Online Problem, Offline Solution: An Ethical Examination of Online Counter-Extremism Policy

Zac Thomson

Handa Centre for the Study of Terrorism & Political Violence, University of St Andrews, St Andrews, Scotland

ABSTRACT

This paper critically examines online counter-extremism policies that conflate radicalism with extremism, and extremism with violence, thereby suppressing non-violent dissent without addressing the core drivers of terrorism. Drawing on evidence of risk and protective factors, it argues for a shift from regulating ideas to preventing harmful actions. The proposed Fostering Resilience through Education, Empowerment, and Dialogue (FREED) framework prioritises targeted legal action against incitement while investing in community-based interventions that tackle socioeconomic inequalities and social isolation. By focusing on pathways leading to violence rather than stifling controversial beliefs, this approach preserves freedom of expression, fosters resilience, and provides an ethical strategy for preventing violent extremism.

Introduction

In many parts of the world, legislative and regulatory frameworks continue to evolve in an attempt to tackle the phenomenon of online extremist content. Policymakers who pursue these efforts often cite harrowing instances of terrorist violence, radicalisation clusters emerging on social media, or hate-based harassment of minority communities.[1,2] They contend that firm legal tools are urgently needed to minimise the availability of extremism, especially as it circulates with agility through global digital platforms.[3] Yet for all the moral clarity that may underlie these intentions, there remains a profound and unresolved tension between suppressing extremist material and preserving the breadth of free expression.[4,5] Some critics suspect that even well-intentioned bans on "harmful" or "extremist" content can morph into paternalistic or draconian measures that chill legitimate discourse, deter political opposition, and centralise the power to define public morals in the hands of regulators or platform executives.[6,7]

Approaching this issue, this paper offers an examination of the complex ethical dilemmas involved in preventing radicalisation towards violent extremism —particularly within the context of digital policy responses. Rather than asserting the superiority of any single viewpoint, this paper advocates for a reorientation of counter-extremism efforts—shifting the focus from cognitive extremism, which targets 'bad ideas,' to enhancing community resilience to violent narratives. This reframing prioritises the societal susceptibility to violent extremism over the regulation of speech. This paper is guided by the central question: *In addressing online radicalisation, is it possible to prevent harm without suppressing lawful and dissenting expression?* To address these questions, this paper first unpacks foundational concepts—radicalism, extremism, and terrorism—before examining the drivers of radicalisation, the efficacy and ethics of current content moderation and legislative approaches. It then presents a community-focused framework, which channels resources toward education, empowerment, and social investment while reserving punitive measures for explicit advocacy or acts of violence.

Unpacking Core Concepts

Extremism is a concept frequently invoked in policy documents, public debates, and scholarly literature, yet its exact contours remain remarkably elusive.[8] At its core, the term attempts to capture beliefs or practices that deviate sharply from accepted norms. However, policy frameworks that often conflate the terms radicalisation with extremism and extremism with violence and terrorism, serving only to blur important distinctions between non-violent beliefs and violent acts.[8] This is exemplified in the UK governments report titled *Tackling Extremism in the UK* where you are swiftly greeted with the following headings:

> *"The UK deplores and will fight terrorism of every kind, whether based on Islamist, extreme right-wing or any other extremist ideology. We will not tolerate extremist activity of any sort, which creates an environment for radicalising individuals and could lead them on a pathway towards terrorism."*[9(p.1)]

And:

> *"Extremist propaganda is too widely available, particularly online, and has a direct impact on radicalising individuals. The poisonous messages of extremists must not be allowed to drown out the voices of the moderate majority."*[9(p.3)]

In these headings, we see the government unequivocally linking a wide range of *"extremist ideologies"*[9(p.1)] to terrorism, emphasising *"the poisonous messages of extremists"*[9(p.3)] as a direct threat that *"could lead [individuals] on a pathway towards terrorism."*[9(p.1)] The stance is to treat entire categories of belief as inherently radicalising. It is a familiar stance worldwide: officials lean on sweeping language that paints ideology itself as the chief culprit, even though a wide body of research underscores that personal grievances, social factors, and situational triggers—not simply beliefs—commonly fuel the path to violence. The only truly 'bad idea' here, then, is the persistent policy assumption that extremist ideology alone is what drives individuals to terrorism, a misconception that risks overshadowing more tangible causes and prompts misguided interventions that target expression rather than the advocacy of violence.

To challenge the state-centric assumption that counterterrorism necessitates ideological intervention, it is essential to distinguish the core concepts of radicalism, extremism, and terrorism before turning to research that explores the underlying drivers of terrorist involvement.

Radicalism

To be radical is, at its core, to reject the status quo—to demand fundamental change in how society operates.[10] But that rejection is not, in and of itself, dangerous. Many radicals do not seek to harm or destabilise society; rather, they confront systems of exclusion or injustice through disruptive but non-violent means.[10] The women's suffrage movement offers a striking example. Their calls for gender equality were, by the standards of the time, undeniably radical. Yet the movement's methods—protest, political pressure, and civil disobedience—were oriented towards expanding democratic participation, not dismantling it.[11] Framing radicalism as a precursor to terrorism obscures these distinctions and invites a policy response that risks suppressing legitimate calls for reform. In doing so, it not only distorts the historical record, but narrows the space for dissent in the present.

