CTRL+ALT+COLLABORATE

PUBLIC-PRIVATE PARTNERSHIPS TO PREVENT EXTREMISM IN GAMING

Galen Lamphere-Englund and Menso Hartgers, RAN Policy Support

February 2024

Radicalisation Awareness Network
RAN Policy Support

Luxembourg: Publications Office of the European Union, 2023

TABLE OF CONTENTS

LIST OF ACRONYMS

EGRN	Extremism and Gaming Research Network (EGRN)
EUIF	EU Internet Forum (EUIF)
GIFCT	Global Internet Forum to Counter Terrorism
P/CVE	Preventing and Countering Violent Extremism
PPP	Public-Private Partnership: The European Union Agency for Cybersecurity (formerly the European Network and Information Security Agency, ENISA) defines a Public-Private Partnership (PPP) as “an organised relationship between public and private organisations, which establishes common scope and objectives, and uses defined roles and work methodology to achieve shared goals” (ENISA, 2011).
TAT	Tech Against Terrorism (TAT)

KEY FINDINGS

This report aims to guide policymakers and private sector companies in understanding and addressing the increasingly urgent issue of online extremism within gaming platforms. Through a multidisciplinary approach, this report analyses key challenges and recommends actionable steps to foster partnerships across both public and private stakeholders. To do so, the report draws on 17 primary interviews carried out by the authors in 2023 with experts from industry, the public sector, and civil society, along with a comprehensive review of existing literature and case studies.

KEY FINDINGS

- **The Pervasiveness of Extremism:** Gaming platforms are not immune to extremist content and radicalising activities, and widespread toxicity contributes to a risk of radicalisation of gamers. The problem is far-reaching and cannot be ignored.
- **Public Sector's Limited Awareness:** There is a noticeable gap in the public sector's understanding of online gaming spaces, contributing to inefficiencies in counter-extremism efforts.
- **Private Sector's Role and Limitations:** Private gaming companies often either ignore the issue or are inadequately equipped to address extremism. Many lack specific policies and operational guidelines.
- **Complexity of Content Moderation:** Neither human-led nor algorithmic moderation alone can sufficiently manage extremist content. A hybrid approach is essential.
- **Lack of Regulatory Clarity**: Inconsistent definitions of extremism, hate speech, and terrorist activities across jurisdictions hinder effective policy development.

RECOMMENDATIONS

1. **Policy Formulation and Implementation:** Both public and private sectors should work on developing clear policies and trust-building dialogues targeting extremism.
2. **Trust and Transparency:** Adopt a transparent reporting system of toxic, targeted harassment, or extremist behaviours with comparable metrics and accessible data between gaming platforms, authorities, and civil society, mandated by regulatory bodies and regulations such as the DSA (in particular, Art. 17).
3. **Educating about Gaming:** Leverage expertise from both sectors to design educational curricula that empower users, regulators, and platform employees to recognise extremist behaviours.
4. **Multi-Stakeholder Partnerships:** Foster collaboration between the public, private, and civil sectors, focusing on areas like strategic communications, education, and technical assistance.
5. **Standardisation Across Platforms:** Develop industry standards to streamline content moderation practices, similar to ISO frameworks.
6. **Balance Privacy and Safety:** Promote the implementation of human rights safeguards in the moderation of online content. Strict moderation can compromise user privacy and freedom of speech. Ethical and human rights considerations are therefore paramount.

EXECUTIVE SUMMARY

With the advent of gaming as a popular medium of entertainment, extremist activities in these online spaces are becoming increasingly prevalent. Online games in particular offer extremists with opportunities to communicate and engage in recruitment and outreach. Although there is no evidence to suggest that gaming spaces are markedly more prone for exploitation by extremists, policymakers should be attuned to the unique challenges that come with the hyper-interactive nature and the global reach of gaming platforms. The mutual exchange of transparent research and enhanced avenues for dialogue can contribute to a unified understanding of both challenges and viable solutions, thereby strengthening the collective impact of counter-extremism measures.

This is where public policies can serve as a complement to private initiatives. For example, public authorities can (co-)fund research (together with the private sector) that provides a nuanced understanding of how extremist ideologies proliferate on these platforms. Similarly, inter-agency cooperation and cross-platform standardisation of terms of services or measures aimed at mitigating online harms amplify the industry's overall capacity to fight extremism, facilitating the swift implementation of cutting-edge technologies and tools. This collaboration between public actors and private entities could manifest itself in several ways: information sharing, co-development of moderation technology, and Public-Private Partnerships (PPPs) to fund digital literacy programmes, among others. The authors also heard from interviewees a pressing need for clear regulatory frameworks that address the nuances of gaming and which also foster accepted guidelines and norms for content moderation activities. Finally, this report highlights an opportunity to augment preventative measures through positive P/CVE interventions and targeted training for both regulators and platforms is essential. In this regard, the private sector can offer valuable expertise in developing educational programmes, while law enforcement and regulatory agencies can help industry players grasp the nuances of extremist risks. Altogether, this integrated strategy emphasises the indispensable roles both public and private actors play in fostering a more secure and inclusive digital gaming landscape. As expanded upon in Chapter Two, this research has identified four core pillars for Public-Private Partnership involvement on the part of public entities:

1. **Understand** through research-focused PPPs that better evaluate risks and challenges of extremism on gaming platforms.
2. **Support** dialogue efforts for public and private actors so they can better come together and jointly decide how to address challenges.
3. **Regulate** via improved and nuanced requirements for limiting extremist exploitation of gaming platforms.
4. **Prevent** through developing partner-led P/CVE programming that can be deployed against exploitation taking place in gaming. Moreover, collaboration can be encouraged in domains like strategic communications, positive interventions, and technical assistance to platforms.

This research also finds that while the gaming industry needs to take far more action to enhance moderation and reporting systems, their efforts can only go so far. They require external assistance in the form of nuanced policy, positive intervention programming, technical support, and standardised guidelines that can be most effectively provided by public bodies or third-party organisations specialised in counter-terrorism and extremism. That said, these companies can start fostering a secure gaming atmosphere by openly recognising and condemning extremist activities. This public stance serves as the bedrock for comprehensive strategies, the allocation of resources, and educational initiatives focused on tackling extremist material. Three main pillars constitute potential recommendations for platforms' engagement:

1. **Sponsoring partnerships to create better in-game environments via education** that empower users to recognise and report extremist content. This education should include information on critical thinking, fact-checking, and responsible online behaviours. Governments should collaborate with industry and civil society to create educational curricula around digital literacy, focusing on the risks posed by extremist content.
2. **Cooperate with external entities**, advocating for increased collaboration with external organisations, such as NGOs, research institutions, and government agencies, can secure additional expertise and resources in the fight against extremism.
3. **Cooperation and standardisation across platforms**, where private gaming companies should work together to share best practices, insight, and technologies for combatting extremism. This collaborative approach can lead to more effective solutions across the industry.

Having a multi-sectoral, synergistic approach to countering extremism in online gaming is not just a 'good-to-have,' but a strategic imperative. Extremists are often ahead of the curve in utilising emerging technologies for recruitment and radicalisation, from print media used by anarchists in the late 1800s, to novel online influence operations by al-Qaida 20 years ago, to livestreaming far right attacks from Christchurch in 2019 to today (Hoffman. 2017; West, 2021). Public and private sectors, operating in silos, are perpetually catching up. By working in tandem, they stand a better chance of pre-empting extremist strategies rather than being in a constant state of reaction.

The central tenet of the report is the collective responsibility that private firms and governments alike bear in the fight against online extremism, especially in the digital playgrounds that gaming platforms have become. While it might be tempting for policymakers to put the onus on the private sector alone, or vice versa, such approaches are fundamentally flawed. They are reductive in treating a complex, multifaceted problem as one that can be solved through unilateral actions. This report provides a roadmap to forge such partnerships and make online gaming spaces resilient against extremist exploitation.

INTRODUCTION

The nexus between gaming and extremism has become increasingly apparent over the last few years. While extremists have built and exploited games since at least the late 1990s (Lamphere-Englund and White, 2023), violent extremists have stepped up their abuse of gaming and gaming-adjacent platforms since 2019. They post manifestos, livestream attacks, and share or create harmful content of an extremist nature, such as memes, bespoke video games, and modifications ('mods') to existing games. In response, the corpus of knowledge on this phenomenon has steadily grown. In the context of the Radicalisation Awareness Network (RAN), a number of papers have recently been written on the role that games, gaming culture, and gaming(-adjacent) platforms play in the proliferation of extremism. A selection of these publications include: Verdegaal's et al. (2020) *Extremists' Use of Video Gaming – Strategies and Narratives*; Schlegel's (2021b) *Extremists' use of gaming (adjacent) platforms: Insights regarding primary and secondary prevention measures* and (2021b) *The gamification of violent extremism & lessons for P/CVE*; Lakhani's (2022) *Video Gaming and (Violent) Extremism*; and Lakhani, White, and Wallner's (2022) *The Gamification of (Violent) Extremism*.

Outside the RAN, several policy, practitioner, and research institutes work to identify, prevent, and combat extremism in video games and related platforms. The EU Internet Forum (EUIF), Europol's Internet Referral Unit (IRU), the Extremism and Gaming Research Network (EGRN), Tech Against Terrorism (TAT), and the Global Internet Forum to Counter Terrorism (GIFCT) are but a few prominent examples of these. While capacity-building and knowledge-sharing efforts are on the rise, knowledge gaps in the existing body of research certainly exist. Questions related to understanding multiplayer games and platforms as communication channels, terrorism finance through gaming platforms, and the behaviour of younger gamers and the implications for VE and P/CVE figure among these gaps (Lamphere-Englund & Bunmathong, 2021). Nevertheless, a consolidated understanding of the threat landscape is slowly emerging, thus paving the way for formulating actionable policy recommendations.

One particular issue concerns the collaboration of public actors and private gaming(-adjacent) companies to address extremism in and around video games. Hartgers (one of the authors of this current paper) and Leidig (2023) note that moderation and monitoring of gaming spaces cannot be relegated to public entities like intelligence and law enforcement agencies alone. Gaming spaces are incredibly opaque – lacking public application programming interfaces (APIs) and data access mechanisms – and are often relatively hidden away from such agencies and academic researchers. The recent case of a U.S. Airman leaking classified intelligence over the course of several years in a Discord server popular with a group of gamers demonstrates this issue (Lamphere-Englund and White, 2023). Moreover, with around an estimated three billion gamers worldwide, public actors lack the resources to monitor gaming spaces comprehensively, and privacy concerns abound even if the technical capacity can be found (Newzoo, 2022a). Therefore, this report argues that impactful P/CVE efforts in the gaming sphere are contingent on the effective collaboration of public and private entities.

To this end, this RAN Policy Support paper explores the concept of Public-Private Partnerships and provide recommendations on how these can be adapted to gaming contexts.[1] The report draws on the authors' experience working on prevention and research initiatives on gaming platforms and findings from 17 in-depth interviews with knowledgeable public and private actors across law enforcement and regulatory agencies, social media companies, game developers, and mental health researchers. Their expertise and a review of successful

[1] A definitional note: The European Union Agency for Cybersecurity (formerly the European Network and Information Security Agency, ENISA) defines a Public-Private Partnership (PPP) as "an organised relationship between public and private organisations, which establishes common scope and objectives, and uses defined roles and work methodology to achieve shared goals" (ENISA, 2011). While other definitions of PPPs exist, in this report we apply the ENISA working definition as authoritative, given the organisation's status as an EU agency with a remit in cyberspace.

PPP approaches inform the policy recommendations in this paper, setting out comprehensive principles for establishing effective PPPs. This report is structured so as to provide an overview of the field. This is followed by guidance targeting the public sector, and lastly, the private sector.

METHODOLOGY

This paper applies a two-part methodology. First, the authors carried out a comprehensive overview of the existing literature on Public-Private Partnerships. This literature review aimed to first define what a PPP is and what configurations of PPPs may be most relevant for gaming contexts. The literature was selected using a keyword search strategy on Google Scholar and online university libraries. 'Pearl growing' and citation-chasing methods (see: Cooper et al., 2018) were subsequently applied to derive a robust corpus of literature. The authors supplemented this corpus with additional opportunistically chosen academic and grey literature relevant to the topic of PPPs, as well as extremism and gaming. Where appropriate, best practices, lessons learnt, and key challenges were identified to inform recommendations on PPP models that may be adapted to gaming contexts.

The second step encompassed the key informant interviews (KIIs). 20 interviewees were identified using purposive and opportunistic sampling based on the authors' networks and additional snowball sampling recommendations during the interviewing process (Palinkas et al., 2015). The pool of participants was divided into two segments: one operating in the public sector, working as policymakers and practitioners in law enforcement, intelligence, regulatory, or national CT and P/CVE fusion centres. The other comprised participants working in the private sector in social media, gaming-adjacent, or game development companies.

Ultimately, the authors conducted 17 interviews over the course of just over two months – from July to early September 2023. Each interview – conducted virtually over Zoom – lasted from around 45 minutes to over an hour. The authors prepared semi-structured questions in advance through the drafting of an interview tool. Questions were shared in advance when requested by participants. The authors spoke with seven female and 10 male respondents from 11 countries. Of these, nine were from the private sector, two from non-profits or research entities, and six from governmental or public entities.

To ensure that participants felt comfortable sharing information and to maximise candour during interviews, the analysis of the interviews has been made wholly anonymised and non-identifiable. The authors occasionally identify the sector of an interviewee where relevant and where doing so does not risk identification. Interviews were transcribed using an automated transcription service, then saved on encrypted services accessible only by the research team. The authors then analysed the transcripts using grounded theory approaches, developing a priori codes from initial interview topics, then expanding the coding to pull relevant sections from the transcripts and group common topics across the complete interview corpus. This report draws on analytical assistance from generative AI tools, but the analysis and all responsibilities outlined herein are the authors' own.

