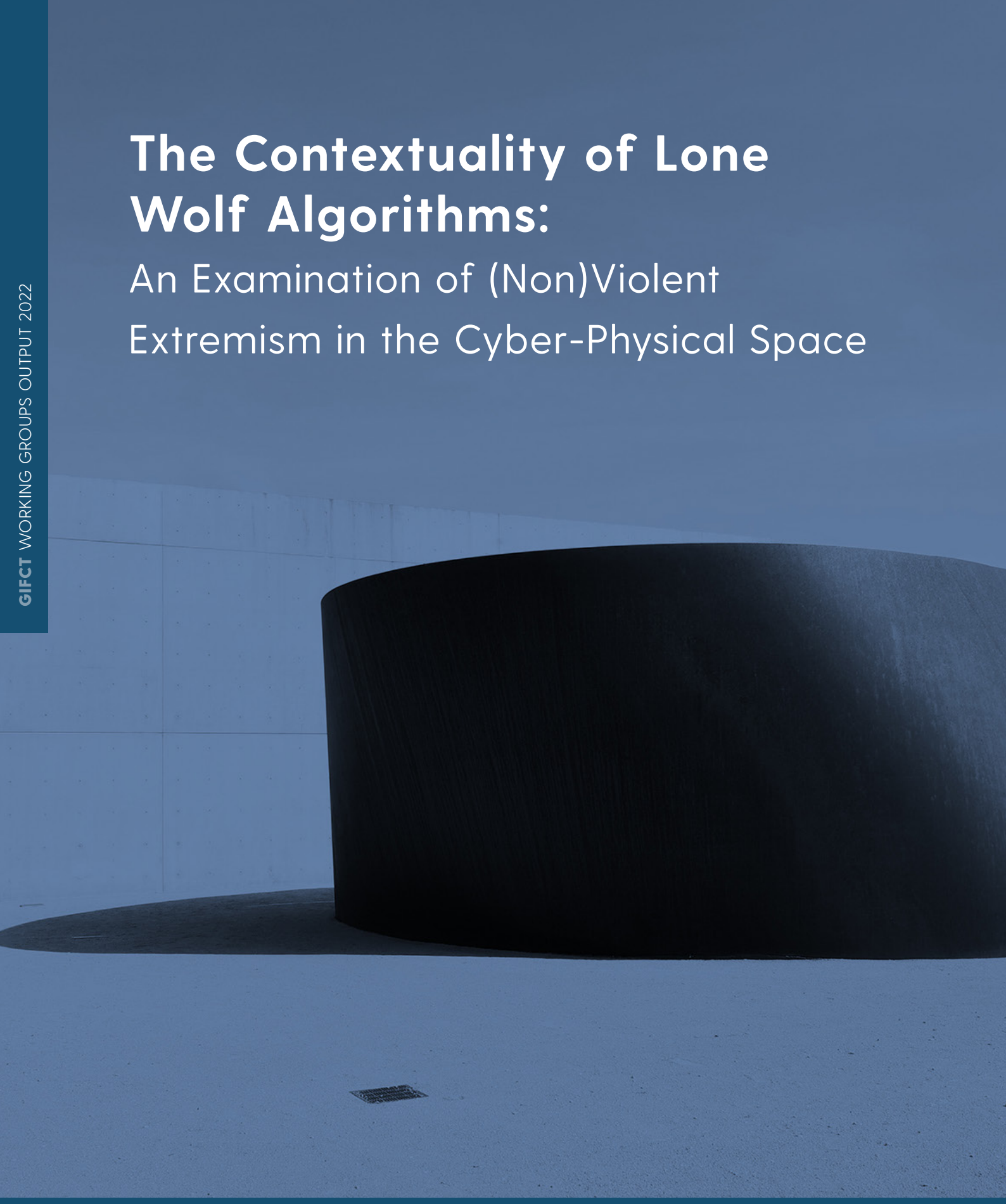
GIFCT WORKING GROUPS OUTPUT 2022

The Contextuality of Lone Wolf Algorithms:

An Examination of (Non)Violent Extremism in the Cyber-Physical Space



GIFCT
Global Internet Forum to Counter Terrorism

Dr. Jazz Yvonne Rowa

GIFCT WORKING GROUPS OUTPUT 2022

Introducing 2022 GIFCT Working Group Outputs

GIFCT
Global Internet Forum to Counter Terrorism

Dr. Erin Saltman
Director of Programming,
GIFCT

In July 2020, GIFCT launched a series of Working Groups to bring together experts from across sectors, geographies, and disciplines to offer advice in specific thematic areas and deliver on targeted, substantive projects to enhance and evolve counterterrorism and counter-extremism efforts online. Participation in Working Groups is voluntary and individuals or NGOs leading Working Group projects and outputs receive funding from GIFCT to help further their group's aims. Participants work with GIFCT to prepare strategic work plans, outline objectives, set goals, identify strategies, produce deliverables, and meet timelines. Working Group outputs are made public on the GIFCT website to benefit the widest community. Each year, after GIFCT's Annual Summit in July, groups are refreshed to update themes, focus areas, and participants.

From August 2021 to July 2022, GIFCT Working Groups focused on the following themes:

- Crisis Response & Incident Protocols
- Positive Interventions & Strategic Communications
- Technical Approaches: Tooling, Algorithms & Artificial Intelligence
- Transparency: Best Practices & Implementation
- Legal Frameworks

A total of 178 participants from 35 countries across six continents were picked to participate in this year's Working Groups. Applications to join groups are open to the public and participants are chosen based on ensuring each group is populated with subject matter experts from across different sectors and geographies, with a range of perspectives to address the topic. Working Group participants in 2021–2022 came from civil society (57%), national and international government bodies (26%), and technology companies (17%).

Participant diversity does not mean that everyone always agrees on approaches. In many cases, the aim is not to force group unanimity, but to find value in highlighting differences of opinion and develop empathy and greater understanding about the various ways that each sector identifies problems and looks to build solutions. At the end of the day, everyone involved in addressing violent extremist exploitation of digital platforms is working toward the same goal: countering terrorism while respecting human rights. The projects presented from this year's Working Groups highlight the many perspectives and approaches necessary to understand and effectively address the ever-evolving counterterrorism and violent extremism efforts in the online space. The following summarizes the thirteen outputs produced by the five Working Groups.

Crisis Response Working Group (CRWG):

The GIFCT Working Group on Crisis Response feeds directly into improving and refining GIFCT's own Incident Response Framework, as well as posing broader questions about the role of law enforcement, tech companies, and wider civil society groups during and in the aftermath of a terrorist or violent extremist attack. CRWG produced three outputs. The largest of the three was an immersive virtual series of Crisis Response Tabletop Exercises, hosted by GIFCT's Director of Technology, Tom Thorley. The aim of the Tabletops was to build on previous Europol and Christchurch Call-led Crisis Response events, with a focus on human rights, internal communications, and external strategic communications in and around crisis scenarios. To share lessons learned and areas for

improvement and refinement, a summary of these cross-sector immersive events is included in the 2022 collection of Working Group papers.

The second output from the CRWG is a paper on the Human Rights Lifecycle of a Terrorist Incident, led by Dr. Farzaneh Badii. This paper discusses how best GIFCT and relevant stakeholders can apply human rights indicators and parameters into crisis response work based on the 2021 GIFCT Human Rights Impact Assessment and UN frameworks. To help practitioners integrate a human rights approach, the output highlights which and whose human rights are impacted during a terrorist incident and the ramifications involved.

The final CRWG output is on Crisis Response Protocols: Mapping & Gap Analysis , led by the New Zealand government in coordination with the wider Christchurch Call to Action. The paper maps crisis response protocols of GIFCT and partnered governments and outlines the role of tech companies and civil society within those protocols. Overall, the output identifies and analyzes the gaps and overlaps of protocols, and provides a set of recommendations for moving forward.