Of course, radicalism can shift toward extremism when the drive for change is accompanied by a rejection of democratic processes and a willingness to use violence.[10] The women's suffrage movement illustrates this trajectory: while its core ideology—advocating for women's rights and political equality—is now broadly embraced, some elements of the movement, particularly among the suffragettes, turned to more militant tactics, including arson and property damage.[11] These actions

marked a departure from radicalism into what might now be described as violent extremism. Crucially, however, it is not the ideology of women's rights that is remembered as dangerous, but the methods used by a radical minority to advance it, highlighting the importance of distinguishing between belief and behaviour—between ideas that challenge power and the actions that sometimes accompany them.

Extremism

Although there is no universally agreed definition of extremism, Lawlor[12] offers a valuable practitioner-informed model that captures the subtle transition from radicalism to extremism. By outlining a four-stage process—exposure, acceptance, idée fixe, and dissemination— Lawlor[12] illustrates how individuals might move from healthy radical thought toward deeper identity reconstruction and, ultimately, extremism. According to Lawlor,[12] *exposure* marks the initial moment a person encounters a radical worldview, while *acceptance* signals a growing embrace of that worldview as it begins to shape the individual's beliefs and perceptions. As internalisation deepens, *idée fixe* marks a turning point where commitment hardens into obsession, narrowing the individual's worldview, closing off alternative perspectives, and weakening their capacity for critical reflection. At this stage, the individual may no longer be seen as simply holding radical thought, but rather embracing extremism. Finally, practitioner concern peaked when signs of active dissemination emerged, signalling not only a deep internal shift but also a deliberate attempt to influence the thinking of others. Taken together, Lawlor's[12] findings offer a practitioner-informed perspective that highlights how extremism is not solely defined by the content of a person's beliefs, but by the intensity with which they are held, the loss of openness to other perspectives, and the effort to impose those beliefs on others—revealing how radical thought can harden into extreme belief, and in some cases, impact behaviour. While this may appear to contradict the earlier critique of ideology-focused approaches, the tension is intentional and will be revisited later, where a closer look at contributing factors offers a more nuanced understanding.

While extremism can include violence it does not mean violence is inevitable, an individual can be extreme and never turn to violence. The Scottish National Party (SNP) in the UK is a prime example with supporters of Scottish independence being labelled extremist in 2024 by then Prime Minister Rishi Sunak.[13] Despite misplaced intentions, Sunak was accurate in his assessment as some supporting Scottish independence exhibit features of extremism: a singular focus on independence, opposition to alternative constitutional arrangements, and the active dissemination of a radical political vision. Despite this, the SNP consistently rejects violence, opposing the presence of nuclear weapons, denouncing genocidal regimes, and advancing policies aimed at dismantling structural inequality.[14-16] In contrast, austerity measures implemented by successive UK governments have been linked to over 190,000 excess deaths.[17,18] Here, it is the non-extreme actor becomes the source of harm, while those labelled extremist champion ethical, non-violent alternatives. By framing extremism in this way, it becomes possible to recognise that not all unconventional or provocative beliefs are automatically a threat, reinforcing the argument that ideology is not the sole contributor to violence and that counter-extremism programmes should focus on behaviour rather than belief.

Terrorism

The distinction between behavioural extremism (acts) and cognitive extremism (beliefs) is particularly important when discussing terrorism. Like extremism, terrorism lacks a universally agreed definition. Although the definitional debate lies beyond the scope of this paper, terrorism is best understood not as an ideology or state of mind, but as an act of violent extremism. This paper adopts the definition of terrorism best described by English[19(p.24)] as:

> *"heterogeneous violence used or threatened with a political aim; it can involve a variety of acts, targets, and of actors; it possesses an important psychological dimension, producing terror or fear among a directly threatened group and also a wider implied audience in the hope of maximising political communication and*

achievement; it embodies the exerting and implementing of power, and the attempted redressing of power relations; it represents a subspecies of warfare, and as such it can form part of a wider campaign of violent and non-violent attempts at political leverage."

To complicate matters further, the pathways through which individuals become involved in extremism or violent extremism are highly individualised and context dependent. While early models conceptualise radicalisation as a linear progression, contemporary research consistently portrays it as a fluid and dynamic process, shaped by a complex interplay of factors that push individuals away from social conformity, pull them toward violent extremism, or protect them from becoming susceptible in the first place. Because of this, I will refrain from depicting a 'radicalisation process,' instead favouring a radicalism grid, as shown by Figure 1A.
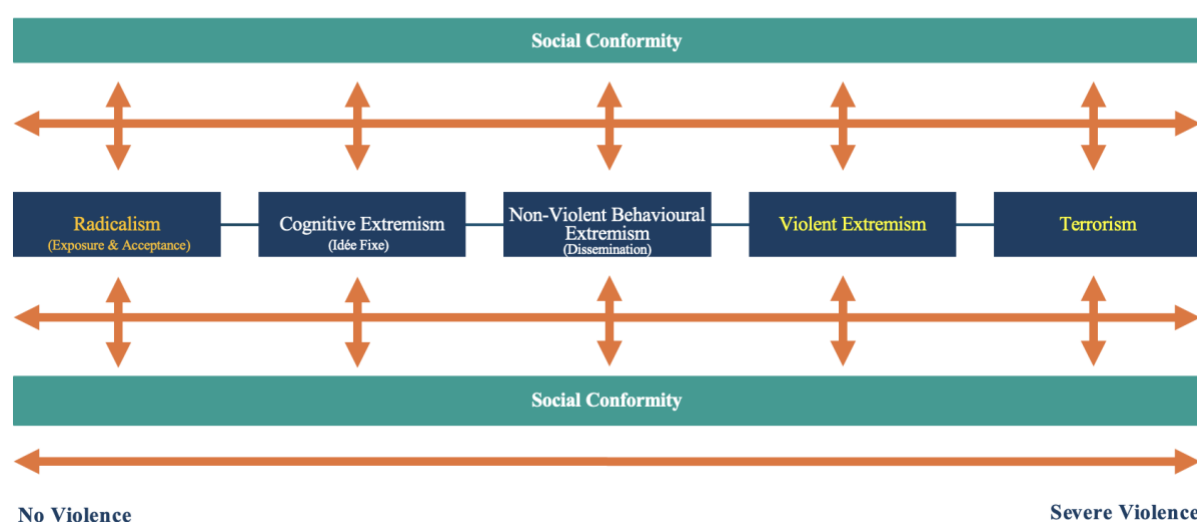
Radicalism Grid