CHAPTER 1: ACKNOWLEDGING THE THREAT

Gaming platforms, given alternative social and interactive functionalities, pose a different set of challenges regarding online harms compared to those found on other social media or web applications. Accordingly, the authors think a comprehensive understanding of the sector should be a pre-requisite when designing Public-Private Partnerships to counter extremist risks across the gaming space. While a comprehensive review of video games – and the ecosystem surrounding them – falls outside the remit of this paper, the authors believe a short overview may be helpful in framing the coming chapters. Previous RAN publications, in addition to recent publications, such as those from the Extremism and Gaming Research Network (EGRN), provide broader information on extremist trends in the sector for those interested in deeper reading (see, for example, Lamphere-Englund and White, 2023; Lakhani, 2022; Schlegel, 2021b and 2021c).

Digital gaming has emerged as the leading form of entertainment across the EU and the globe. Over three billion individuals are engaged in video gaming, a statistic that surged amidst the global lockdowns resulting from the COVID-19 crisis (Newzoo, 2022). By 2025, it is projected that half a billion more will join this community (Newzoo, 2022b). The projected revenue for 2022 was a staggering $184.4 billion, surpassing the combined revenues from movies, TV, and music (Newzoo, 2022b). The growth trajectory for this sector seems upward, both in terms of financial gains, user count, and pervasiveness across life stages. Today's gaming communities are more varied, including a balanced gender distribution. Women constitute nearly half (48 %) of the gaming community in the U.S. – the EU holds a similar share (ESA, 2022; Laaninen and Wessman, 2023). The typical gamer is around 33 years old. The passion for gaming often begins in youth and persists through adulthood. In the U.S., 71 % of children and 65 % of adults enjoy video games (ESA, 2022). The age breakdown reveals that 24 % of gamers are below 18 years of age, 36 % fall between the ages of 18 and 34, and the remaining 40 % are aged 35 and above (ESA, 2022). Although precise global statistics vary, most other regions in the world show a similar age distribution among gamers compared to the U.S. Given their youth, gamers are often cynically viewed as potential recruits to groom by militant and extremist groups globally.

Additionally, contemporary games are far more than immersive experiences: they are social spaces where players build very real identities. The hyper-interactive nature of multiplayer games, which often fuse layers of audio and text chat on top of gameplay, mix with gaming spaces such as livestreaming (where popular gamers often film themselves playing their favourite titles and interact with fans), chat forums centred around games, and other services, such as item marketplaces. Simultaneously, games are pivoting towards live services and are more encompassing in terms of their engagement than non-gaming social media platforms. Unlike social media posts, which can be reviewed after they have been posted, the dynamic nature of multiplayer games makes immediate content moderation a challenge. Additionally, many multiplayer games feature voice chat or integrate with additional apps to offer voice capacities (such as the voice channels in Discord, which can be harder to monitor and moderate than text). Automated tools for identifying problematic content in audio and video content are less developed than those for text, making those potential avenues for extremist narratives.

Not only do games incorporate the same features that more traditionally established social media platforms have (profile pages, private chat functions, and so on), but you can also purchase in-game and out-game items (skins, avatars, custom weapons) that potentially impact a player's status within the community or reflect their commitment to a particular game or platform. Potential financial flows and money laundering through these platforms to finance terrorism or extremist causes are a serious concern yet remain under-assessed (Davis, 2022; Kelly, 2021).

Despite its demographic sway and financial prominence, regulators and policymakers have often overlooked the sector. In interviews, respondents suggested that various causes may be at play. These ranged from outdated perceptions of games as solitary single-player endeavours – lacking social aspects – to a fixation on outdated stereotypes such as a belief that violence in video games causes offline violence. It bears reminding that there is no causal evidence of this (APA, 2015). Most people who play games have overwhelming positive, pro-social experiences that can be profoundly beneficial (Kaye et al., 2017; Kowert et al., 2014). That being said, extremist exploitation of gaming platforms is a clear risk (Lamphere-Englund and White, 2023) and one which is furthered by a few specific attributes:

- The live, interactive, and social nature of multiplayer gaming spaces facilitates socialisation and identity formation alongside risks of grooming and radicalisation.
- Cross-platform engagement and out-linking[2] are used by malicious actors to bypass platform-specific rules or engage in harmful activities via strategic movements across platforms.
- In-game imagery, motifs, and symbols can be manipulated for propaganda purposes. Meanwhile, bespoke games[3], mods, and symbols made by players can be used to communicate extremist narratives. These can be difficult to monitor and may require a nuanced understanding of the symbols being used and their potential meaning.
- A normalisation of extremist rhetoric: games pose unique challenges in that aspects of gamer culture(s) have normalised toxic rhetoric, be that white supremacist, racist, far right, or otherwise (ADL, 2022; Kowert et al., 2022). While such rhetoric is found elsewhere in digital spaces, the authors acknowledge that much of gaming culture is steeped with problematic speech – racist, homophobic, and misogynistic – that would be socially stigmatised nearly anywhere offline.

At the same time, increasing numbers of players are harassed, exposed to extremist content, and bullied to the point of leaving gameplay entirely (Kowert & Kilmer, 2023; ADL, 2022 & 2021). Fully six in 10 gamers polled in North America in 2023 had quit a game because of the abuse they experienced, while some 77 % of gamers in the U.S. were exposed to white supremacist content in 2022 (Kowert & Kilmer, 2023; ADL, 2022). Forthcoming polling that the authors are working on in several EU countries indicates similarly concerning trends. Parts of this may be explained by what has been referred to in the research literature as identity fusion, which posits that the conflation of individual identity with group identity ('being a gamer') can be a predictor of antisocial and potentially extremist endorsing behaviours (Gomez, et al., 2011; Kowert et al., 2022).

ADMITTING THE PROBLEM IS THE FIRST STEP

However, as the interviewees pointed out, most gaming platforms often do not want to admit they have a problem with extremist content. Many platforms – including their management teams – are afraid of giving any credibility to a new 'moral panic' around gaming, given previous and erroneous conceptualisations of gaming and offline violence (RAN Interview #1, 2023). Interviewees also commented on the industry's occasional naivety about its audience: contrary to the belief that children are exempt from radicalisation, extremist elements are

[2] Cross-platform engagement refers to the interaction and participation of users across multiple platforms or devices, specifically pertaining to operating systems, software applications, social media networks, and hardware devices. Out-linking refers Out-linking, also known as outbound linking, refers to the practice of including hyperlinks in a web page or content that direct users to external websites or resources.

[3] Bespoke games are custom-made or tailored video games. Bespoke games are different from mods in that they are typically entire games, rather than pieces of content from an existing game that have been modified to suit user needs.

targeting younger and younger audiences who are finding their social milieus through gaming (Koehler et al., 2022). Simultaneously, games also serve a far wider audience than children alone (RAN Interviews, 2023).

Ultimately, effectively addressing issues related to extremism and terrorism on online platforms requires an acknowledgement from all stakeholders. This hesitance from industry poses significant obstacles to ensuring safer online gaming environments. In general, many of the gaming industry's conversations around extremism mirror those of social media platforms during their nascent stages around 10 years ago (around 2010-2013). As Facebook grew, and newer platforms like Instagram emerged, they faced challenges in recognising and combatting extremist content. Public and regulatory pressure, in addition to internal motivations related to reputation and marketability , then moved these social media companies to progress the issue by developing improved trust and safety[4] responses, such as increased content moderation. The authors see a pattern of hesitancy across the digital space, particularly strong among emerging platforms, to address and admit to problems linked to extremism, terrorism, or related concerns. Several factors contribute to this hesitance in the gaming sphere:

- **Historical Trauma:** Past associations of video games with offline violence in the 1990s and 2000s, largely since disproven (APA, 2015), have left the industry wary of revisiting topics that could potentially stir another moral panic. The interviewees noted that many in the industry often believe they need to distance themselves from any negative connotations of gaming, denying any problems in public.

- **Knowledge Gaps:** Internal understanding of extremist content and radicalisation risks on gaming platforms is also lacking. Expertise in extremism and terrorism is often found in government or academia, yet knowledge-sharing with gaming trust and safety teams is minimal between most companies. Other game studios are simply too small to reckon with potential risks, yet their user bases can increase extremely quickly if a game becomes popular, creating issues of scale without internal guardrails in place.

- **Fear of Regulation:** The gaming industry is, as would be the case with any industry, generally cautious about regulation and fears a recurrence of previous moral panics from the '90s and early 2000s around video games and violence. Regulatory awareness is especially present in the EU context as regulations such as the Digital Services Act (DSA) come into full implementation.

- **(Lack of) Industry Engagement:** Keeping regulators at arm's length might protect companies in the short run, but it leaves them without a say in decision-making processes affecting the gaming ecosystem in the long run. This is especially true for smaller game companies, which may not have the ability or bandwidth to consider engagement with policymakers.

At the same time, politicians, policymakers, and law enforcement agents often do not know much about gaming communities or the companies that make games and adjacent platforms. The interviewees from regulatory bodies and law enforcement highlighted significant internal knowledge gaps, with one regulator noting that "the degree of misunderstanding could partly be due to how [we] as an organisation recruit and the people that we have" (RAN Interview #2, 2023). It is crucial for both gaming platforms *and* governments to not only be aware of potential extremism risks but also to have robust policies to manage and counteract these challenges.

[4] "Trust and safety" is a broad concept that encompasses measures and practices implemented to create a secure and reliable online environments.

COMMON NEEDS OF PUBLIC AND PRIVATE ACTORS

As with other low-prevalence and high-risk behaviours, a lack of data makes it difficult to ascertain exactly what this online threat landscape looks like. However, open dialogue, recognising challenges, and active collaboration can help to ensure that the gaming industry and government can effectively address concerns related to extremism while providing safe spaces for citizens and users.

As interviews with government and industry representatives alike attested, there is a clear overlap between private companies and public sector interests. Both sectors share an interest in preventing the spread of extremist narratives for societal and security reasons. Towards that end, Public-Private Partnerships in relation to national security may be attractive models of collaboration for national authorities as they provide them with the ability to involve industry in national security strategies with said help from industry. For the private sector, PPPs may be attractive as industries can experience certain problems – e.g., extremism in video games – the impact and possible solutions of which are beyond the expertise and means of addressing by industry alone., Meanwhile, recent regulations including the EU 2021 Terrorist Content Online (TCO) regulation and aspects of the 2022 Digital Services Act (DSA) illustrate the rapidly evolving statutory landscape that companies will be required to navigate. PPPs can help ensure both compliance and successful, mutually beneficial aspects such as voluntary referral mechanisms.

At the same time, many industry partners and policymakers alike lack awareness of extremist content on gaming platforms. According to interviewees, better internal and cross-sectoral knowledge and education, or capacity-building, are key to resolving this. Information gathering and data sharing mechanisms may need to be created as platforms often do not have specialised knowledge in-house on violent extremist issues. That means that signals of radicalisation, especially less overt ones, are often missed. More robust monitoring and data-sharing systems already in place for other types of illicit content such as CSAM (Childhood Sexual Abuse Material) provide a template for what could be done to thwart extremist or terrorist exploits.

Both sectors could benefit from shared knowledge, tools, and practices to identify and mitigate these narratives, including collaboration with academia and civil society organisations. Equally, those interviewed pointed out that governmental agencies and gaming platforms can possess similarly complex bureaucratic structures with various divisions and branches, each holding competing interests. This complexity can hinder the straightforward establishment of best practices in the realms of trust and safety, making public disclosure and internal buy-in difficult tasks. Platform representatives the authors spoke with said they face a hard balancing act when it comes to publicly acknowledging their efforts in trust and safety. On the one hand, such acknowledgement can demonstrate proactive steps towards addressing problems. On the other, it necessitates an admission that challenges exist. This situation calls for policy adjustments in both internal and external communications. Such adjustments should enable public and private actors to openly discuss their initiatives without instigating panic or admitting to any inadequacy on their part.

In short, to start the conversation with industry and to move forward effectively with creating models of Public-Private Partnerships to address extremism on gaming platforms, there are several foundational principles that must be borne in mind:

1. **Acknowledge common ground and common aims.** By understanding the common risks and challenges both societies and platforms face, partnerships can be launched from the standpoint of collaborative problem-solving instead of blaming private sector actors.
2. **Understand that building safe online spaces is good for public interest and for business.** Recent polling shows that more people are leaving games because they are often full of hate and harassment (Kilmer

and Kowert, 2023). There is a clear monetary argument for building inclusive, strong communities in games, just as there is for building such communities in democratic societies.

3. **Frame the conversation for partnerships in a shared language.** Industry insiders suggested avoiding discussing only extremism and terrorism and instead framing the discussion around broader topics such as hate and harassment, online child safety, or grooming: "I find that they've been more open to talking about hate and harassment more generally, which is kind of a way in the door and finding ways to target hate and harassment generally, which in a way will trickle down and also impact extremism" (RAN Interview #4, 2023).

These approaches are corroborated by existing literature on Public-Private Partnerships – elaborated on below – which indicates that the use of clear, coherent, and concise definitions, frameworks, and vocabulary is a necessary condition for effective collaboration. A shared understanding of the threat landscape – whether pertaining to child sexual exploitation or other trust and safety concerns like extremism – helps outline goals and objectives, communicate expectations, and provide measurable success metrics (Beutel & Weinberger, 2016).

CHAPTER 2: PUBLIC SECTOR GUIDANCE

Public sector regulators play an essential role in shaping the online safety landscape. According to industry insiders interviewed, successful regulatory regimes need to be nuanced and help to shift the perception of trust & safety efforts from a cost center that can be cut to an operational necessity. Meanwhile, beyond improved regulatory regimes and in keeping with the theme of this paper, this research identified several innovative PPP opportunities for public sector investment and engagement. Accordingly, the authors have split the guidance in this chapter into five groupings below:

- **Models**, which examines promising PPP formulations related to Counter-Terrorism (CT) and Preventing and Countering Violent Extremism (P/CVE), as well as other sectors.
- **Understand**, which provides suggestions for PPPs to better evaluate the risks and challenges of extremism on gaming platforms.
- **Dialogue**, or guidance for how public and private actors can better come together and jointly decide how to address these challenges.
- **Regulate**, with specific recommendations for regulating extremist exploitation of gaming platforms.
- **Prevent**, with ideas and practical suggestions on P/CVE programming that can be deployed to tackle exploitation taking place in gaming.

