tier 3, cross-platform.** Requires inter-agency data sharing: platform migration patterns, network connections on other platforms.

Steps requiring Tier 2 or Tier 3 data are conditional on ATKM establishing the relevant agreements with hosting providers and intelligence partners.

Step 1. Initial content assessment

- Does the content meet the legal threshold for terrorist content?
- If no: the assessment is complete, no further action.
- If yes or uncertain: go to Step 2.

Step 2. Assessment of immediate threats

- Does the content actively call for immediate violence?
- Does the material discuss operational planning for future violence?
- Does the content give directions or methods for carrying out violence?
- Does the material attempt to recruit for or support a terrorist entity?

- **If any of the above:** the content meets the operational threshold for mobilisation risk, meaning the transition into behavioural radicalisation and operational preparation for violence (Scarcella et al., 2016). Immediate removal, report to police, and coordinate with intelligence services.
- **If none of the above:** go to Step 3.

Step 3. Assessment of contextual indicators, subject to data access

Four dimensions are assessed. Dimensions A to C are carried over from the original framework. Dimension D is new and is the addition this report makes. Dimensions A and B rest on information about an account over time rather than on a single item of content. They are therefore available only at the Tier 2 or Tier 3 data access described above, and only where a legal basis for that processing has been established. Dimensions C and D can be assessed from the referred content itself.

A. Escalation

- Has the account's rate of posting spiked recently?
- Has the extremity of the profile's messages worsened over time?
- Has the content shifted from political discussion into open violent threats?

B. Network

- Does the profile interact frequently with accounts identified as extremist?
- Does the account indicate recognised membership of fringe extremist groups?
- Is the content cross-posted into extremist enclaves such as Telegram or Discord channels?

C. Ideology

- Are the messages coherent enough to reflect an identifiable extremist ideology?

- Does the content use accelerationist rhetoric such as race war or impending societal collapse?
- Does the content show alignment with known jihadist, far-right, or other extremist groups?

D. Capability and operational behaviour (new dimension, weighted highest)

- Does the content use operational or logistical language, such as planning, timing, access, or materials?
- Is there evidence of tactical learning or rehearsal, including virtual training material, weapons or tactics research, or documented practice?
- Has the account migrated to encrypted channels specifically for logistical planning rather than for general privacy? Migration indicates deepening commitment and operational security rather than capability in itself (Kenyon et al., 2022a).
- Is there evidence of acquisition, construction, or documentation of means?
- Is there evidence of target selection or reconnaissance?

Step 4. Determine risk category

Capability and operational behaviour, dimension D, carries the greatest weight, because operational and tactical signals differentiate those who act from those who do not more reliably than ideological expression alone (Brown et al., 2024; Schumann et al., 2025). Ideology is no longer treated as one of three equal points.

- **D plus any one other dimension: HIGH.** Removal, report to police, coordinate with AIVD and MIVD.
- **D alone: MEDIUM as a floor.** Escalate to HIGH where the indicator concerns means, targeting, or timing.

- **Three or more dimensions without D: HIGH.** Removal plus law enforcement referral.
- **Two dimensions without D: MEDIUM.** Removal plus monitoring; consider a de-radicalisation referral.
- **One dimension without D: LOWER.** Remove where the legal threshold is met; log for pattern development.
- **No dimensions present: MINIMAL.** Case-by-case assessment; document and close; re-open if a pattern is detected.

Calibration note. The thresholds above are heuristic starting points derived from theoretical reasoning rather than cut-points validated against known outcomes, and they do not weight qualitatively different indicators within a dimension equally. Content giving specific attack instructions should carry more weight than a posting-frequency spike, even though each counts as one indicator. The thresholds should be treated as a provisional floor pending a structured pilot on historical cases.

Step 5. Special considerations

- **Content depicting minors.** Follow child safeguarding referral procedures under applicable Dutch child protection law. This operates independently of the risk tier assessment.
- **Content shared by a minor under 18.** The same removal and assessment criteria apply, but the default HIGH-risk response should prioritise non-criminal-justice pathways, consistent with UNCRC Article 3 best-interest obligations, unless specific indicators of operational violence planning are present. Lewis et al. (2025) review non-criminal justice interventions for this age group and found only one study eligible for their effectiveness analysis, so this default rests on proportionality rather than on demonstrated effectiveness. Police coordination should be a secondary rather than a default response.

- **Repeated cross-posting.** Where a single piece of content has been shared repeatedly across multiple sites, the distribution is likely deliberate and raises the potential harm.
- **Com and 764 network content.** Often mischaracterised as a formal ideology. Better described as nihilistic and violence-centred rather than politically ideological, with euphemisms such as lorebooks and fansigning standing in for real offences (Europol, 2026b; Global Network on Extremism and Technology, 2026). On coded and euphemistic extremist language more generally, see Weimann and Ben Am (2020).