Figure 1A

Although not a process, the grid is organised to reflect the varying degrees of violence associated with each concept of radicalism, with the act of terrorism representing the most severe manifestation of violent extremism. However, it is important to recognise that individuals may enter the grid at any point—there is no fixed sequence or linear pathway, and movement between positions can be fluid, fragmented, or even reversible.

Factors

As discussed in earlier sections, various factors can draw individuals away from social conformity toward radicalism and (violent) extremism, just as other influences can protect against such pathways. Understanding these dynamics is essential to any discussion of radicalisation and efforts to prevent political violence. This sub-section explores the key factors that shape an individual's susceptibility or resilience, challenging the claim that radical ideas are the primary drivers of violent extremism.

Risk & Protective Factors (RPF)

Understanding violent extremism requires an appreciation of the complex interplay of multiple risk and protective factors spanning different layers of an individual's social ecology.[20] At the individual level, research identifies a range of significant risk factors, including prior criminal history, thrill-seeking tendencies, unemployment, and authoritarian or fundamentalist beliefs.[21] Individuals prone to violent extremism often exhibit low self-control, impulsivity, and strong perceptions of injustice—factors that

fuel grievances and may motivate extremist actions.[21] Conversely, protective factors at this level, such as higher education, employment, intensive religious practices, self-control, adherence to law, and a nuanced value system (value complexity), play a crucial role in mitigating risks and fostering resilience.[21,22] Importantly, protective characteristics like law legitimacy, police legitimacy, and positive school experiences further bolster resistance to radicalisation.[23]

Moving to the relationship level, risk factors primarily involve negative influences within the immediate social circle, such as exposure to parental abuse, violent family dynamics, and radicalised peers, whether encountered in-person or online.[21,23] Critical family events, like divorce or bereavement, significantly heighten susceptibility by disrupting essential social support systems.[23] In contrast, relationship-level protective factors include family cohesion, parental involvement and monitoring, and positive interactions with prosocial peers. Having non-violent peers, significant others uninvolved in violence, and contact with diverse groups or foreigners significantly reduces the likelihood of engaging in violent extremism.[22,23]

While individual and relationship-level factors are comparatively well documented, De Silva et al.,[20] highlight the scarcity of empirical evidence addressing community and societal-level factors. Nevertheless, they suggest that community-level risk factors may include concentrated neighbourhood poverty, limited economic opportunities, community violence, and diminished social cohesion.[24] Conversely, community connectedness, social support networks, and proactive engagement can serve as protective factors against violent extremism.[24]

At the broadest, societal level, macro-level dynamics that can exacerbate radicalisation, such as ethnic tensions, discrimination, long-term socioeconomic inequalities, repressive governmental policies, weak state structures, and the exclusion of certain groups from political processes.[20,25] These factors reflect systemic issues that often underpin collective grievances, motivating extremist ideologies. Although explicit protective factors at the societal level are less clearly defined in existing literature, addressing structural inequalities and promoting inclusive governance are likely beneficial strategies for mitigating extremism.[25]

Critically, no single factor alone is sufficient to either cause or prevent violent extremism.[20] Instead, it is the cumulative and interactive effects of these multiple factors across different levels that shape individual trajectories toward or away from radicalisation and violent action. Consequently, effective prevention and intervention strategies must adopt an integrative, multi-level approach, recognising that fostering resilience at the individual level must be complemented by efforts to enhance relational bonds, strengthen community resources, and advocate systemic change at societal levels.[20]

Push & Pull Factors (PPF)

Understanding the factors that draw individuals towards violent extremism involves recognising the complex interplay of "push" and "pull" elements, which vary widely from person to person.[26,27] "Push factors" refer to negative social and personal influences that create susceptibility or openness towards extremist ideologies. At the societal level, these include experiences such as marginalisation, discrimination, limited access to opportunities, perceived injustice, and human rights abuses.[26] Such societal grievances may foster feelings of alienation and frustration, creating fertile ground for extremist narratives.[26,27] At the individual level, personal vulnerabilities—such as loneliness, isolation, grief, low self-esteem, a sense of disconnection, or unresolved trauma—can similarly heighten susceptibility to extremist appeals. These personal experiences often leave individuals searching for meaning, acceptance, or solutions to deeply felt grievances.[26,27]