In the third chapter, which follows this one, this report also provides tailored guidance to the private sector on potential avenues for successful collaboration with governments.

MODELS: EXISTING FRAMEWORKS FOR PPP

PPPs can take multiple shapes. Based on a case study analysis, the European Union Agency for Cybersecurity (ENISA) broadly defines three categories of Public-Private Partnerships: (i) response-focused PPPs; (ii) prevention-focused PPPs; and (iii) umbrella PPPs. Response-focused PPPs are, as the name suggests, typically organised to respond to a particular event. In security terms, this can be in response to a natural disaster or, more pertinent to the topic at hand, a terrorist attack. Response-focused PPPs have a clear technical and operational focus (ENISA, 2011). An example of such a PPP is the Dutch Counter-Terrorism Alert System (*Alerteringssysteem Terrorismebestrijding*) which is designed to mobilise public and private sectors in case of an imminent terrorist attack. It covers a wide range of actions, from detection (e.g., a chemical company flags a suspicious purchase) to coordination and information-sharing, and is designed to create resilient information and decision-making infrastructure that explicitly involves private industry (Nationaal Coördinator Terrorismebestrijding en Veiligheid, 2018).

Prevention-focused PPPs, on the other hand, have a long-term strategic focus aimed at preventing a particular security threat. Such partnerships leverage awareness-raising, knowledge-generation, knowledge-sharing, capacity-building, and readiness exercises to anticipate and preventively address security threats before (and immediately after) they manifest – in addition to offering responses to threats after they have transpired. An example of this is the Slovenian initiative VARNI NA INTERNETU, which raises awareness on digital hygiene and safety through the publication of reports on sextortion, scams, and disinformation

(https://www.varninainternetu.si/). The initiative also publishes handbooks for law enforcement practitioners and members of the armed forces. It offers best practices for private companies who have been the subject of defrauding or scamming attempts.

Another example of a prevention-focused PPP is, of course, the EU Commission-led Radicalisation Awareness Network (RAN). Both RAN Practitioners and RAN Policy Support engage in awareness-raising, knowledge-sharing, and capacity-building, with an explicit focus on P/CVE. Both strands also involve the private sector (either as practitioners or official partners) and public sector (practitioners and policymakers) to produce cutting-edge research, share best practices and lessons learnt, and support the formulation of evidence-based policy to prevent and counter violent extremism.

Umbrella PPPs are those that "have the ability to deliver capability across the full security life cycle," meaning that they are comprehensive PPPs that are active in both respond and prevent stages (ENISA, 2011). They may even be key implementers of national security strategies, thus operating at strategic and tactical levels. Of course, umbrella PPPs may be enacted in response to a particular threat, but their remit encompasses collaboration that extends across public sectors and industries. For example, the COVID-19 PPP Rapid Response Umbrella Program is a partnership between the Public-Private Infrastructure Advisory Facility and the World Bank that manages global PPPs to assess responses to the disruptions caused by the pandemic, advise governments, and assist sectors most affected by COVID-19.

A study by the Verwey-Jonker Institute (Van Steden & Meijer, 2016) identified additional PPPs focused on the aftermath or recovery phase of a security threat. The report, however, acknowledges that the actions of these particular PPPs and those of the response PPPs described above are blurred. One salient distinction listed in the report pertinent to developing PPPs in the online space, however, relates to 'market-PPPs' versus 'authoritative-PPPs.' The former is a type of government-led, transactional partnership whereby national authorities may tender services to private entities. As Newman (2017) notes, however, such arrangements should not be called PPPs unless some risks, responsibilities, or rewards are shared between the private contractor and the public authorities. Government simply hiring a private company to deliver a service such as catering should not be construed as a PPP, because the engagement is transactional and limited in scope, focusing on a specific service (catering) without involving the caterer in broader project responsibilities. Conversely, constructing and maintaining a toll road could be crudely construed as a PPP if both contractor and government share risks (e.g., no one drives on road so no toll is collected), responsibilities (maintenance for the road, even after the contractor relinquishes it back to government), and rewards (toll revenue that the contractor collects and the government can tax).

'Authoritative PPPs' constitute a national authorities compel private entities to undertake certain actions by, for example, imposing industry standards while reserving the right to sanction in case of non-compliance. A topical example of this is the EU Terrorist Content Online (TCO) regulation. Though not strictly a PPP in the sense that it is a legal instrument imposing clear obligations on Member States and Hosting Service Providers, it provides the obligation for private companies to remove terrorist content within the hour upon receipt of a removal order from Member States' competent authorities.

Type	Description	Scope
Response-focused PPP	A PPP established in respond to an imminent security threat.	Tactical and operational
Prevent-focused PPP	A PPP established to provide long-term strategic capacities to prevent a security threat.	Strategic

Umbrella-focused PPP	An overarching PPP established to achieve tactical, operational, and strategic objectives to respond to and prevent a security threat, or to implement a national security strategy.	Tactical, operational, and strategic
Market PPP	A PPP based on material or service agreements between public and private entities, with at least some shared risks, responsibilities, and rewards between public and private actors.	Transactional or contractual
Authoritative PPP	A top-down PPP imposed by national authorities in the form of regulation and establishment of industry standards.	Regulatory

Table 1. *A summary overview of Public-Private Partnerships (PPP) adapted from ENISA (2011), Van Steden & Meijer (2016) and Newman (2017).*

Challenges and lessons learnt - PPPs in P/CVE and adjacent fields

Public-Private Partnership configurations come with different challenges, risks, and rewards. In a general sense, PPPs focused on P/CVE share similar issues as those faced by inter-/multiagency partnerships between public institutions (e.g., Danish Info-houses or the Dutch *Veiligheidshuizen*), for example, related to information-sharing, interagency trust, different organisational cultures, and the use of different terminology and language. Examples of this can be found in collaboration models between law enforcement and mental health practitioners, where patient confidentiality and distrust towards security agencies can sometimes stifle the sharing of timely and pertinent information. Similarly, the involvement of a wide array of actors within a partnership may obfuscate roles and responsibilities, and communications may consequently be hindered. Partnerships with a focus on countering violent extremism should clearly outline roles and responsibilities and adopt a uniform approach to information-sharing.

A report from the University of Maryland's National Consortium for the Study of Terrorism and Responses to Terrorism, or START, (Beutel & Weinberger, 2016) outlines a number of field principles that may guide the successful facilitation and implementation of PPPs focused on P/CVE. In brief, these are:

Principles for Government Actors to Facilitate PPPs

1. Identify the division of labour.

   This principle advocates that partners determine what tasks are "inherently governmental" – i.e., can only be executed by governmental agencies – and what is "inherently non-governmental" and should be executed by industry and civil society partners.

2. Reduce barriers to entry for non-governmental partners.

   This principle prescribes that the barriers for participation of industry and civil society partners should be as few as possible. Such barriers include funding opportunities (including but not limited to financial incentives), informational and relational barriers (understanding the scope of violent extremism in communities and the resources that exist to combat it), and political and policy-oriented barriers (creating trustworthy relationships between governmental and non-governmental actors).

3. Foster an organisational culture that makes partnering a top priority.

This principle denotes a working culture that seeks to maximise the partnership's effectiveness by fostering a collaborative mindset, often through individuals who champion this principle.

4. Act as an 'innovation catalyst.'

   This principle prescribes that governments create environments for industry and civil society partners in which innovation of CVE-related measures and technologies can be rapidly developed and deployed, for example by drawing inspiring practices from adjacent government fields.

Principles for Developing, Implementing, and Sustaining Successful PPPs

1. Have clear goals, focus on results, and measure progress.

   Denotes having shared and measurable aims and objectives with appropriate success metrics.

2. Involve consumers in developing programmes.

   Successful interventions know and understand their audiences. Involving consumers from the target audience in partnerships helps shape successful interventions.

3. Involve diverse stakeholders from the start.

   Facilitate cross-sectoral collaboration and multiagency working (MAW) by involving diverse stakeholders with a range of applicable skills and competencies.

4. Identify and utilise champions for support.

   Champions are individuals who act as change agents and can foster partnerships, promote working culture shifts, and drive innovation.

5. Establish clear governance structures.

   Governance structures help define roles and responsibilities, coordinate stakeholders, and manage diverse and competing interests within a PPP.

6. Adapt to changing conditions.

   These conditions can include a change in governance and funding, and shifting threat landscapes.

7. Enable all partners to benefit.

   Enabling partners to benefit creates incentives for them to engage in partnership.

8. Work to maintain momentum and sustain efforts.

   This may include planning for financial support and build up resources to ensure longevity of the PPP.

9. Balance transparency and confidentiality.

   This can be achieved by considering legal obligations to disclose certain information (e.g., funding sources); personal safety of partners (especially when considering threats made by terrorist and violent extremist offenders (TVEOs) against CVE-specific PPP partners), efficacy of interventions (sometimes, it might be desirable that governments do not disclose their involvement in PPPs to ensure other industry or civil society partners remain credible); programme evaluation (data needs to be sufficiently transparent between partners to evaluate the partnerships).

Table 2. Principles for CVE-focused Public Private Partnerships. Adapted from Beutel & Weinberger (2016).

UNDERSTAND

One issue raised by nearly all of the interviewees, as also demonstrated in the existing literature (Lamphere-Englund and Bunmathong, 2021; Lakhani et al., 2022), is that of an evidentiary gap. While a substantial and increasing body of evidence points to the real harms and risks of radicalisation and extremism present in the gaming ecosystem, nuanced research needed to inform policy and regulatory development is still insufficient. The authors see reasons for this, and most notably the lack of accessibility of internal data on gaming platforms, which makes understanding the social workings inside games extremely difficult. Imagine, for instance, an anthropologist who cannot observe one group for more than a few minutes of a game, an urban planner who cannot access road maps of the city, or a sociologist who cannot survey their communities of interest with any precision. As digital spaces evolve into complex ecosystems with various platforms and communities, so does the challenge of effectively countering extremist behaviour. There is a pressing need for greater data transparency and information-sharing among stakeholders, including platforms, researchers, and law enforcement agencies.

"I […] would say one is that game platforms, unlike social media or forums or any other kind of online platform, are significantly less searchable and less visible to researchers and journalists than any other kind of platform that I'm aware of." (RAN Interview #7, 2023).

The authors and other researchers frequently advocate for access to Application Programming Interfaces (APIs), which can provide valuable insight into extremist content dissemination and social interactions on platforms: Meta, TikTok, and other social media platforms have begun to offer such access. The DSA also has provisions to require data access and transparency, yet as of writing no gaming platforms and services have been designated as Very Large Online Platforms (VLOPs) or Search Engines (VLOSEs) to which the full onus of obligations placed from the DSA would be applied based on their role, size, and impact online. Additionally, the interviewees raised the issue of the compartmentalisation of conversations and bad actors' behaviours across multiple platforms. As each platform only has a view of its users' activity, yet online users bridge across many sites and platforms, a lack of information sharing prohibits any effective intervention. To truly grasp the scale and nuances of extremist narratives on gaming platforms, the barrier of visibility must be addressed.

As gaming platforms are significantly less accessible for research and data extraction, this report sees several issues that can be addressed through PPPs and public sector action. Drawing from the existing literature, it is evident that governments can act as innovation catalysts. Governments should foster environments conducive for industry and civil society to innovate, specifically regarding measures to counter violent extremism. For example, initiatives like Tech Against Terrorism Europe (TATE)[5] are prime examples of how industry, backed by civil society and financially supported by the European Commission through the Internal Security Fund (ISF), can devise flexible solutions to address regulatory challenges. Through innovative PPPs and dialogue efforts, which are discussed below, it is critical to improve the following:

- **Data Access:** For effective research, there is a need for data access. Gaming and adjacent platforms need to provide structured access, possibly via APIs, to their data repositories. This will allow vetted – non-profit motivated – researchers to understand the trends and nature of extremist content dissemination.

- **Cooperative Frameworks:** Identifying threats requires a timely exchange of information. There is a potential to develop a clearing house for data, standardising the information gathering and sharing process. This not only ensures consistency but can also help in maintaining privacy by allowing for

[5] https://tate.techagainstterrorism.org/

anonymised information sharing. Precedents in sharing agreements exist for preventing Childhood Sexual Exploitation and Abuse (CSEA) content where, for example, behavioural patterns and flags for offending material can be shared between tech platforms to quickly remove content and offending users who might otherwise take advantage of siloed information between platforms.

- **Shared Datasets:** Extremist groups often employ covert symbols and terminologies, making their detection challenging. To address this, shared datasets that catalogue extremist symbols and content, and which are frequently updated can be invaluable. These can help platforms, especially smaller ones without in-house extremism expertise, to identify and address extremist content. With ideologies blurring lines and activism taking various forms, recognising these symbols can offer insight into potential threats. Precedents for this exist in the GIFCT Hash Sharing database (GIFCT, 2021), which allows tech platform members to share terrorist-related content flagged on one platform with another to enable rapid takedowns of material like the Christchurch perpetrator's manifesto. Similarly, the EU Internet Forum (EUIF) has released a manual on far-right symbols that has, according to the interviews, been well received by the tech sector but which respondents would like to see updated frequently. Improving on such collaborative efforts can bridge awareness gaps and help platforms stay ahead of emerging threats.

- **Funding Cross-Sector Research:** There is a significant need to fund research into the spread and nature of extremist narratives on gaming platforms. Unfortunately, funding is often only dedicated to academic groups, which can be slow and compartmentalised in their findings. More innovative approaches to funding – borrowing from R&D and innovation approaches taken elsewhere by governments such as the Five Eyes R&D Network or the Defense Advanced Research Projects Agency (DARPA) – can be designed that pool research funding across civil society, the private sector, and academics alike for the public good.

Data is often closely guarded due to its nature as intellectual property, as well as due to privacy legislation such as the General Data Protection Regulation (GDPR), creating another hurdle for collaborative efforts. Balancing the need for data privacy with the urgency of identifying and mitigating extremist behaviour remains a critical issue. However, information sharing is not optional: it is a necessity in the fight against online extremism. From data accessibility and standardised reporting to inter-platform cooperation, multiple pathways exist to make digital spaces safer. Given the dynamic nature of online communities, these efforts must evolve continuously, encompassing the technological, academic, and human elements involved in combatting extremism.