Positive Interventions & Strategic Communications (PIWG):

The Positive Interventions and Strategic Communications Working Group developed two outputs to focus on advancing the prevention and counter-extremism activist space. The first is a paper led by Munir Zamir on Active Strategic Communications: Measuring Impact and Audience Engagement. This analysis highlights tactics and methodologies for turning passive content consumption of campaigns into active engagement online. The analysis tracks a variety of methodologies for yielding more impact-focused measurement and evaluation.

The second paper, led by Kesa White, is on Good Practices, Tools, and Safety Measures for Researchers. This paper discusses approaches and safeguarding mechanisms to ensure best practices online for online researchers and activists in the counterterrorism and counter-extremism sector. Recognizing that researchers and practitioners often put themselves or their target audiences at risk, the paper discusses do-no-harm principles and online tools for safety-by-design methodologies within personal, research, and practitioner online habits.

Technical Approaches Working Group (TAWG):

As the dialogue on algorithms and the nexus with violent extremism has increased in recent years, the Technical Approaches Working Group worked to produce a longer report on Methodologies to Evaluate Content Sharing Algorithms & Processes led by GIFCT's Director of Technology Tom Thorley in collaboration with Emma Llanso and Dr. Chris Meserole. While Year 1 of Working Groups produced a paper identifying the types of algorithms that pose major concerns to the CVE and counterterrorism sector, Year 2 output explores research questions at the intersection of algorithms, users and TVEC, the feasibility of various methodologies and the challenges and debates facing research in this area.

To further this technical work into Year 3, TAWG has worked with GIFCT to release a Research Call

for Proposals funded by GIFCT. This Call for Proposals is on Machine Translation. Specifically, it will allow third parties to develop tooling based on the gap analysis from last year's TAWG Gap Analysis. Specifically, it seeks to develop a multilingual machine learning system addressing violent extremist contexts.

Transparency Working Group (TWG):

The Transparency Working Group produced two outputs to guide and evolve the conversation about transparency in relation to practitioners, governments, and tech companies. The first output, led by Dr. Joe Whittaker, focuses on researcher transparency in analyzing algorithmic systems. The paper on Recommendation Algorithms and Extremist Content: A Review of Empirical Evidence reviews how researchers have attempted to analyze content-sharing algorithms and indicates suggested best practices for researchers in terms of framing, methodologies, and transparency. It also contains recommendations for sustainable and replicable research.

The second output, led by Dr. Courtney Radsch, reports on Transparency Reporting: Good Practices and Lessons from Global Assessment Frameworks. The paper highlights broader framing for the questions around transparency reporting, the needs of various sectors for transparency, and questions around what meaningful transparency looks like.

The Legal Frameworks Working Group (LFWG):

The Legal Frameworks Working Group produced two complementary outputs.

The first LFWG output is about Privacy and Data Protection/Access led by Dia Kayyali. This White Paper reviews the implications and applications of the EU's Digital Services Act (DSA) and the General Data Protection Regulation (GDPR). This includes case studies on Yemen and Ukraine, a data taxonomy, and legal research on the Stored Communications Act.

The second LFWG output focuses on terrorist definitions and compliments GIFCT's wider Definitional Frameworks and Principles work. This output, led by Dr. Katy Vaughan, is on The Interoperability of Terrorism Definitions. This paper focuses on the interoperability, consistency, and coherence of terrorism definitions across a number of countries, international organizations, and tech platforms. Notably, it highlights legal issues around defining terrorism based largely on government lists and how they are applied online.

Research on Algorithmic Amplification:

Finally, due to the increased concern from governments and human rights networks about the potential link between algorithmic amplification and violent extremist radicalization, GIFCT commissioned Dr. Jazz Rowa to sit across three of GIFCT's Working Groups to develop an extensive paper providing an analytical framework through the lens of human security to better understand the relation between algorithms and processes of radicalization. Dr. Rowa participated in the Transparency, Technical Approaches, and Legal Frameworks Working Groups to gain insight into

the real and perceived threat from algorithmic amplification. This research looks at the contextuality of algorithms, the current public policy environment, and human rights as a cross-cutting issue. In reviewing technical and human processes, she also looks at the potential agency played by algorithms, governments, users, and platforms more broadly to better understand causality.