In contrast, "pull factors" represent positive attractions offered by extremist groups that resonate with these individual vulnerabilities. Central among these is the sense of belonging, identity, and purpose provided by extremist communities.[26] Extremist narratives frequently offer romanticised portrayals of heroism, empowerment, or the illusion that personal and societal injustices can be effectively

understood and addressed through their worldview. Peer relationships, including charismatic influencers and group dynamics, play a significant role in enhancing the attractiveness of these narratives, particularly in digital spaces where online grooming, peer validation, and echo chambers amplify extremist beliefs and reinforce commitment.[27]

Engagement with radicalism and extremism typically arises from the interaction between Risk & Protective Factors and Push & Pull Factors. Rather than following a linear cause-and-effect trajectory, each individual carries a unique combination of influences (sequential key) that, under certain conditions, may align to create an environment conducive to (violent) extremism.[26,27]

Ultimately, the very concept of 'counter-extremism' is deeply problematic. By design, it involves state intervention into the ways individuals form their worldviews and identities when those views are deemed too extreme. Rather than targeting beliefs, policymakers should focus on addressing all forms of violence within society—especially violent extremism and, most critically, terrorism—through a holistic understanding of the diverse factors that contribute to, or protect against, the adoption of violence. Only, by approaching violent extremism with clarity and an openness to its many grey areas, can policymakers and practitioners hope to prevent violence while respecting the diversity of thought that underpins democratic life.

Extremism Online

Extremist movements ranging from far-right nationalists and jihadists to other violent groups—have cultivated multi-layered strategies for leveraging digital platforms to disseminate propaganda, radicalise sympathisers, and coordinate operations.[28,29] They produce polished, high-volume content designed to attract and engage audiences through multiple digital channels, including streaming platforms, forums, messaging apps, and social media.[2,30] Often relying on dramatic visuals and emotional narratives—whether in the form of glorified battle footage, martyrdom stories, or stylised memes—these materials blend anger, fear, or grievance toward targeted out-groups with coded references, ensuring content resonates with "insiders" while appearing benign to casual observers or automated detection systems.[31,32] By circulating messages that reinforce their worldview, they heighten the collective identity among sympathisers and reduce immediate detection of overtly hostile material.[4,33]

Additionally, groups can also cultivate one-on-one relationships with vulnerable individuals—often youths in precarious social or emotional conditions—to provide a personalised pathway into extremism.[34,35] Through private chats, direct messaging, and invitation-only groups, recruiters establish trust under the guise of mentorship or camaraderie.[2,7] They ultimately funnel individuals into secluded digital communities where the extremist worldview is validated and normalised through symbolism, shared rituals, and constant group reinforcement.[29,36] Researchers note that once such bonds form, radical beliefs crystallise and can spill offline.[1,35,36]

This online immersion also assists with operational coordination, wherein activities ranging from protest and propaganda distribution to more clandestine violent attacks are planned.[1,34] Through closed or encrypted channels, violent groups exchange operational knowledge such as weapon-making instructions, bomb assembly guides, and secure communication tactics.[37,38] Research on jihadi networks show how bots are utilised to propagate extremist messages and draw unsuspecting users from relatively moderate extremist posts into more radical material.[39] By restructuring or reposting propaganda, these automated accounts reduce the human work required for extremist campaigns while reinforcing a sense of collective mission for participants.[39]

A critical factor in extremist success is the algorithmic architecture of popular platforms, which often promote provocative and emotionally charged posts to increase user interaction.[29,40] Extremist actors exploit recommendation systems by crafting sensational content that algorithms may amplify, drawing susceptible users toward increasingly radical material.[2,33] As people engage with such content, they quickly become enclosed within echo chambers, where the same narratives reverberate, reinforcing

extreme worldviews over time.[4,32] This cyclical nature of repeated exposure, confirmation bias, and group validation is crucial to encouraging extremist mindsets.[31,35]

These tactics are further strengthened by exploiting the internet's transnational nature, which allows extremists to forge global alliances.[37,41] Far and extreme-right networks unite around conspiratorial beliefs—such as white supremacy or anti-immigrant rhetoric—while jihadist organisations rally around calls for a global ummah, each employing narratives which exploit risk and push factors.[6,28] This transnational dimension consolidates extreme individuals, enhancing emotional solidarity, while easing propaganda circulation.

Finally, anonymity and the use of synthetic identities allow extremists to orchestrate large-scale disinformation or infiltration efforts.[42,43] Many extreme individuals operate under pseudonyms or deploy multiple accounts, making it difficult for platforms and law enforcement to identify real actors.[6,41] Such covert measures let them infiltrate mainstream conversations, portray themselves as average users, and manipulate trending topics, especially when others cannot distinguish genuine discourse from orchestrated propaganda.[34,36] By navigating these digital ecosystems with agility, extremists maximise their reach, evade detection, and strengthen both their operational security and narrative power.

All of these overlapping strategies demonstrate how adept extreme groups are at harnessing online opportunities to expand their influence.[28,40] The exploitation of bots and algorithms ensures content remains visible, while encrypted messaging services sustain a sense of community.[1,3] Consequently, law enforcement, technology companies, policymakers, and civil society must collaborate constructively to address the multi-level tactics that underpin online extremist behaviour. Merely removing and criminalising inflammatory content is insufficient: deeper policy and societal responses are required to counteract the factors that fuel extremist recruitment and violence.