DIALOGUE

Based on robust information, and even while research is ongoing, dialogue between public and private actors must continue to address extremism online. Such dialogue plays a crucial role in policy implementation, setting norms, and navigating the challenges posed by new technology. From an *Umbrella PPP,* which might be established to "achieve tactical, operational, and strategic objectives to respond to and prevent a security threat, or to implement a national security strategy," (Van Steden & Meijer, 2016; Newman, 2017) to a nimbler *Response-Based PPP,* established in response to an imminent security threat, the interviewees provided a range of tips to foster such dialogue.

Public-Private Dialogue General Tips

Firstly, partnerships aimed at P/CVE online need to **share clearly delineated aims and objectives** while protecting against undue levels of influence. As one regulator put it, *"*I think industries should be more involved in this process, not to the point where they can guide it, but where at the very least they can have some sort of input into it because feigning technical knowledge in an area is just going to end up making the [regulatory]

organisation look stupid further down the line" (RAN Interview #3, 2023). Preventing and countering violent extremism can be a fuzzy concept: actors seeking to establish a partnership should first establish "what is it that we are trying to prevent?" In the context of gaming and online spaces writ large, this question is notoriously difficult to answer.

Secondly, **establishing trust** is foundational. Both public and private entities must approach discussions with openness – and development of shared needs, as discussed in Chapter 1, is key to nurturing this trust. It is essential that both public and private actors commit to transparent communication, ensuring decisions are driven by shared goals rather than individual interests. A lack of blame, where possible, may be helpful here, though public sectors may also need to use a mildly coercive approach to clearly name the issue of extremism present on many gaming platforms. Additionally, the interviewees noted that it is essential to understand the intricacies of tech firms, as they often operate in low trust and bureaucratic settings internally with different incentives between teams. Inside, a trust and safety or community team may be focused on user safety, management may be focused on profit and shareholder engagement, and product teams may lack the engineering time to spend on non-game related developments.

EXAMPLE OF FACILITATING DIALOGUE

The Global Internet Forum to Counter Terrorism (GIFCT) serves as version of private-sector led collaboration with public entities. While private sector led, platforms like GIFCT have afforded space for tech giants and governments to discuss and develop pragmatic guidance such as frameworks of regulatory regimes, workshops, and so on. Such engagement requires both industry involvement and input from technical experts. It is not about firms leading the discussions, but about ensuring they provide critical insight, thus ensuring that solutions are both technologically feasible and effective.

Towards that end, some interviewees emphasised the **utility of more informal, or at least semi-confidential dialogue** to build trust, such as those modelled on Track Two Diplomacy, which are unofficial dialogues and initiatives conducted by non-state actors or individuals to resolve conflicts or build trust between actors outside of formal diplomatic channels. These are often used in challenging negotiations while using Chatham House rules for discussions. These approaches prioritise understanding over confrontation, promoting a more collaborative spirit. These diplomatic efforts, rooted in informal and confidential dialogue, can foster mutual understanding, particularly in areas as delicate as terrorism and public policy. By adhering to these diplomatic efforts, an atmosphere conducive to 'constructive creation' emerges, where stakeholders feel empowered to invest their resources and political capital (RAN Interview #17, 2023).

Additionally, the **inclusivity** of these efforts is pivotal: mediators and conveners should ensure that all necessary stakeholders are present. Not just high ranking leaders or policy makers, but also frontline workers including community managers and UX engineers from gaming companies, or intelligence analysts and cyber-crime enforcement agents from government.

A few other points on dialogue bear consideration. The interviewees also noted that a **collaborative tone** set by governments geared towards constructive collaboration is pivotal. If the engagement strategy is purely coercive and not collaborative, it can deter actors from investing in their efforts. Lastly, they identified concerns about balancing informal conversations with the need for public **accountability** and **transparency**. For example, Chatham House rules dialogues can be beneficial to have candid conversations between government and

industry, but can also raise valid concerns of regulatory capture and the type of transparency fostered by public record hearings or meetings.

In terms of pragmatic steps for engagement during dialogue efforts, informants held that a few specific efforts may be particularly fruitful:

- **Establish Communication Channels with Smaller Platforms:** Regular dialogue with platform developers and service providers is crucial, especially with smaller platforms and game studios that may not be on public sector actors' radar. Proactive, helpful engagement by government that seeks out smaller actors and helps to provide support and frameworks for online safety can be of use in fostering such engagement.

- **Share Best Practices**: Dialogue can focus on providing technological solutions, guidelines, and best practices for content and community moderation – these are elaborated on later in this paper.

- **Appreciate Confidentiality:** Public sector actors should understand that some platforms may wish to cooperate without public acknowledgement, given the political sensitivities surrounding specific topics. Others may wish to show off their good work – but that is not a given and should not be taken for granted.

Whole-of-Society Engagements

Beyond bi-directional engagements between public and private sector actors, addressing extremism necessitates a wider, whole-of-society approach. As such, multistakeholder forums can promote information sharing, research partnerships, and transparency. Taking a holistic angle on dialogue involves including representatives from diverse sectors such as the gaming industry, civil society, education, academia, and law enforcement. By organising roundtables or even establishing formal foundations, the depth and breadth of discussions can be enriched. With the increasing influence of global gaming platforms, such collaboration should really extend across borders – i.e., also beyond the EU – to address challenges associated with online extremism. For instance, some of the respondents suggested:

- Creating 'tripod' platforms between government, gaming companies, and civil society that could be jointly funded by the state and private sectors to minimise potential biases and provide funding to civil society with less of a perceived risk of capture by any one set of interests while mitigating the cost to taxpayers. The independent Extremism and Gaming Research Network (EGRN) of which, in full disclosure, the authors are a part is one such initiative that could be strengthened.

- Forming foundations or similar non-profit legal entities to work on extremism online, especially in gaming communities steered by civil society and representatives of gamers themselves, ensuring those affected by policies have a voice in their creation and implementation. These foundations would work on both research and prevention activities.

- Sharing local insight. Local prevention work, including online and through gaming platforms such as via digital street work efforts with social workers online, are worth bringing into wider dialogue to share. The authors heard mention of specific municipalities in Denmark and Norway in multiple interviews.

With the right approach and stakeholders being present in the room, such collaboration can result in innovative and effective solutions to combatting online hate on gaming platforms. In the sections ahead, the report elaborates on particular examples of insight that can be useful for regulators and prevention practitioners, as well as platforms.

REGULATE

PPPs for Improving Regulatory Frameworks

As illustrated by the manifest creation of new legislation across the EU and the globe, there is a pressing need for robust regulatory and policy frameworks to address online harms, including the sort posed by extremist manipulations of gaming platforms. Two which immediately came to mind for respondents are the **Online Harms Bill,** initially presented by the UK Government in 2019 and gradually making its way to implementation, and the **Digital Services Act (DSA)** which came into force on 16 November 2022. The **Terrorist Content Online** regulation **(TCO)** to address the dissemination of terrorist content online that came into effect in June 2022 was also mentioned positively by industry insiders and public policy informants alike as being well-shaped and particularly clear regarding definitions, roles, and responsibilities for platforms. The DSA and TCO provide measures against illegal content and terrorist content respectively, setting transparency requirements, and safeguarding users' rights online.

FROM SPORTS ARENAS TO DIGITAL PLATFORMS: GOVERNING ONLINE PUBLIC SPACES

Digital gaming platforms, akin to public community areas, play a crucial role in shaping social interactions and fostering community connections. Drawing parallels between these platforms and the modus operandi of sports teams offers an insightful perspective on their governance and responsibility.

Sports arenas, such as football stadiums, often attract a concentration of (mainly) young men, making them potential grounds for extremist recruitment, especially for extremist ideologies such as far-right movements. Recognising this, top-tier sports clubs in many countries, especially in the UK and Germany, have proactively initiated community programmes linking “fan workers,” supporters and police authorities to educate and inform their fans about radicalisation and violence risks (Marsh et al., 1996). While the sports clubs do not necessarily police these spaces themselves, their close collaboration with law enforcement ensures a safe environment for all attendees.

Much like the community initiatives of these sports teams, digital platforms can also deploy an ‘ethical net’ — a system that not only engages with its user base to increase awareness about potential threats but also partners with law enforcement to understand and adopt effective strategies for combatting these threats.

The willingness to openly address and discuss these challenges is vital. The dynamics between larger publishers and platforms can be complex, especially considering the global nature of many digital businesses. For instance, a significant portion of gaming businesses are owned by international entities. Thus, while the business may operate in one country, the decisions may be influenced by stakeholders from another. Despite these intricacies, the overarching objective remains the same: to create a safe, inclusive digital space that is cognisant of potential threats and actively works towards mitigating them.

A comprehensive review of online harms legislation is beyond the purview of this paper, but the role of Public-Private Partnerships in shaping well-informed legislation came up regularly during conversations for this paper. As such, the authors have developed a brief set of principles and tactics from our research that may help to guide future efforts at regulating online gaming spaces.

While regulation can be an effective driver of change, these frameworks must take into account the nuances of gaming ecosystems. A broad-strokes approach that fits commerce platforms and social media might overlook essential elements like in-game chat functionalities, voice chat dynamics, and cross-platform engagements that gamers often take. On that point, a pertinent concern raised by industry experts is the inherent challenges posed by the vast volume of content generated in real-time interactions. Smaller platforms and gaming studios, in particular, need to have clear guidance that also assists them in preventing extremist and terrorist content from appearing on their platforms.

Building on those critiques and on best practices from the regulators and policymakers with whom the authors spoke, a clear set of **Principles for Effective Regulation** of online (gaming) platforms emerges:

- **Nuance and Flexibility:** Regulation should be crafted with keen attention given to the diverse designs of platforms, both in terms of their operational scale and the nature of their services. This flexibility should also accommodate the swift evolution of online threats from extremists and other malicious actors.

- **Holistic Approach:** Addressing extremist content and narratives on gaming platforms should not be limited to content moderation. Underlying societal issues that manifest online should also be tackled. Merely focusing on the removal of harmful content will not resolve the root causes – regulation should recognise this and also attempt to incentivise positive community practices beyond just takedowns.

- **Proportionality and Transparency:** As underscored by the experts spoken to, the development of regulation should be grounded in thorough research and be informed by pragmatic consultation with technical experts or platforms, ensuring proportionality of response to the threats evidenced in research. Additionally, any legislative process should be transparent, ensuring inclusivity in consultations and decision-making, with both industry and civil society alike.

- **Human Rights and Due Process:** Regulatory frameworks need to ensure compliance with fundamental human rights and established legal processes.

- **Fairness in Implementation:** Regulatory principles should provide guidance for how fairly apply regulation to all users when transposing into platform policies, guidelines, content review processes (both manual and automated), and content appeal procedures. It is crucial to ensure that platform implementation does not target specific communities based on protected characteristics, and that there is sufficient recourse in event of challenges to content flags, bans, shadow-bans, or other enforcement actions. Both automated and human content review systems have the potential to perpetuate racial and other biases (see, for example: Chen, 2023; Zack et al., 2024) – regulatory guidance should be aware of this concern and provide principles to mitigate against it.

- **Broadened Perspectives:** Drawing inspiration from the past can offer valuable insight. Historical regulatory approaches, such as those applied to industries like steel and railroads, or telecom networks can provide valuable lessons for today's digital landscape.

- **Transparent Reporting:** Platforms should adopt reporting mechanisms that serve multiple stakeholders. This transparency is not just essential for governmental oversight but also benefits the general public, researchers, and civil society. Such a framework should enable these groups to analyse the provided data, derive insight, and advocate for necessary modifications. Regulation should be bolstered by provisions for data audits that have clear reporting metrics.

Collaborating between platforms and governments is crucial to ensure a safer gaming environment that respects users' rights and addresses the challenge of extremist manipulation. Public policymakers need to understand the technical limitations of platforms, while platforms need to have clear requirements given to them. One comment consistent in interviews was how software engineers should be brought in to provide expert opinions on the technical underpinnings of gaming platforms and how these can be adapted to meet emerging regulations. Challenges concerning content moderation and flagging are just as much *policy problems* as they are *engineering problems*: "In any tech company, engineers are gold dust. You can feel like the lawyers and public policy people have a role to play. But let's be honest about where we sit…even about some of the things you do not necessarily think about as [being] in engineering spaces. It is not just about detection technology: this is about building all of the architecture that we need to do these things." (RAN Interview #8, 2023).

Respondents suggested three areas of focus where regulators and public sector actors can provide more specific guidance on terrorist and extremist content applicable to gaming platforms and companies: **content guidelines, including those addressing borderline content, algorithmic amplification limitations, and transparency reporting requirements.**

Borderline Content Guidelines

As this report has shown, while online gaming platforms continue to evolve and attract diverse user bases, governments and platforms must collaboratively navigate the nuanced challenge of regulating extremist narratives. The TCO stands as a noteworthy attempt, receiving general endorsement from industry and public sector representatives with whom the authors spoke. Yet, a key challenge identified across both sectors concerns the regulation of borderline content: narratives that are legally permissible but morally questionable. The 'lawful but awful' type of content. While governments should be the ultimate arbiter to delineate legal and illegal content, the growing reliance on private companies to regulate online content complicates the matter and has resulted in many removing borderline content from their platforms. There is a pressing need to clearly define who bears the responsibility for determining the permissibility of content and behaviours on gaming and other platforms online. While governments define what is legal and what is illegal, private companies clearly play a role in making content removal decisions in the absence of clear guidance on borderline content. For example, multiple social media platforms, as noted earlier, acted to remove content and adopt internal policies banning terrorist and extremist content following massive public backlash against ISIS-related content on their services during the period from 2012-2015. A similar wave of removals of far-right related content followed the January 6th riots in the US in 2021. However, compared with social media companies, fewer gaming and gaming-adjacent platforms have community guidelines or Terms of Service (TOS) which prohibit terrorist or violent extremist activities and content.