We at GIFCT hope that these fourteen outputs are of utility to the widest range of international stakeholders possible. While we are an organization that was founded by technology companies to aid the wider tech landscape in preventing terrorist and violent extremist exploitation online, we believe it is only through this multistakeholder approach that we can yield meaningful and long-lasting progress against a constantly evolving adversarial threat.

We look forward to the refreshed Working Groups commencing in September 2022 and remain grateful for all the time and energy given to these efforts by our Working Group participants.

Participant Affiliations in the August 2021 - July 2022 Working Groups:

Tech Sector	Government Sector	Civil Society / Academia / Practitioners	Civil Society / Academia / Practitioners
ActiveFence	Aqaba Process	Access Now	Lowy Institute
Amazon	Association Rwandaise de Défense des Droits de l'Homme	Anti-Defamation League (ADL)	M&C Saatchi World Services Partner
Automattic	Australian Government - Department of Home Affairs	American University	Mnemonic
Checkstep Ltd.	BMI Germany	ARTICLE 19	Moonshot
Dailymotion	Canadian Government	Australian Muslim Advocacy Network (AMAN)	Moduslzad - Centre for applied research on deradicalisation
Discord	Classification Office, New Zealand	Biodiversity Hub International	New America's Open Technology Institute
Dropbox, Inc.	Commonwealth Secretariat	Bonding Beyond Borders	Oxford Internet Institute
ExTrac	Council of Europe, Committee on Counter-Terrorism	Brookings Institution	Partnership for Countering Influence Operations, Carnegie Endowment for International Peace
Facebook	Department of Justice - Ireland	Business for Social Responsibility	Peace Research Institute Frankfurt (PRIF); Germany
JustPaste.it	Department of State - Ireland	Centre for Analysis of the Radical Right (CARR)	PeaceGeeks
Mailchimp	Department of State - USA	Center for Democracy & Technology	Point72.com
MEGA	Department of the Prime Minister and Cabinet (DPMC), New Zealand Government	Center for Media, Data and Society	Polarization and Extremism Research and Innovation Lab (PERIL)
Microsoft	DHS Center for Prevention Programs and Partnerships (CP3)	Centre for Human Rights	Policy Center for the New South (senior fellow)
Pex	European Commission	Centre for International Governance Innovation	Public Safety Canada & Carleton University
Snap Inc.	Europol/EU IRU	Centre for Youth and Criminal Justice (CYCJ) at the University of Strathclyde, Scotland.	Queen's University
Tik Tok	Federal Bureau of Investigation (FBI)	Cognitive Security Information Sharing & Analysis Center	Sada Award, Athar NGO, International Youth Foundation
Tremau	HRH Prince Ghazi Bin Muhammad's Office	Cornell University	Shout Out UK
Twitter	Ministry of Culture, DGMIC - France	CyberPeace Institute	Strategic News Global
You Tube	Ministry of Foreign Affairs - France	Dare to be Grey	S. Rajaratnam School of International Studies, Singapore (RSIS)
	Ministry of Home Affairs (MHA) - Indian Government	Dept of Computer Science, University of Otago	Swansea University
	Ministry of Justice and Security, the Netherlands	Digital Medusa	Tech Against Terrorism
	National Counter Terrorism Authority (NACTA) Pakistan	Edinburgh Law School, The University of Edinburgh	The Alan Turing Institute