Online Counter-Extremism

One prominent approach to tackling extremism online is the criminalisation of speech. These laws extend existing hate speech or terrorism provisions to cyberspace, where content, once published, may reach audiences on a global scale within seconds. This approach reflects heightened anxiety by states about the role of digital platforms in amplifying radical ideas.[4,44] Germany's Network Enforcement Act (NetzDG), for instance, obliges large social media companies to remove "manifestly unlawful" material within 24 hours or face substantial fines.[4,44] Its supporters contend that such rapid action disrupts real-time virality and deters the formation of extremist echo chambers.[32] Yet, critics highlight the risk of over-removal by private firms who lack the judicial mandate to distinguish harmful but lawful expression from that which incites violence.[2,29] Moreover, operators located abroad can resist compliance with such directives, revealing how enforcement challenges transcend borders in an era when extremist ideas travel online with ease.[4]

In the United Kingdom, the Online Safety Act epitomises another regulatory pathway by imposing a "duty of care" on platforms which require proactive monitoring to shield users from harmful or radical content.[5,34] Proponents emphasise the protective rationale, particularly where children or vulnerable groups may be exposed to extremist materials that normalise hatred.[34] Despite such intentions, detractors question whether the Act's scope may be overly broad, encompassing forms of "offensive speech" that do not cross a threshold of incitement.[2,29] Questions also remain as to whether automated detection tools can reliably parse the evolving language of hate, given that extremists frequently mask their rhetoric with coded references or stylised symbols.[1,45]

Legislative urgency often spikes following major terrorist incidents or harrowing hate crimes, prompting governments to enact laws with minimal public consultation.[46] As a result, these statutes sometimes rest on limited empirical evidence, complicating any assessment of whether they effectively reduce extremism or merely shift it elsewhere.[6,47] In practical terms, even rigorous regulations hinge on

robust enforcement, requiring well-funded oversight agencies, extensive platform cooperation, and cross-industry data-sharing.[2,44] Absent such mechanisms, content can reappear on platforms that lack the resources or motivation for strict moderation.[1,45]

This adaptive behaviour highlights the complexity at the heart of digital counter-extremism. Extreme groups modify their language, adopt ambiguous terms, and employ insider codes to camouflage hateful messages from content flags.[31,32] When platforms do enforce takedowns, extremist networks typically respond by establishing backup accounts, copying content, and moving to alternative online spaces with lighter moderation.[6,40] These repeat appearances underscore that removal alone, however swift, cannot guarantee long-term disruption of extremist ecosystems.[48,49] Instead, it spurs a perpetual cycle of detection and evasion, placing significant pressure on moderators who must remain attuned to changing rhetorical trends.[1]

Against this backdrop, scholars point to a deeper ethical quandary: whether it is legitimate to conflate extreme belief with incitement, and if penalising the former jeopardises core principles of freedom of expression.[2,3] Traditional democratic frameworks prioritised criminalising concrete acts—such as violent plots or direct incitement—rather than the mere possession of radical ideas.[30,50] Yet, the offline impact of online extremist networks prompted a legislative shift against broader categories of content deemed harmful to society.[34] Critics argue that this change risks chilling legitimate debate and investigative journalism, as well as discouraging community-led counter-speech and open debate against extreme narratives.[40]

Discussions around "thought crime" become all the more urgent in authoritarian or semi-authoritarian contexts, where governments cite counter-extremism to clamp down on political dissent, labelling opposition figures or independent journalists as "extremist" threats.[43,51] Russia's frequent invocation of "extremism" charges against critics and Zimbabwe's broad application of comparable statutes illustrate how vaguely defined laws can be manipulated to suppress dissent.[40,43] Even in democratic nations, ambiguous categories like "gross offensiveness" or "harmful but not illegal" raise fears that lawful expressions might be curtailed on the basis of subjective interpretation.[5,36] These developments prompt reflection on whether such far-reaching regulations can be reconciled with foundational commitments to open, pluralistic discourse.[41,44]

It is, however, equally clear that online extremism can pose tangible risks to public safety and social cohesion.[2,5] Terror plots are repeatedly linked to online radicalisation, persuading policymakers that pre-emptive regulation is necessary.[34] Yet, the central challenge remains: laws targeting violent conspiracies or direct incitement appear justifiable, whereas approaches that encroach controversial, even extreme opinion risks eroding free expression and open debate.[30,50] Moreover, driving extremist networks into clandestine spaces might hamper law enforcement and researchers' ability to understand the trajectory of radicalisation, potentially making it harder to intervene effectively.[3,31]

Ultimately, the consensus emerging in much of the literature is that robust yet narrowly tailored legal measures against the incitement or glorification of violence, complemented with societal resilience building, offer the best prospect for curbing susceptibility to extreme worldviews without stifling legitimate public debate.[6,46] Approaches that penalise overt calls to violence and active conspiracies align more closely with traditional legal safeguards and, crucially, do not treat all hateful or offensive ideas as criminal in themselves.[30,50] In this view, striking a balance between safety and liberty is neither a purely legal nor purely technological endeavour, but rather an ongoing ethical and social challenge in a digital landscape that perpetually evolves.