Defining Borderline Content

Defining terrorism and violent extremism on a regional or global scale poses a significant challenge. Due to diverse international perspectives, establishing a universally accepted definition remains elusive. The scope of the TCO is limited to terrorist content– and does not cover borderline extremist content. Distinguishing between violent extremism and terrorism, meanwhile, adds another layer of complexity. While the TCO covers the domain of terrorism, it will not cover content that leans more towards violent extremism if it does not fall under the definition of terrorist content in the TCO.. Such distinctions become especially pertinent when considering the liminal content spread on video gaming platforms or private servers on platforms like Discord.

Another significant hurdle in defining extremist content – especially in gaming settings – is the role of context. The same piece of information or imagery could be interpreted differently based on contextual factors. For

instance, a historical image might be educational or game-related in one context but could be used to propagate extremist ideologies in another. Add to this the pervasive toxicity of many gaming environments, and the automated monitoring and regulation becomes incredibly challenging. As one of the interviewees said:

"Borderline content…is not really covered by [the TCO], and if we go into especially the right-wing or the violent right-wing sphere, we do not deal a lot with clearly labelled content, which is more often violent and extremist than terrorist. This is a point where the TCO regulation, for example, is not very strong." (RAN Interview #9, 2023)

Borderline content occupies a grey area in content moderation. It frequently does not violate terms of service or legal guidelines outright but hovers at their edges. This might include subtly worded hate speech, dog whistles, coded language, or content that has plausible deniability. Meme-based, or memetic, propaganda from extremist organizations often also falls into this category by employing humour to lower inhibitions to radicalising content. The ambiguity surrounding borderline content makes it particularly challenging to regulate, especially when considering free speech protections. Platforms often find it tough to justify action against such content without clear violations and might be reluctant to take action against borderline content due to freedom of speech concerns, fears of backlash, accusations of bias, or potential legal challenges.

Extremist views are increasingly finding their way into mainstream conversations online, especially in gaming social spaces rife with toxic, racist, misogynistic, and homophobic content. Efforts such as the recent EUIF handbook on borderline content were viewed in positive terms by interviewees as they help platforms identify extremist content. However, the static nature of such documents is challenging. Platforms, governments, researchers, and civil society need to continue to collaborate on future documentation efforts, making this an area ripe for Public-Private Partnerships. By pooling knowledge, especially about emerging trends, symbols, and language used by digital extremists, the digital ecosystem can become more adept at recognising and addressing potential threats. Shared databases and definitions, including hateful or extremist language indices that platforms can use for moderation efforts, are pragmatic contributions that can be built towards that end.

Reporting Protocols

Evaluating the effectiveness of platform action against extremist and terrorist content is critical for effective regulation and public safety. Upon the identification of such content, platforms should be mandated to act within a reasonable specific timeframe. For example, the 24-hour timeline for removal of alleged illegal content incentivised under the EU Code of Conduct on Countering Illegal Hate Speech Online (2016) and referenced as precedence setting in the DSA (Regulation 2022/2065(62); (87); Ars. 9; 17) is widely seen as unrealistic given the volume of content. A potential other tactic – discussed in the PrOCoM textbox below – draws on Anti-Money Laundering (AML) regulations to identify long-term trends and actions, further to applying punishments as required to incentivise prompt action internally at firms under the compliance regime.

In any case, this report finds that platforms should release more comprehensive transparency reports that are detailed and comparable across platforms. While these are a laudable component of the DSA, respondents noted that many are not sufficiently disaggregated at present. Clarity is needed on what constitutes meaningful transparency: the 2023 GIFCT Working Groups have sought to improve efforts in this regard, but more work is needed (gifct.org/year-three-working-groups). To be maximally useful to civil society and academics, these should detail the quantity and nature of extremist content identified and how they have addressed it. Percentages should be avoided (what does a 99 % removal rate mean, after all?), and expressed instead as full numbers of content by type and by platform involved (grouping a game platform data inside a larger company's corporate report blurs distinctions). For instance, while Xbox releases data on hate speech, it does not always distinguish between general hate speech and extremism-driven hate speech. Consistency in reporting such

distinctions is crucial. Standardised and comparable metrics across platforms are also critical. Instead of offering fragmented data, platforms should provide uniform metrics for cross-platform comparison.

Furthermore, platforms need to be subject to more stringent oversight and accountability mechanisms, which the Commission controls at present along with relevant national authorities. These independent oversight bodies should also assess the effectiveness of moderation efforts on gaming platforms – through public-private dialogue to ensure maximal compliance. Penalties – more nuanced than what are currently in place via the DSA – should also be in place for platforms that recurrently fail to tackle extremist content efficiently. For example, offenders should be held financially accountable for extremist and terrorist content if they have been alerted to its presence and have taken no action. Prolonged negligence in this regard should be penalised to the utmost possible – while pushing for immediate takedowns, unless constituting an imminent threat to life, should be minimised. On this point, the challenge of non-compliant platforms is significant. Some platforms, despite being prominent hubs for extremist content, may be less inclined to cooperate. A comprehensive framework is needed to ensure that even these platforms adhere to set standards.

PROCOM (THE NETHERLANDS)

The Online Content Moderation Project (PrOCoM) provides a public-private framework which may also be used to engage with gaming platforms within which citizens, the government and the internet sector can more easily take action with regard to online material that is criminal, causes harm or has socially undesirable effects. To achieve this, the PrOCoM has four operational goals:

- Public-private cooperation: setting up a structure for public and private parties to cooperate on online content moderation.
- Low-threshold reporting facility for citizens: establishing a low-threshold facility for citizens to be able to report unlawful online content.
- Local Approach:
    - > Having tools available for local administrators to prevent or mitigate public order problems instigated online.
    - > Providing advice for how local government can legally request for content to be removed.
- Knowledge & Skills: building and sustainably maintaining knowledge and skills on the workings of the internet, online phenomena and content moderation.

CONTEXT

The introduction of the Digital Services Act (DSA) expands the scope of responsibilities related to illegal online content beyond what was outlined in the Electronic Commerce Directive. While freedom of expression is valued, it is acknowledged that limits can be set, provided they are legally justified, proportionate, and subsidiary. The challenge lies in striking a balance between allowing the internet sector to self-regulate and preventing a scenario where it dictates the boundaries of public discourse. The Dutch hybrid model, combining public and private efforts, has proven effective in addressing online child sexual abuse, serving as a potential blueprint for regulating various forms of online content that require standardisation, monitoring, and enforcement. The absence of explicit standards and enforcement

measures online has created opportunities for malicious actors to disseminate harmful material, potentially leading to significant societal consequences. To counteract this, the Proactive Online Content Moderation (PrOCoM) initiative seeks to establish a public-private framework for citizens, the government, and the private internet sector to collaboratively address punishable, harmful, or socially undesirable online content. This initiative encompasses the development of a fundamental infrastructure for processing removal requests based on criminal, civil, or administrative law, as well as the provision of tools and knowledge to facilitate timely interventions in a socially responsible manner. Ultimately, PrOCoM aims to serve as a model of effective online content regulation both nationally and internationally.

PrOCoM is based on lessons learnt from the banking sector and Anti-Money Laundering (AML) enforcement techniques that could be mirrored for gaming and other online platforms. Suspect Activity Reports (SARs) have emerged as an essential tool for guiding law enforcement's information-gathering process. The underlying principle is to collate data, often proactively, before an act translates into a verifiable illegal action. This pre-emptive approach, initially launched by financial institutions, has morphed over time from a largely investigative procedure to a more automated one. Collaboration between banks and law enforcement has streamlined the reporting process, enhancing efficiency while ensuring that the necessary data reaches the appropriate authorities promptly.

An important recommendation, informed by this AML framework, is the adoption of a similar risk assessment practice for digital platforms, echoing the protocols in AML and Counter-Terrorism Financing (CTF) and Know Your Customer (KYC) modalities. Instead of solely focusing on singular content takedowns, the emphasis could be on assessing and reprimanding platforms that consistently fail to address extremist content. This strategy would encourage platforms to develop systematic, holistic mechanisms to combat such content rather than sporadically addressing individual instances.

Algorithmic Amplification

Algorithmic amplification, as it pertains to the dissemination of extremist content, is an area of rising concern for EU and international policymakers. For example, discussions in various fora among those interviewed for this paper indicate an increasing acknowledgement of the impact of recommendation algorithms, especially when they unintentionally bolster extremist content. There is a recognised need among most of the actors reached for this study to establish frameworks for 'Algorithmic Accountability.' These frameworks would audit recommendation algorithms – including those inside games – to ensure they do not inadvertently promote extremist or harmful content. While the DSA does provide provisions on recommender systems (Regulation 2022/2065 (84); Ars. 27; 34), the content would need to fall under the DSA for those provisions to apply.

Additionally, as the world has learnt over the last several years from new AI and machine learning scandals, incorrect or biased data can create algorithms that perpetuate stereotypes or false narratives. As such, policymakers should push for open access and legibility of training data used for algorithm recommendation engines, including use cases particular to the gaming sphere. In the gaming sphere, these include:

- Game match recommendations, whereby users are paired with other people to play with online.
- Game recommendation engines where users are shown games they might want to play or purchase.
- Server recommenders that match users with particular servers to play a given game in.

- As games may increasingly be designed or co-produced with generative AI, understanding the training data sets used for the generative models will be critical, as flawed or biased datasets have been shown to reproduce racial stereotypes, extremist narratives, and disinformation (Bunker & Bunker, 2023).

By implementing regulatory frameworks to monitor and manage these algorithms, the inadvertent promotion of harmful content can be significantly reduced. For instance, the U.S. proposed Algorithmic Accountability Act offers a potential approach to tackle such issues (Clarke, 2022). This legislation emphasises the importance of transparency and accountability for automated decision systems. In any event, future regulatory efforts and PPPs need to take into account how algorithmic recommendations and machine learning models are used in games and gaming spaces.

PREVENT

Beyond facilitating a better understanding of the threat, improving regulatory clarity, and enhancing dialogue efforts, this report sees ample room for Public-Private Partnerships to prevent violent extremism on gaming platforms. Broadly speaking, four avenues for prevention programming became clear during the interviews and research:

- Education and capacity-building efforts targeted at industry, policymakers, law enforcement, and educators alike.
- Positive interventions, such as digital street work models that shift social work and prevention professions into online gaming environments and which work on addressing online toxicity in games.
- Strategic communications campaigns, including PSAs and behavioural nudges on gaming platforms designed in partnership with public safety actors.
- Technical assistance for smaller platforms and studios to enhance their content moderation and trust and safety efforts, as well as local municipal governments and law enforcement to enhance their existing prevention programming.

These four prevention pillars for PPPs are discussed in depth throughout the sections below.

Education and Capacity-Building

There is an evident need for education and capacity-building across multiple sectors to combat the growing threat of extremism within gaming platforms. On the one hand, those interviewed for this report expressed a clear call to fund digital literacy programmes, which would equip users with the tools to recognise and report extremist narratives (this approach is expanded on in the strategic communications portion below). On the other hand, many of the respondents emphasised a knowledge gap faced by industry and public actors alike: gaming companies may not know when and what information is significant enough to warrant alerting authorities about potential threats, while law enforcement may not know enough to take alerts from online games or adjacent platforms seriously. To counter this, there is a need to channel knowledge about gaming platforms and the tactics extremists use into the hands of several key audiences: regulators and policymakers, industry, law enforcement and social workers, and educators/parents.

For regulators and policymakers: The regulators reached for this study emphasised how imperative internal expertise in the gaming sector is for their efforts. This is not merely about knowing how games work, but

comprehending the nuances of in-game communication, player communities, in-game economies, and how these platforms can be exploited. To fill knowledge gaps, regulators and policymakers can consider hiring individuals who have hands-on experience within the gaming sector or consult with subject matter experts and trainers. Engaging with these professionals can offer insight into the complexities of the gaming environment and the challenges faced in moderating content. Alternatively, cross-training PPPs can be set up with the industry to share granular knowledge on platform functionalities.

For industry: Gaming companies might not be fully equipped to identify and deal with violent extremism and terrorism. Interviewees suggested that these companies are often not even aware of the kind of information vital for law enforcement actors. They need training and collaboration with experts from law enforcement, academia, and civil society. Industry representatives also shared the significance of learning from experts to understand the nature and scope of threats and the best measures to counteract them. Cross-training efforts with regulators and law enforcement can help towards this end as well.

Law enforcement and social workers: In many European contexts, local law enforcement and social workers play a crucial role in early detection of radicalisation, as some of the interviewees explained. However, there is a challenge in the form of apprehension towards the digital and gaming realm: reporting threats online does not always result in action from the police, indicating a significant gap in awareness and capability, while frontline social workers are often not engaged with regarding digital gaming environments. This knowledge gap may be bridged via educational exchanges between industry, law enforcement, and experts on extremism and gaming.

Educators and parents: Experts also advocated for preventative education on online harms to be provided in classrooms. Early intervention can prevent an individual from heading down a dangerous path, while educators and parents are often best placed to help. This involves not just identifying grooming or recruitment for extremism but broader online threats. Teachers need better information to help their students and parents understand online harms and pro-social benefits found in gaming settings. Curricula about these risks could be incorporated into teacher training or deployed via innovative PPPs with tech firms and non-profits working on digital literacy.

Strategic Communications

Shared communication strategies, formulated collaboratively between public entities and private companies, can play a pivotal role in P/CVE work (Lamphere-Englund and Vugteveen, 2022). Interviewees drew parallels from the British public transport's awareness campaign, which uses the straightforward mnemonic "see it, say it, sorted," for public safety awareness on buses, tubes, and other mass transit, an equivalent strategy might be developed for online platforms (RAN Interview #17). Such a strategy – designed between public safety actors and a gaming platform or industry body – could empower community managers at gaming companies and users alike. By adopting simple, non-controversial language, platforms can encourage users to report suspicious activities, such as extremist or suspicious terms in chats or attempts to move discussions to external private forums.