Fostering Resilience through Education, Empowerment, and Dialogue (FREED)

A compelling alternative to expansive online speech controls emerges from empirical insights into the drivers of radicalisation. Rather than criminalising broad categories of non-violent expression, this approach restricts legal sanctions to explicit incitement or glorification of violence.[52] Research consistently shows that criminalising hateful but non-violent content neither prevents radicalisation nor upholds ethical standards around free expression, and instead displaces extremist activity into hidden corners of the internet, making it more difficult to monitor.[1] By recalibrating legislation to focus on threats of violence, governments can redirect resources toward tackling actual conspiracies or plots rather than waging endless battles against vague "offence" categories.[30] This reorientation also diminishes the risk of misuse, whereby regimes justify suppressing dissent by labelling legitimate opposition or inconvenient activism "extremist."[43]

Narrower legal thresholds do not equate to a laissez-faire stance on hateful content. Social media companies retain a responsibility to remove direct advocacy of violence, terrorism, or explicit threats.[2] Early observations suggest that when platforms rely on transparent moderation standards and accessible appeals procedures, users become more confident in reporting genuinely dangerous content, and extremist narratives of state persecution lose their potency.[1] Such nuanced governance ensures that legal interventions remain sharply focused on violent agendas.[4,44]

A second, more substantive pillar of this model calls for robust community-focused programmes grounded in research on risk and protective factors, as well as the push and pull elements that together steer individuals towards—or away from—violent extremism.[20,21,27] Evidence confirms that alienation, economic exclusion, identity crises, and social isolation are key pathways through which extremist rhetoric resonates.[23] Addressing these pressures requires tangible social investment in communities, from mental health support and employment opportunities to mentoring schemes and civic initiatives that cultivate belonging and resilience.[21] When governments and civil society prioritise social care, education, and equality, they address the vital foundations of extremist susceptibility rather than fixating on the content of any specific ideology.[8] This strategy reinforces the notion that although ideology may be present, it is not the principal motive that draws individuals to violent extremism.[8]

Programmes that equip young people and adults with online critical skills are especially important, given the agility with which extremist networks exploit digital tools.[26,50] Several pilot initiatives show that citizens who are forewarned about manipulative tactics—such as coded propaganda, friendship-based recruitment, and algorithmic echo chambers—are less likely to be drawn in.[53] States can launch free, immediately accessible courses that teach users how to recognise misinformation and hate narratives, and to distinguish these from legitimate forms of political, religious, or social critique.[36] Such initiatives operate at minimal cost and enhance social resilience by creating informed communities capable of challenging extremist claims in the open.[28]

This strategy of de-regulating speech and investing in communities not only makes law enforcement more effective by allowing it to focus on impending threats of violence, but also preserves the open discourse essential to democratic life.[5] Rather than suppressing radical claims into invisibility—and thereby playing into extremist narratives of persecution—states can encourage constructive public critique that exposes the flaws of such ideas.[32] By upholding transparent, narrowly defined regulations, states offer fewer pretexts for the crackdown on dissent under the guise of counter-extremism.[43]

Ultimately, the FREED framework harnesses evidence-based findings on radicalisation to advance two mutually reinforcing outcomes. First, it concentrates regulatory power where it is most necessary—on genuine advocacy of violence—rather than on intangible offence. Second, it mobilises social investment to tackle the well-established risk and protective factors that drive or deter extremism.[54] Moreover, these proposals are by no means revolutionary, yet governments, fully aware of this evidence, continue to pursue punitive measures against ideology. This combination not only reduces the chances of overreach and undue censorship but also promotes community resilience and quality of life by focusing on the barriers people face.[20] FREED represents a shift from state-centred ideological

oversight towards a more humane and pragmatic effort to build robust, united societies where hateful narratives fail to take root. Through this rebalanced emphasis, societies can tackle violent extremism at its behavioural core whilst safeguarding the free exchange of ideas upon which healthy democracies depend.

Conclusion

In navigating the nexus of radicalism, extremism, and terrorism, this paper has highlighted the critical importance of making clear distinctions between belief and behaviour, non-violent dissent and violent action. Common policy approaches, shaped by state-centric assumptions that ideology alone motivates terrorism, risk obscuring the complex interplay of personal, social, and structural factors that steer individuals toward or away from violence.

By reframing radicalism not as an inevitable precursor to terrorism but as a potentially constructive force for challenging injustice, we create space for differentiating legitimate dissent from coercive extremism. Similarly, the concept of extremism need not be viewed as synonymous with violence: an individual or group may hold intense, unorthodox beliefs without ever crossing the threshold into terrorism. An overreliance on criminalising non-violent thought can thus serve to push vulnerable individuals into the shadows, undermining efforts to mitigate violence.

Ultimately, the evidence underscores that violent extremism and terrorism emerge through a constellation of interrelated push and pull factors, amplified by social contexts and online ecosystems. In this light, a focus purely on the content of one's beliefs—devoid of the lived realities that drive alienation, exclusion, and frustration—misses the root causes of harm. Policymakers and practitioners stand to benefit from centring their efforts on tangible, community-level solutions: addressing economic marginalisation, strengthening supportive social networks, and investing in education and mental health interventions that fortify individual and collective resilience. A more impactful model emerges through the FREED framework, which channels resources toward education, empowerment, and social investment while reserving punitive measures for explicit advocacy or acts of violence.

Notes

1.

Scrivens R, Wojciechowski TW, Freilich JD, Chermak SM, Frank R. Comparing the Online Posting Behaviors of Violent and Non-Violent Right-Wing Extremists. Terrorism and Political Violence. 2023;35(1):192–209.

2.

Kim M. Online gender-based hate speech in South Korea. International Review of Law, Computers & Technology. 2024;38(2):140–9.

3.

Kohl U. Toxic recommender algorithms: immunities, liabilities and the regulated self-regulation of the Digital Services Act and the Online Safety Act. Journal of Media Law. 2024;16(2):301–35.

4.

Katsirea I. "Fake news": reconsidering the value of untruthful expression in the face of regulatory uncertainty. Journal of Media Law. 2018;10(2):159–88.

5.

Kalliris K. Online harm, free speech, and the "legal but harmful" debate: an interest-based approach. Journal of Media Law. 2024;16(2):390–416.

6.

Price L. Platform responsibility for online harms: towards a duty of care for online hazards. Journal of Media Law. 2021;13(2):238–61.

7.

Mangan D. Regulating for responsibility: reputation and social media. International Review of Law, Computers & Technology. 2015;29(1):16–32.

8.

Richards A. The Broadening Remit of UK Counter-Terrorism - from Terrorism to "Radicalization" to "Extremism": C-T Imperative or Loss of Focus?. International Affairs. 2015;91(2):371–80.

9.

Task Force on Tackling Radicalisation and Extremism. Tackling extremism in the UK Report from the Prime Minister's Task Force on Tackling Radicalisation and Extremism [Internet]. https://assets.publishing.service.gov.uk. HM Government; 2013 [cited 2025 May 28]. Available from: https://assets.publishing.service.gov.uk/media/5a7c4ff6ed915d338141df89/ETF_FINAL.pdf

10.

Bartlett J, Miller C. The Edge of Violence: Towards Telling the Difference Between Violent and Non-Violent Radicalization. Terrorism and Political Violence. 2012;24(1):1–21.

11.

Bearman CJ. An Examination of Suffragette Violence. The English Historical Review. 2005 Apr 1;120(486):365–97.

12.

Lawlor E. "But what if I get it wrong?" Exploring practitioners' understanding of preventing and countering violent extremism and radicalisation duty guidance. In: Silva TC, Kordaczuk-Wąs M, editors. Policy Press eBooks. Bristol: Policy Press; 2024. p. 216–43.

13.

Andrews K. Sunak includes independence supporters in list of extreme groups [Internet]. Thetimes.com. The Times; 2024 [cited 2025 May 28]. Available from: https://www.thetimes.com/uk/scotland/article/sunak-includes-independence-supporters-in-list-of-extreme-groups-j9wvkw6ml

14.

BBC. Scottish government suspends meetings with Israel. BBC News [Internet]. 2024 Aug 19 [cited 2025 May 28]; Available from: https://www.bbc.co.uk/news/articles/c5ydg5lqgwxo

15.

SNP. Delivering Progress — Scottish National Party [Internet]. Scottish National Party. SNP; 2019 [cited 2025 May 28]. Available from: https://www.snp.org/record/

16.

SNP. What is the SNP position on nuclear weapons? [Internet]. Scottish National Party. 2025 [cited 2025 May 28]. Available from: https://www.snp.org/policies/what-is-the-snp-position-on-nuclear-weapons/

17.
Balls K, Simmons M. "Austerity is back": Inside Labour's emergency budget. The Spectator [Internet]. 2025 Mar 20 [cited 2025 May 28]; Available from: https://www.spectator.co.uk/article/is-there-any-way-out-of-the-low-growth-trap/

18.
Berman Y, Hovland T. The cost of austerity: How UK public spending cuts led to 190,000 excess deaths [Internet]. LSE Inequalities. 2024 [cited 2025 May 28]. Available from: https://blogs.lse.ac.uk/inequalities/2024/06/19/the-cost-of-austerity-how-spending-cuts-led-to-190000-excess-deaths/

19.
English R. Terrorism : How to Respond. Oxford: Oxford University Press; 2009.

20.
Da Silva C, Amadio N, Sarg R, Domingo B, Benbouriche M. Risk and Protective Factors Associated with Violent Extremism: A Multilevel and Interdisciplinary Evidence-Based Approach. Perspectives on terrorism. 2024 Sep 27;18(3).

21.
Wolfowicz M, Litmanovitz Y, Weisburd D, Hasisi B. Cognitive and behavioral radicalization: A systematic review of the putative risk and protective factors. Campbell Systematic Reviews. 2021 Jul 20;17(3).

22.
Lösel F, Bender D, Jugl I, King S. Resilience against Political and Religious Extremism, Radicalization, and Related Violence: A Systematic Review of Studies on Protective Factors. In: Weisburd D, Savona EU, Hasisi B, Calderoni F, editors. Understanding Recruitment to Organized Crime and Terrorism. Cham: Springer,; 2020. p. 55–84.

23.
Zych I, Nasaescu E. Is radicalization a family issue? A systematic review of family-related risk and protective factors, consequences, and interventions against radicalization. Campbell Systematic Reviews. 2022 Jul 20;18(3).

24.
Armstead TL, Wilkins N, Doreson A. Indicators for Evaluating Community- and Societal-Level Risk and Protective Factors for Violence Prevention. Journal of Public Health Management and Practice. 2018;24:S42–50.

25.
LaFree G, Schwarzenbach A. Micro and macro-level risk factors for extremism and terrorism: Toward a criminology of extremist violence. Monatsschrift für Kriminologie und Strafrechtsreform. 2021 Jul 1;104(3):184–202.