Another pertinent example of this flagged during interviews was a government-led campaign addressing loneliness, particularly pronounced during the COVID-19 pandemic. Recognising the communal nature of online gaming, where players converse and form bonds, the campaign smartly integrated gaming elements. By merely incorporating imagery of video game controllers and generating related content, the campaign effectively engaged a broader audience without altering its core messaging. There have also been initiatives focused on promoting positive behaviour among men in online spaces. The objective is to encourage gamers to identify and call out misogynistic behaviour, further fostering an inclusive and respectful digital environment. The salient

point is that gaming is integral to today's digital discourse. If the industry and public safety actors are not actively engaged in shaping P/CVE narratives, there is a risk of disseminating misinformed content. It is crucial, therefore, for these campaigns to be formed collaboratively, involving stakeholders who understand the nuances of the gaming world.

Partnering with government agencies and gaming companies can help form informational hubs, disseminating critical safety guidelines. Combined with technical backend strategies, such as those used to combat childhood sexual exploitation and abuse online, communications campaigns for P/CVE can be effectively rendered in gaming environments. By merging effective messaging with practical actions and designing these interventions collaboratively with the industry, the chances of wider buy-in and user safety increase substantially. Such PPP initiatives position gaming companies not just as entertainment providers but also as vital channels for disseminating information and promoting online safety.

Positive Interventions

The concept of positive interventions in gaming spaces has been discussed in previous RAN publications (Schlegel, 2021b), although several other potential avenues for PPPs emerged in our conversations. For example, 'digital street work' has emerged as a novel approach in combatting extremism and fostering positive online behaviours. This digital street work can be visualised as the online counterpart of traditional on-the-ground outreach by social workers, specifically aiming at those at risk of radicalisation or engaging in harmful activities.

In the Netherlands, much of the responsibility for counter-radicalisation efforts is vested in local entities, especially municipalities. They are aided by the National Support Centre for Extremism, (*Landelijk Steunpunt Extremism*) which is central to enhancing digital street work practices. While certain best practices have been established, work is ongoing to refine these methodologies with specific regard to online extremism, even within spaces like gaming platforms. This emphasis on tailoring interventions to the online realm underscores the evolving nature of the threat and the necessity of adapting countermeasures to new digital arenas.

Supplementary projects in this field have experimented with game development as an innovative tool for intervention. Games are being developed to address issues related to democratic values, extremism prevention, and historical education. An instance of such an initiative is the *Well Played, Democracy* project in Germany, an initiative of the Amadeu Antonio Stiftung which not only integrates educational elements into its design but engages in digital street work. Such actions underline the potential of gamification as an avenue for positive engagement and intervention.

From a broader perspective, the essence of effective digital intervention might lie in 'nudge theory.' Nudging refers to the making of indirect suggestions or exerting positive reinforcements that can influence individuals' behaviour and decision-making, without coercing them or mandating or restricting their choices. By redirecting users towards more constructive content, especially those bordering on content that violates community guidelines, platforms and public P/CVE actors can play an instrumental role in combatting extremism. This is not about forcing a particular perspective on users but providing them with a diversified array of content that helps them escape echo chambers and exposes them to varied viewpoints. As digital spaces continue to grow and influence behaviours, proactive interventions such as digital street work, gamified education, and the application of nudge theory can serve as robust tools in the ongoing fight against online radicalisation and extremism.

Technical Assistance to Platforms and Local Actors

While it is commonly acknowledged that Public-Private Partnerships can serve as a mechanism to address these challenges, the focus often remains on larger companies. However, it is crucial to recognise the unique vulnerabilities of smaller platforms and consider how government interventions can support them.

A fundamental challenge is that smaller platforms – and especially indie or small game studios – often do not have the resources to establish robust trust and safety measures. While large companies are approached for partnerships by public authorities, smaller organisations might find themselves excluded from such dialogue, simply due to a lack of resources. Even if willing, a small development team is ill-equipped to meet the stringent requirements often set forth in these partnerships, especially without external support.

Public authorities have an opportunity to foster conditions that enable Public-Private Partnerships across the board. For instance, some jurisdictions have funds designed to help smaller organisations comply with regulations. Subsidies could be granted to small platforms for hiring trust and safety experts, thereby bridging the expertise gap. Initiatives such as Tech Against Terrorism (Europe), the ALLIES project,[6] and FRISCO[7] go a long way in this effort and should be supported to expand to help the gaming sector. Others, such as the Extremism and Gaming Research Network (EGRN), of which, in full disclosure, the authors are both members, can also assist towards this end.

In addition to funding, a more proactive approach by public grant-making organisations can make a significant difference. Rather than passively waiting for grant applications, these entities could actively identify and offer support to rapidly growing platforms and multiplayer games that are seeing a rapid increase in popularity. Such a model would relieve the burden on smaller companies, allowing them to focus on scaling their services while ensuring compliance with trust and safety norms.

In our discussions regarding the role of online platforms in ensuring trust and safety, an important aspect that emerged was the variance in resources and capabilities across different platforms. While large corporations have long been the focus of discussions around public-private partnerships for safety, a significant challenge arises when addressing smaller platforms with limited resources. A notable example of this was seen when large platforms like Epic Games, Riot Games, and Activision Blizzard were subject to scrutiny over their (lack) of extremism policies by U.S. lawmakers in 2022 alongside the developers of Among Us, a small team that faced overwhelming challenges given its exponential user base growth (Trahan, 2022). The demand to maintain rigorous safety and security measures, while undeniably crucial, is not always feasible for platforms that have seen rapid growth but have limited resources and manpower.

This scenario underscores the potential for governments to play an active role in creating conditions conducive to the establishment of effective partnerships. One suggestion has been that government bodies could establish additional funds for small organisations to help them navigate and implement necessary safety regulations. Such a proactive approach could be more than just providing financial assistance: it would also involve proactively identifying platforms that are experiencing rapid growth and pre-emptively offering them the support they need. This would address the problem faced by developers like Among Us, which, despite their significant user base, might lack the resources to seek out government grants or support.

[6] https://www.alliesproject.com/

[7] https://friscoproject.eu/

CHAPTER 3: GAMING SECTOR GUIDANCE

The previous chapter outlined how public sector authorities can offer guidance to private industry to address extremism on their gaming(-adjacent) platforms. Private companies can similarly take steps to strengthen public actor efforts to combat this multifaceted problem. To begin with, a comprehensive approach is paramount. Based on interviews with industry experts and public policymakers, publicly acknowledging and denouncing the issue demonstrates a commitment to creating a safer gaming environment, thus acting as a deterrent to potential extremists. Equally important is the continuous effort to attain deeper insight through research, allowing companies to refine their strategies effectively. Improving the overall user experience by fostering an engaging and inclusive atmosphere discourages extremist behaviour, achieved through user-friendly interfaces and community-building initiatives. This effort is strengthened by the implementation of robust reporting mechanisms to empower users and to work towards community-based moderation. These mechanisms should be informed by rigorous content moderation policies, providing clear guidelines to swiftly address extremist behaviour.

Effective content moderation, powered by advanced technology and human oversight, strikes a balance between removing extremist content and safeguarding legitimate expression. Trust and transparency form the bedrock of this endeavour, as companies openly communicate their policies and enforcement actions, fostering a sense of security within the gaming community. Digital literacy education empowers users to identify and report extremist content, equipping them with critical thinking skills and responsible online behaviour. Collaborative efforts with external entities and standardisation across platforms further enhance the industry's collective ability to combat extremism. By adopting this multifaceted and united approach, private gaming companies can create safer, more inclusive gaming experiences, strengthening Public-Private Partnerships (PPPs) to prevent and counter violent extremism. Accordingly, this chapter is divided into the following steps:

1. **Acknowledging and denouncing the issue**, which argues that companies should openly acknowledge the presence of extremism on their platforms and publicly denounce such behaviour.

2. **Improving knowledge of the problem**, by investing in research and data analysis to gain a deeper understanding of the nature and extent of extremism within their platforms.

3. **Improving online user experiences**, deterring potential extremists by creating a more engaging and positive environment. This includes features like intuitive interfaces, fair gameplay, and community-building tools.

4. **Providing easy reporting mechanisms**, establishing comprehensive and user-friendly reporting mechanisms across platforms and how to implement these systems.

5. **Effective content moderation,** employing various methods built on strict moderation policies that ensure that extremist content is swiftly identified and removed, while legitimate expression is protected.

6. **Building trust and transparency**, which notes that establishing trust with the Public-Private Partners, such as government, civil society, industry, and platform users is crucial.

7. **Digital literacy education**, recognizing importance of providing resources and education on digital literacy that empowers users to recognise and report extremist content. This education should include information on critical thinking, fact-checking, and responsible online behaviours.

8. **Cooperation with external entities**, advocating for more intensified collaboration with external organisations, such as NGOs, research institutions, and government agencies, can bring additional expertise and resources to the fight against extremism.

9. **Cooperation and standardisation across platforms**, where private gaming companies should work together to share best practices, insight, and technologies to combat extremism. This collaborative approach can lead to more effective solutions across the industry.

ACKNOWLEDGING AND DENOUNCING THE ISSUE

Acknowledging and denouncing the issue of extremism on gaming(-adjacent) platforms is paramount for private gaming companies due to their pivotal role in shaping the online gaming culture. While extremely diverse, this culture has been subjected to instances of racism and misogyny. Depending on the online space in which they are espoused, these issues have frequently become normalised, creating an environment where discriminatory behaviour can persist unchecked. Gaming companies set a clear and necessary precedent by openly acknowledging and denouncing extremism, signalling that such behaviour is not condoned or accepted within their communities.

These declarations could carry significant weight as they serve as a foundation for developing robust policies. By publicly recognising the existence of extremism, companies establish an imperative to act. This acknowledgement enables them to allocate resources, invest in technology, and implement training programmes to combat extremist behaviour. It also sets a standard for the community, clarifying that discriminatory actions should have no place in the gaming environment.

Furthermore, these declarations serve as a crucial starting point for downstream efforts. They provide a framework for educating and training frontline managers on effectively identifying and addressing extremist content. By openly stating their commitment to combatting extremism, private gaming companies can send a powerful message to their user base and community managers about the seriousness with which they approach this issue. This, in turn, informs the development of training programmes that equip managers with the knowledge, tools, and protocols necessary to navigate and manage extremist content in real time.

Denouncing violent extremism on platforms by companies has economic benefits as well. New, preliminary research has shown that adverse impacts on the revenue streams of gaming companies exist when they fail to address violent extremism on their platforms (Kowert and Kilmer, 2023). When extremist content flourishes unchecked, it creates an unwelcoming and unsafe user environment, deterring potential players and driving away existing ones. This erodes user trust and can lead to declining user engagement and retention rates. Moreover, the negative publicity surrounding extremist incidents can tarnish a company's reputation, potentially alienating advertisers and partners. By actively preventing and countering violent extremism, gaming platforms safeguard their user base and uphold their brand integrity, ultimately bolstering their long-term financial viability and sustainability.

Acknowledging and denouncing extremism in gaming is foundational to driving positive change. It sets the tone for policy development, informs resource allocation, and empowers frontline managers to take effective action. By making a stand against extremism, private gaming companies not only improve the safety and inclusivity of

their platforms but also send a powerful message about the values and standards they uphold within the broader gaming culture.

IMPROVING KNOWLEDGE OF THE PROBLEM

Our interviews with public and private actors evidenced a clear need for investment into knowledge generation surrounding gaming(-adjacent) platforms. As discussed in Chapter 2 (par. Understand), governments can act as 'innovator catalysts' by facilitating data and platform access for vetted researchers. Ultimately, though, it is the private gaming companies that provide this access. Therefore, this report recommends that these companies work with civil society organisations and academia, granting access to data and promoting research into violent extremism on their platforms. Strengthening research efforts into violent extremism on gaming-adjacent platforms is imperative, and the knowledge gained is pivotal in crafting effective on-platform counter-extremism policies and tailored interventions. Providing researchers with access to Application Programming Interfaces (APIs) for these platforms will be instrumental in facilitating comprehensive data collection and analysis. APIs enable researchers to obtain real-time, granular information regarding user interactions, content dissemination, and community dynamics, allowing for a nuanced understanding of extremist networks and strategies.

A frequently cited consideration by private industry representatives regarding platform research was privacy and data protection. Research into violent extremism on gaming-adjacent platforms must comply with privacy and data protection laws, as these are crucial to ensure that the pursuit of security does not infringe upon individuals' rights. There are a number of key steps that can be taken to guarantee compliance:

- **First**, it is imperative to establish strict protocols for data anonymisation and aggregation, ensuring that personally identifiable information is never disclosed. Researchers should only have access to de-identified data that cannot be traced back to specific individuals.

- **Second**, obtaining informed consent from platform users, where feasible, should be a priority, especially when conducting studies involving sensitive content. Transparency in research goals and methodologies is paramount in building trust between the research community, platform operators, and users.

- **Third,** adherence to established legal frameworks, such as the General Data Protection Regulation (GDPR) in the European Union or equivalent laws in other jurisdictions, must be rigorously upheld. This includes conducting thorough privacy impact assessments. Collaboration between researchers, policymakers, and privacy experts is essential in crafting guidelines that strike the right balance between research objectives and individual privacy rights. Regular audits and reviews of research practices can help ensure ongoing compliance with evolving privacy standards. The goal is to harness critical insight while respecting and upholding individuals' privacy and data protection rights in accordance with legal and ethical obligations.

IMPROVING ONLINE USER EXPERIENCES

The improvement of online user experiences will not only improve the overall quality of player engagement with a video game, but the establishment of intuitive interfaces and community-building/moderation tools can also deter extremists from exploiting online gaming spaces. First, providing a user-friendly platform may foster engagement in positive player activities and stymie toxic behaviour. Community-building tools and effective community moderation also actively cultivate a community that values diversity, inclusivity, and positive engagement, increasing platform accessibility and empowering players to express themselves and participate meaningfully. These tools can be pivotal in forging connections, allowing users to form relationships, share experiences, and collectively contribute to a vibrant digital ecosystem. The resulting sense of belonging and

shared purpose is a powerful counterforce against extremist narratives, as individuals find support and validation within the community. These enhancements would bolster the platforms' appeal and foster an environment that actively resists the allure of extremism, thus promoting a safer, more constructive online space for all users.

FAIR PLAY ALLIANCE

The Fair Play Alliance is a coalition of video game companies, developers, and other stakeholders in the gaming industry aimed at encouraging healthy online interactions and combatting disruptive behaviour, including hate speech and harassment, in online multiplayer games. The alliance focuses on sharing research, providing resources, and promoting best practices to help create more inclusive gaming environments. Its members often collaborate on various projects and initiatives, emphasising the value of collective action in achieving its mission. Although the Fair Play Alliance is not focused on extremism, its work could contribute to reducing the incidence or impact of extremist behaviour within gaming spaces.

PROVIDING EASY REPORTING MECHANISMS

Gaming companies can play a pivotal role in mitigating online extremist behaviour by implementing robust reporting mechanisms that empower users to flag content or behaviour swiftly and effectively. First and foremost, companies should prioritise the development of user-friendly reporting tools that are easily accessible within the platform's interface. This includes clear instructions on how to report, along with options for categorising the nature of the concern, such as hate speech, extremist propaganda, or harassment.

Establishing a responsive reporting system that acknowledges receipt of the report and provides regular updates on the investigation process is also imperative. Transparency builds trust and encourages users to actively participate in reporting extremist content. Additionally, companies should commit to timely reviews and acting on reported incidents, with well-defined procedures for escalating and addressing severe cases in place.

To ensure the efficacy of reporting mechanisms, companies should also invest in training and equipping content moderators with the necessary tools and resources to assess and handle extremist content appropriately. This includes providing them with clear guidelines on identifying and responding to extremist behaviour in accordance with community standards and legal frameworks.

Moreover, transparency reports detailing the number and nature of reports received and actions taken should be periodically published, as noted in our section on PPPs and public action above. Companies need not wait for regulatory requirements to enter into force, with such a proactive approach demonstrating a commitment to combatting extremism on their platform.

By fostering a culture of user vigilance and collaboration, along with responsive reporting mechanisms, gaming companies can actively contribute to creating a safer, more inclusive online gaming environment, effectively deterring and mitigating extremist behaviour. Unfortunately, this is currently not always the case. A report published by the UNOCT (Schlegel and Amarasingam, 2022) indicated that many surveyed players do not report toxic behaviour or problematic content, citing that many such reports resulted in no action being taken against the player committing the offence. Instead, a far more common reaction to these behaviours was simply ignoring the issue.

Another consideration in establishing effective reporting mechanisms is that the onus is placed on individual users rather than addressing systemic gaming issues related to toxic or extremist content online (Hartgers & Leidig, 2023). There is a risk of creating a tacit expectation that gaming environments must become entirely self-regulated spaces and that the responsibility of creating safe online spaces falls on individual users rather than on the gaming companies themselves. Nevertheless, establishing mechanisms that empower gamers to take a stance against online harms – such as those found in the EUIF videogaming handbook and inspiring practices – is a step in the right direction, though certainly not the only one to take.

EFFECTIVE CONTENT MODERATION

Private companies can offer additional guidance by establishing effective content moderation. Effective content moderation is a cornerstone in combatting online extremism, achieved through a multifaceted approach founded on stringent moderation policies. These policies serve as the bedrock, setting clear guidelines and standards for identifying and promptly removing extremist content, all the while safeguarding legitimate forms of expression (Roberts, 2019). Leveraging a diverse array of methods, from automated algorithms to human moderators, ensures a comprehensive and dynamic strategy. Automated systems can swiftly flag potentially harmful content, allowing human moderators to conduct nuanced assessments and make contextually informed decisions. By combining technology with human oversight, platforms can strike a balance that upholds free speech rights while effectively curbing the dissemination of extremist ideologies, ultimately fostering a safer and more constructive online environment.

Developments in AI are often looked at for their potential application in content moderation. However, this is not a panacea and human intervention will continue to be necessary for the foreseeable future until algorithmic moderation becomes sufficiently effective in detecting coded language or content and toxic or extremist expressions in voice chat. Ethical considerations related to questions of how particular AI content moderation tools are trained or what implicit or explicit biases may be present in AI also pose challenges to the application of algorithmic moderation (Schwartz et al., 2022). As such, private companies must continue to invest in human resources around trust and safety.

Another area of investment for private gaming(-adjacent) companies are platform policies that are explicitly established to counter (violent) extremism. This ties in with the previous paragraph that discusses acknowledging and denouncing the issue, and it must again be emphasised that without specific policies in place to address extremist content, moderation – be it algorithmic or human-driven – cannot sufficiently address extremist content online. Several respondents lamented the lack of policies that address violent extremism and expressed a desire for private companies to take a more proactive approach.

"I think maybe [one could] think of adding a policy as low-hanging fruit, which in some cases, in some respects it is, but in other respects it needs to and should reflect a bunch of downstream work around defining and putting in place the ways in which educating your frontline managers that this is the new policy…and this is how you deal with it." (RAN Interview #7, 2023).

BUILDING TRUST AND TRANSPARENCY

Another aspect in which private platforms can provide guidance in addressing extremist content online is through trust-building and transparency efforts. This effectively goes two ways: companies engage in content moderation reporting with public authorities as captured in the transparency reporting obligations of the DSA which mandates those very large online platforms (VLOPs) or very large online search engines (VLOSEs) report on takedowns and illegal content; and VLOPs and VLOSEs communicate the terms and conditions of their

platforms to their user base. To achieve these objectives, transparent communication channels should be established to share insight, data, and best practices in content moderation and user behaviour analysis. This fosters a shared understanding of the challenges and potential solutions, enhancing the collective efficacy of counter-extremism efforts. Moreover, regular reporting on progress and outcomes helps build accountability and confidence in the partnership. For example, public awareness campaigns, jointly conducted by private companies and public entities, can play a crucial role in educating users about the risks of extremist content and how to report it. By creating a cohesive, transparent, and accountable framework, public-private partnerships can maximise their impact in combatting extremism on gaming platforms. Several examples can be found in the previous chapter on public actor strategic communication guidance.

DIGITAL LITERACY EDUCATION

We already discussed the importance of cross-cultural education in the previous chapter and the need expressed by industry interviewees to leverage the existing expertise of civil society and law enforcement to build awareness around extremist content and activities on gaming platforms. It should be emphasised, however, that industry can also assist education and capacity-building efforts by lending expertise to public authorities and civil society alike. While it was noted by interviewees that – barring a few examples – private companies often do not employ counter-terrorism or counter-extremism, subject matter experts and private industry can nevertheless strengthen digital education by sharing the necessary information to create educational curricula. For example, by leveraging the experiences of community managers, general trends of harmful content online and user behaviour patterns related to extremism could be detected and analysed accordingly. Moreover, industry experts can inform educational efforts by mapping the technical aspects of gaming platforms and train policymakers on the nuances of gaming culture. It was indicated by several interviewees that public authorities lack a general awareness and understanding of gaming spaces. Industry can address this gap by proactively sharing knowledge of their platforms with these public officials: “the other pillar is that if you talk to politicians, or to police enforcement, they actually don't know anything about the gaming communities.” (RAN Interview #2, 2023).

In addition, the private industry can support the implementation of education and dissemination of materials developed by civil society partners. For example, tools and features can be developed to empower players to recognise and report extremist behaviour: the EUIF videogaming handbook and a RAN paper by Linda Schlegel on *Countering the misuse of gaming-related content & spaces* have good practices on this. Public awareness campaigns co-designed by industry, government, and civil society can also be delivered in gaming spaces. As discussed previously, one participant cited an example of gaming companies and a government having worked together to stop loneliness during the COVID-19 pandemic. This was achieved by promoting the playing of video games and connecting online to counteract physical and emotional separation due to lockdowns (RAN Interview #17, 2023). The private industry should continue to work with public partners to address social issues, such as mental health, bullying, and, importantly, extremism.

COOPERATION WITH EXTERNAL ENTITIES

This paper strongly encourages public and private sectors to collaborate with a view to addressing extremism on gaming platforms. In Chapter 2 (par. Prevent), the authors elaborated on the domains where Public-Private Partnerships are possible, specifically noting: Education and Capacity-Building, Strategic Communications, Positive Interventions, and Technical Assistance to Platforms and Local Actors. From an industry point-of-view, many of these domains were already outlined in the discussion above: the role of private companies in the provision of expertise that strengthens education efforts, as well as their delivery; the establishment of

communication campaigns and positive interventions to address social issues, such as loneliness; and researcher access to platform data to further study and analyse trends and user behaviours. The question, nevertheless, remains how such partnerships should be established and what they should look like.

Overall, the willingness to engage with public partners was shared by all industry respondents. A number of areas were identified in which such collaboration may occur, including but not limited to: information-sharing of trends and threat analyses between public and private actors, multi-stakeholder engagement across public and civil society sectors, civil society engagement through funding of research and data-sharing, resilience-building of player bases by generating awareness and improving online user experiences and reporting mechanisms, and stakeholder convening through international forums like the EUIF and GIFCT. However, several concerns interfacing with each of these areas of potential partnerships were also expressed, specifically related to regulatory clarity and balancing privacy and safety.

Regulatory Ethics

A concern regarding regulation pertained to perceived overreach and regulatory abuse. To be clear: the question of the extent to which governments should be involved in content removal and moderation was raised by public sector participants as well. Overall, there was a common understanding that governments should apply some constraint in the moderation of content, with human rights and free speech as crucial guiding principles. The regulatory abuse was not necessarily raised in relation to the EU – rather, it was brought into question with regard to non-EU countries that are found to frequently transgress human rights. The establishment and implementation of EU legislation often have far-reaching consequences, and the externalisation of European laws – sometimes referred to as the 'Brussels effect' – may result in such non-EU countries adopting similar legislative frameworks. Without proper respect of fundamental human rights, however, these adopted frameworks may be instrumentalised to remove dissident speech and prosecute political opposition on the grounds of spreading 'misinformation', sharing 'terrorist content', or other offences outlined in adopted legislation. While not necessarily applicable to the EU context, private industry and public regulatory partners alike considered this a challenge.

Balancing Privacy and Safety

Stricter content moderation and regulatory measures, while imperative in the fight against extremism, raise legitimate concerns about potential encroachments on the right to freedom of speech (Gillespie, 2018). This apprehension is well-founded, as such measures might be perceived as biased or censorious. Thus, ensuring that any forthcoming policies and guidelines are characterised by transparency, impartiality, and a steadfast commitment to upholding users' fundamental right to express themselves freely is crucial. Private sector participants noted that data-sharing agreements prove to be challenging, given that companies prioritise user privacy and data protection. The dual commitment must be to identify and address harmful content while treading cautiously to avoid intrusive examinations into potential networks. Striking the delicate balance between safeguarding user privacy and upholding community safety involves navigating a complex terrain of policy and legal considerations. Platforms with social features, where the emphasis lies on nurturing genuine connections and forming communities, are particularly confronted with this. Maintaining this nuanced balance in such spaces is paramount to cultivating a secure and inclusive environment.

COOPERATION AND STANDARDISATION ACROSS PLATFORMS

Cooperation and standardisation across platforms can support the fight against extremism on gaming platforms for a number of reasons. Given the cross-platform nature of extremist content and outlinking practices, extremist content and activities are spread across platforms. Setting industry standards and engaging in cross-industry collaboration allows private gaming companies to pool their collective knowledge and resources to develop more effective and cohesive strategies against extremism. Sharing best practices, insight, and technologies allows for a comprehensive approach to identifying, moderating, and preventing extremist content, thereby safeguarding online gaming communities.

Moreover, establishing industry standards is a critical step towards creating a unified and consistent framework for content moderation. Participants noted good practices such as the International Organisation for Standardisation (ISO) and its creation of standards, which may be similarly applicable to gaming spaces. Standardisation can moreover promote transparency, accountability, and uniformity in content moderation practices, ensuring that all gaming platforms adhere to a shared set of guidelines. Additionally, industry standards can enhance interoperability between platforms, facilitating the exchange of critical information and resources. This collaborative approach not only streamlines the sharing of expertise but also allows for the rapid adoption of innovative technologies and tools designed to combat extremism.

CHAPTER 4: CONCLUSION AND IMPLICATIONS FOR THE FUTURE

The landscape of online gaming platforms is becoming increasingly complex, serving as spaces for both community building and, regrettably, extremist exploitation. As shown in this paper, to mitigate the spread of extremist activities and enhance the overall safety and integrity of these digital ecosystems, a multi-pronged strategy involving both public and private sectors is imperative. Interviews with private and public representatives have indicated the need for a multifaceted approach to addressing extremism on gaming platforms, catering to both public sector authorities and private industry.

Public actors have a way to go in improving building trust and transparency in the fight against extremism, with companies engaging in content moderation often not practically engaging with public authorities. Shared, transparent research, coupled with improved dialogue channels foster a common understanding of the challenges and potential solutions, enhancing the collective efficacy of counter-extremism efforts. Similarly, collaboration with public entities and standardisation across platforms further enhance the industry's collective ability to combat extremism, allowing for the rapid adoption of innovative technologies and tools. Next, regulatory clarity, balanced with privacy and safety concerns, is a major area for improvement, necessitating common definitions and principles for content moderation. Lastly, improved prevention work through digital education, along with specialised training for regulators and platforms, is paramount, and private industry can contribute by sharing expertise to create educational curricula, just as law enforcement and regulators can help private actors understand extremism risks. Overall, this comprehensive strategy underscores the crucial role of both public and private sectors in creating a safer and more inclusive online gaming environment. As illustrated in Chapter Two, **for public actors**, there are at least four main pillars of PPP engagements:

1. **Understand** through research-focused PPPs that better evaluate risks and challenges of extremism on gaming platforms.
2. **Support** dialogue efforts for public and private actors so they can better come together and jointly decide how to address challenges.
3. **Regulate** via improved and nuanced requirements for limiting extremist exploitation of gaming platforms.
4. **Prevent** through developing partner-led P/CVE programming that can be deployed against exploitation taking place in gaming. Encourage collaboration in domains like strategic communications, positive interventions, and technical assistance to platforms.

Private companies, as this paper contests, also play a pivotal role in creating a safer gaming environment by publicly acknowledging and denouncing extremism, thereby setting a clear precedent against such behaviour. This acknowledgement forms the foundation for robust policies, resource allocation, and training programmes aimed at combatting extremist content. Leveraging a combination of automated algorithms and human moderators allows for effective content moderation, striking a balance between safeguarding free speech and ultimately building trust and transparency with Public-Private Partners, such as government, civil society, industry, and platform users.