26.
Step Together. Push and pull factors in violent extremism [Internet]. Step Together - Countering Violent Extremism (CVE). 2024 [cited 2025 May 24]. Available from: https://steptogether.nsw.gov.au/violent-extremism/push-and-pull-factors-in-violent-extremism.html

27.
Meines M, Ranstorp M. The Root Causes of Violent Extremism -THE BASICS. European Commission ; 2024.

28.
Miller-Idriss C. Chapter 6. Weaponizing Online Spaces. In: Hate in the Homeland: The New Global Far Right. Princeton University Press; 2020. p. 138–60.

29.
Munn L. More than a mob: Parler as preparatory media for the U.S. Capitol storming. First Monday. 2021 Feb 7;26(3):1–16.

30.
Schmitt JB, Rieger D, Rutkowski O, Ernst J. Counter-messages as Prevention or Promotion of Extremism?! The Potential Role of YouTube Recommendation Algorithms. Journal of Communication. 2018;68(4):780–808.

31.
Odag Ö, Leiser A, Boehnke K. Reviewing the Role of the Internet in Radicalization Processes. Journal for Deradicalization. 2019;21:261–300.

32.
Kohl U. Islamophobia, "gross offensiveness" and the internet. Information & Communications Technology Law. 2017;27(1):111–31.

33.
Siegel AA. Online Hate Speech. In: Tucker JA, Persily N, editors. Social Media and Democracy: The State of the Field, Prospects for Reform. Cambridge: Cambridge University Press; 2020. p. 56–88.

34.
Kenyon J, Binder J, Baker-Beall C. Understanding the Role of the Internet in the Process of Radicalisation: An Analysis of Convicted Extremists in England and Wales. Studies in Conflict & Terrorism. 2024;47(12):1747–71.

35.
Lancheros M, Atanasova T, Laganaro M. When does speech planning rely on motor routines? ERP comparison of speech and non-speech from childhood to adulthood. Language, Cognition and Neuroscience. 2023;38(8):1115–32.

36.
Nemes I. Regulating Hate Speech in Cyberspace: Issues of Desirability and Efficacy. Information & Communications Technology Law. 2002;11(3):193–220.

37.
Moro N, Aikat D "Deb". Liberty v. Libel: Disparity and Reconciliation in Freedom of Expression Theory. First Amendment Studies. 2013;47(1):58–83.

38.
Reeve Z. Repeated and Extensive Exposure to Online Terrorist Content: Counter-Terrorism Internet Referral Unit Perceived Stresses and Strategies. Studies in Conflict & Terrorism. 2023;46(6):888–912.

39.
Alrhmoun A, Winter C, Kertész J. Automating Terror: The Role and Impact of Telegram Bots in the Islamic State's Online Ecosystem. Terrorism and Political Violence. 2024;36(4):409–24.

40.
Tworek HJS. Fighting Hate with Speech Law: Media and German Visions of Democracy. The Journal of Holocaust Research. 2021;35(2):106–22.

41.
Marsoof A, Luco A, Tan H, Joty S. Content-filtering AI systems–limitations, challenges and regulatory approaches. Information & Communications Technology Law. 2023;32(1):64–101.

42.
Smith BH. To Filter or Not To Filter: The Role of the Public Library in Determining Internet Access. Communication Law and Policy. 2000;5(3):385–422.

43.
Mugari I. The dark side of social media in Zimbabwe: Unpacking the legal framework conundrum. Cheng K, editor. Cogent Social Sciences [Internet]. 2020 [cited 2025 May 24];6(1). Available from: https://doi.org/10.1080/23311886.2020.1825058

44.
Kohl U. Platform regulation of hate speech – a transatlantic speech compromise? Journal of Media Law. 2022;14(1):25–49.

45.
Chang B. From Internet Referral Units to International Agreements: Censorship of the Internet by the UK And EU. Columbia Human Rights Law Review. 2017;49(2):114–212.

46.
Banchik AV. Disappearing acts: Content moderation and emergent practices to preserve at-risk human rights–related content. New Media & Society. 2021;23(6):15271544.

47.
Bunting M. From editorial obligation to procedural accountability: policy approaches to online content in the era of information intermediaries. Journal of Cyber Policy. 2018;3(2):165–86.

48.
Siyuan C. Regulating online hate speech: the Singapore experiment. International review of law computers & technology. 2024;38(2):119–39.

49.
Montalbano K. Reimagining Section 230 and Content Moderation: Regulating Incivility on Anonymous Digital Platforms. Communication Law and Policy. 2023 Feb 2;27(3-4):187–219.

50.

Amarasingam A. Belonging is just a click away: Extremism, radicalisation, and the role of online communities. In: Malkki L, Busher J, Marsden S, editors. The Routledge Handbook on Radicalisation and Countering Radicalisation. Routledge; 2023. p. 196–212.

51.
Workneh TW. Digital cleansing? A look into state-sponsored policing of Ethiopian networked communities. African Journalism Studies. 2015;36(4):102–24.

52.
Neumann P. Countering Online Radicalization in America [Internet]. biplrtisanpolicy.org. Homeland Security Project; 2012 [cited 2025 May 23]. Available from: https://bipartisanpolicy.org/download/?file=/wp-content/uploads/2019/03/BPC-_Online-Radicalization-Report.pdf

53.
Littler M, Lee B. Introduction. In: Littler M, Lee B, editors. Digital Extremisms: Readings in Violence, Radicalisation and Extremism in the Online Space. 1st ed. Cham: Springer International Publishing; 2020. p. 1–11.

54.
Meines M, Ranstorp M. The Root Causes of Violent Extremism -THE BASICS. European Commission ; 2024.