1. **Sponsoring partnerships for literacy education** that empower users to recognise and report extremist content. This education should include information on critical thinking, fact-checking, and responsible online behaviours. Governments should collaborate with industry and civil society to create educational curricula around digital literacy, focusing on the risks posed by extremist content.

2. **Cooperate with external entities**, advocating for more intensified collaboration with external organisations, such as NGOs, research institutions, and government agencies, can bring additional expertise and resources to the fight against extremism.
3. **Cooperation and standardisation across platforms**, where private gaming companies should work together to share best practices, insight, and technologies for combatting extremism. This collaborative approach can lead to more effective solutions across the industry.

By embracing a holistic approach that combines regulatory clarity, policy development, digital literacy, and multi-stakeholder collaboration, both gaming platforms and governments can substantially mitigate the risks of extremist activities online. The authors also strongly believe that all work in this space should be rooted in the principles of fundamental human rights and free speech. This will, in turn, contribute to the development of more secure, inclusive, and resilient digital communities.

BIBLIOGRAPHY

Anti-Defamation League (ADL). (2022). Hate Is No Game: Hate and Harassment in Online Games 2022. In *www.adl.org*. https://www.adl.org/resources/report/hate-no-game-hate-and-harassment-online-games-2022

Anti-Defamation League (ADL). (2023). *Caught in a Vicious Cycle: Obstacles and Opportunities for Trust and Safety Teams in the Games Industry | ADL*. www.adl.org. https://www.adl.org/resources/report/caught-vicious-cycle-obstacles-and-opportunities-trust-and-safety-teams-games

APA Task Force on Violent Media. (2015). *Technical report on the review of violent video game literature*. https://www.apa.org/pi/families/review-video-games.pdf

Beutel, A., & Weinberger, P. (2016). *Public-Private Partnerships to Counter Violent Extremism: Field Principles for Action. National Consortium for the Study of Terrorism and Responses to Terrorism (START).* https://www.start.umd.edu/pubs/START_STate_PublicPrivatePartnershipstoCounterViolentExtremismFieldPrinciplesforAction_June2016.pdf

Bunker, R., & Bunker, K. (2023). *The Terrorism Potentials of ChatGPT & Related Generative AI Models*. C/O Futures. https://www.cofutures.net/post/the-terrorism-potentials-of-chatgpt-related-generative-ai-models

Chen, Z. Ethics and discrimination in artificial intelligence-enabled recruitment practices. Humanitieies and Social Sciences Commununications 10, 567 (2023). https://doi.org/10.1057/s41599-023-02079-x

Clarke , Y. (2022, February 7). *Clarke, Wyden And Booker Introduce Algorithmic Accountability Act Of 2022 To Require New Transparency And Accountability For Automated Decision Systems*. Congresswoman Yvette Clarke. https://clarke.house.gov/clarke-wyden-and-booker-introduce-algorithmic-accountability-act-of-2022-to-require-new-transparency-and-accountability-for-automated-decision-systems/

Cooper, C., Booth, A., Varley-Campbell, J., Britten, N., & Garside, R. (2018). *Defining the process to literature searching in systematic reviews: A literature review of guidance and supporting studies.* BMC Medical Research Methodology, 18(1). https://doi.org/10.1186/s12874-018-0545-3

Entertainment Software Association (ESA). (2022). *Essential facts about the video game industry 2022*. https://www.theesa.com/wp-content/uploads/2022/06/2022-Essential-Facts-About-the-Video-Game-Industry.pdf

European Union Agency for Cybersecurity. (2011). *Public Private Partnerships (PPPs)*. ENISA. https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ppps.

Davis, J. (2022, January 10). *National security expert Jessica Davis on terrorist financing in the 21st century*. The Hub. https://thehub.ca/2022-01-10/terrorist-financing-in-the-21st-century

Gillespie, T. (2018, January). Custodians of the internet: *Platforms, content moderation, and the hidden decisions that shape social media.* ResearchGate. https://www.researchgate.net/publication/327186182_Custodians_of_the_internet_Platforms_content_moderation_and_the_hidden_decisions_that_shape_social_media

Global Internet Forum to Counter Terrorism (GIFCT). (July 2021). *Broadening the GIFCT Hash-Sharing Database Taxonomy: An Assessment and Recommended Next Steps.* GIFCT. https://gifct.org/wp-content/uploads/2021/07/GIFCT-TaxonomyReport-2021.pdf

Gómez, Á., Brooks, M. L., Buhrmester, M. D., Vázquez, A., Jetten, J., & Swann, W. B. (2011). On the nature of identity fusion: Insights into the construct and a new measure. *Journal of Personality and Social Psychology*, *100*(5), 918–933. https://doi.org/10.1037/a0022642

Gómez, Á., Chinchilla, J., Vázquez, A., López-Rodríguez, L., Paredes, B., & Martínez, M. (2020). Recent advances, misconceptions, untested assumptions, and future research agenda for identity fusion theory. *Social and Personality Psychology Compass*, *14*(6). https://doi.org/10.1111/spc3.12531

Hartgers, M., & Leidig, E. (2023). *Fighting extremism in gaming platforms: a set of design principles to develop comprehensive P/CVE strategies. International Centre for Counter-Terrorism.*- ICCT. https://www.icct.nl/publication/fighting-extremism-gaming-platforms-set-design-principles-develop-comprehensive-pcve

Hoffman, B. (2017). *Inside Terrorism*. Columbia University Press.

Kaye, L. K., Kowert, R., & Quinn, S. (2017). The role of social identity and online social capital on psychosocial outcomes in MMO players. *Computers in Human Behavior*, *74*, 215–223. https://doi.org/10.1016/j.chb.2017.04.030

Kelly, S. (2021). Money Laundering Through Virtual Worlds of Video Games: Recommendations for a New Approach to AML Regulation. *Syracruse Law Review*, *71*(1487). https://lawreview.syr.edu/wp-content/uploads/2022/01/1487-1512-Kelly.pdf

Koehler, D., Fiebig, V., & Jugl, I. (2022). From Gaming to Hating: Extreme-Right Ideological Indoctrination and Mobilization for Violence of Children on Online Gaming Platforms. *Political Psychology*. https://doi.org/10.1111/pops.12855

Kowert, R., Domahidi E., & Quandt, T. (2014). The Relationship Between Online Video Game Involvement and Gaming-Related Friendships Among Emotionally Sensitive Individuals. *Cyberpsychology, Behavior, and Social Networking*: 447-453. http://doi.org/10.1089/cyber.2013.0656

Kowert, R., Kilmer, E., & Authors. (2023). *Toxic Gamers are Alienating Your Core Demographic*. Take This. https://www.takethis.org/wp-content/uploads/2023/08/ToxicGamersBottomLineReport_TakeThis.pdf

Kowert, R., Martel, A., & Swann, W. B. (2022). Not just a game: Identity fusion and extremism in gaming cultures. *Frontiers in Communication*, *7*. https://doi.org/10.3389/fcomm.2022.1007128

Lakhani, S. (2022). *Video Gaming and (Violent) Extremism: An exploration of the current landscape, trends, and threats*. Radicalisation Awareness Network (RAN). https://home-affairs.ec.europa.eu/system/files/2022-02/EUIF%20Technical%20Meeting%20on%20Video%20Gaming%20October%202021%20RAN%20Policy%20Support%20paper_en.pdf

Lakhani, S., White, J., & Wallner, C. (2022). *The Gamification of (Violent) Extremism: An exploration of emerging trends, future threat scenarios, and potential P/CVE solutions*. Radicalisation Awareness Network (RAN). https://home-affairs.ec.europa.eu/system/files/2022-09/RAN%20Policy%20Support-%20gamification%20of%20violent%20extremism_en.pdf

Lamphere-Englund, G., & Bunmathong, L. (2021). *State of Play: Reviewing the literature on gaming & extremism*. Extremism and Gaming Research Network (EGRN). https://lovefrankie.link/extremismandgamingreview

Lamphere-Englund, G., & Vugteveen, M. (2023). *Understanding Strategic Communications for P/CVE: Audience Segmentation and Message Testing Approaches*. Radicalisation Awareness Network (RAN). https://home-affairs.ec.europa.eu/system/files/2023-02/RANPS_Stratcomms_Consolidated_Audience%20segmentation_ICCT.pdf

Lamphere-Englund, Galen, and White, J. *"The Pentagon Leak: Childish Stunt or Dangerous Trend?"* Royal United Services Institute (RUSI) Commentary, 23 Apr 2023. https://www.rusi.org/explore-our-research/publications/commentary/pentagon-leak-childish-stunt-or-dangerous-trend?trk=public_post_comment-text.

Lamphere-Englund, G., & White, J. (2023). *The Online Gaming Ecosystem: Assessing Digital Socialisation, Extremism Risks and Harms Mitigation Efforts*. Extremism and Gaming Research Network (EGRN) and the Global Network on Extremism and Technology (GNET). https://gnet-research.org/wp-content/uploads/2023/05/GNET-37-Extremism-and-Gaming_web.pdf

Marsh, P., Fox, K., Carnibella, G., McCann, J. and Marsh, J. (1996) *Football Violence in Europe*. The Amsterdam Group. http://www.sirc.org/publik/football_violence.html.

Nationaal Coördinator Terrorismebestrijding en Veiligheid. (2018). Handreiking Alerteringssysteem Terrorismebestrijding (ATb) Publiek-Private Samenwerking bij een terroristische dreigingssituatie. https://www.nctv.nl/binaries/nctv/documenten/publicaties/2018/06/27/handreiking-atb/Handreiking+ATb+-+update+juni+2018.pdf

Newman, J. (2017). *Governing Public-Private Partnerships.* McGill-Queen's Press - MQUP. https://books.google.com/books/about/Governing_Public_Private_Partnerships.html?id=Pc8mDwAAQBAJ

Newton, C. (2019, December 16). *Google and YouTube moderators speak out on the work that gave them PTSD*. The Verge. https://www.theverge.com/2019/12/16/21021005/google-youtube-moderators-ptsd-accenture-violent-disturbing-content-interviews-video

Newzoo. (2022a). *The Games Market in 2022: The Year in Numbers*. Newzoo. https://newzoo.com/resources/blog/the-games-market-in-2022-the-year-in-numbers

Newzoo. (2022b, July 26). *Newzoo Global Games Market Report 2022*. Newzoo. https://newzoo.com/resources/trend-reports/newzoo-global-games-market-report-2022-free-version?utm_campaign=GGMR2022&utm_source=press

Radicalisation Awareness Network (RAN). (2020). *Extremists' Use of Video Gaming - and Narratives*. Radicalisation Awareness Network (RAN). https://home-affairs.ec.europa.eu/system/files/2020-11/ran_cn_conclusion_paper_videogames_15-17092020_en.pdf

Radicalisation Awareness Network (RAN). (2021). *Digital Grooming Tactics on Video Gaming & Video Gaming Adjacent Platforms: Threats and Opportunities*. https://home-affairs.ec.europa.eu/system/files/2021-05/ran_c-n_conclusion_paper_grooming_through_gaming_15-16032021_en.pdf

Roberts, S. T. (2019). *Behind the Screen.* Yale University Press.

Schlegel, L. (2021a). Connecting, Competing, and Trolling: "User Types" in Digital Gamified Radicalization Processes. *Perspectives on Terrorism*, *15*(4). https://www.universiteitleiden.nl/binaries/content/assets/customsites/perspectives-on-terrorism/2021/issue-4/schlegel.pdf

Schlegel, L. (2021b). *Extremists' use of gaming (adjacent) platforms: Insights regarding primary and secondary prevention measures*. Radicalization Awareness Network (RAN). https://home-affairs.ec.europa.eu/system/files/2021-08/ran_extremists_use_gaming_platforms_082021_en.pdf

Schlegel, L. (2021c). *The gamification of violent extremism & lessons for P/CVE*. Radicalisation Awareness Network (RAN). https://home-affairs.ec.europa.eu/system/files/2021-03/ran_ad-hoc_pap_gamification_20210215_en.pdf

Schlegel, L. (2022). *Why extremists are gaming and how P/CVE can leverage the positive effects of video games to prevent radicalization*. GameD. https://www.scenor.at/_files/ugd/ff9c7a_9f5f3687937b4f3384e2b0a7eac8c33f.pdf

Schlegel, L., & Amarasingam, A. (2022). *Examining the Intersection Between Gaming and Violent Extremism*. United Nations Office of Counter-Terrorism (UNOCT). https://www.un.org/counterterrorism/sites/www.un.org.counterterrorism/files/221005_research_launch_on_gaming_ve.pdf

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). Towards a Standard for Identifying and Managing Bias in Artificial Intelligence. *National Institute of Standards and Technology (NIST)*, *NIST Special Publication 1270*. https://doi.org/10.6028/nist.sp.1270

Steden, R. van & Meijer, R. (2016). *Public-Private Partnerships in times of diffuse threat: A study on the diversity of working methods and opportunities in the dutch and flemish context*. Verwey Jonker.

Trahan, L. (2023). *Summary of Responses from Gaming Companies*. U.S. House of Representatives. https://trahan.house.gov/uploadedfiles/summary_responses_to_letter_game_companies_online_harassment_extremism.pdf

White, J., & Lamphere-Englund, G. (2023). A View from the CT Foxhole: Jessica White and Galen Lamphere-Englund, Co-Conveners, Extremism and Gaming Research Network. *CTC Sentinel*, *16*(3). https://ctc.westpoint.edu/a-view-from-the-ct-foxhole-jessica-white-and-galen-lamphere-englund-co-conveners-extremism-and-gaming-research-network/

West, L.J. (2021). The impact of technology on extremism. In L. Close & D. Impiombato (Eds.), Counterterrorism Yearbook 2021 (pp. 29–32). Australian Strategic Policy Institute. http://www.jstor.org/stable/resrep31258.9

Zack, T., et al. (2024) Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. *The Lancet Digital Health*, Volume 6, Issue 1, e12 - e22. https://doi.org/10.1016/S2589-7500(23)00225-X