	Organisation for Economic Co-operation and Development (OECD)	European Center for Not-for-Profit Law (ECNL)	The Electronic Frontier Foundation
	Office of the Australian eSafety Commissioner (eSafety)	Gillberg Neuropsychiatry Centre, Gothenburg University, Sweden,	The National Consortium for the Study of Terrorism and Responses to Terrorism (START) / University of Maryland
	Organization for Security and Co-operation in Europe (OSCE RFoM)	George Washington University, Program on Extremism	Unity is Strength
	Pôle d'Expertise de la Régulation Numérique (French Government)	Georgetown University	Université de Bretagne occidentale (France)
	North Atlantic Treaty Organization, also called the North Atlantic Alliance (NATO)	Georgia State University	University of Auckland
	Secrétaire général du Comité Interministériel de prévention de la délinquance et de la radicalisation	Global Network on Extremism and Technology (GNET)	University of Groningen
	State Security Service of Georgia	Global Disinformation Index	University of Massachusetts Lowell
	The Royal Hashemite Court/ Jordanian Government	Global Network Initiative (GNI)	University of Oxford
	The Office of Communications (Ofcom), UK	Global Partners Digital	University of Queensland
	UK Home Office	Global Project Against Hate and Extremism	University of Salford, Manchester, England,
	United Nations Counter-terrorism Committee Executive Directorate (CTED)	Groundscout/Resonant Voices Initiative	University of South Wales
	UN, Analytical Support and Sanctions Monitoring Team (1267 Monitoring Team)	Hedayah	University of the West of Scotland
	United Nations Major Group for Children and Youth (UNMGCY)	Human Cognition	Violence Prevention Network
	United States Agency for International Development (USAID)	Institute for Strategic Dialogue	WeCan Africa Initiative & Inspire Africa For Global Impact
		International Centre for Counter-Terrorism	Wikimedia Foundation
		Internet Governance Project, Georgia Institute of Technology	World Jewish Congress
		Islamic Women's Council of New Zealand	XCyber Group
		JOS Project	Yale University, Jackson Institute
		JustPeace Labs	Zinc Network
		Khalifa Ihler Institute	
		KizBasina (Just-a-Girl)	
		Love Frankie	

GIFCT WORKING GROUPS OUTPUT 2022

GIFCT Executive Summary and Discussion of Dr. Jazz Rowa's Algorithms Research

GIFCT
Global Internet Forum to Counter Terrorism

Dr. Erin Saltman
Director of Programming,
GIFCT

GIFCT recognizes the increasing concern from governments, researchers, technologists, and human rights advocates about the potential link between algorithmic amplification and processes of radicalization towards violence. Increased legislative language around the world has turned to 'algorithmic transparency' and one of the primary themes of the Christchurch Call to Action's Second Anniversary Summit in 2021 was to support methods to better understand user journeys online and the role algorithms may play in processes of radicalization. There is a fear that the nature of online environments may amplify hatred and glorify terrorism and violent extremism in a way that drives others towards violence. To effectively counter terrorism and violent extremism online, GIFCT aims to support research, analysis, and tools to better understand the true nature of the problem so that action can be taken. On the topic of understanding algorithmic processes there remain large knowledge gaps. GIFCT commissioned an extensive research effort by Dr. Jazz Rowa to assist in framing and better understanding the role of algorithms as part of GIFCT's 2022 Working Group outputs. This executive summary of her longer research paper, The Contextuality of Algorithms: An Examination of (Non)Violent Extremism in the Cyber-Physical Space, serves as a briefing document and reflection from GIFCT about some of Dr. Rowa's key findings. As of September 2022, her longer report can also be found on the GIFCT website under Working Group output and under our highlighted resources.

Background

In the first year of GIFCT Working Groups, held September 2020 through July 2021, GIFCT convened a group of global experts focussed on Content-Sharing Algorithms, Processes, and Positive Interventions, with participants from across tech companies, government, and civil society. Since an algorithm can be almost any input online with an output, the group adopted the shared goal of mapping which content-sharing algorithms and processes used by industry had the potential of facilitating consumption of content that may amplify terrorist and violent extremist content, or user interest in such content. The group also mapped and considered positive interventions and risk mitigation points for safety-by-design. The results of this paper honed in on the algorithmically optimized surfaces and tools that could potentially be exploited by bad actors, such as terrorists or violent extremists. This allowed the conversation on algorithms to focus more specifically on three online surfaces: search functions, recommendation features, and ad targeting algorithms.

In Year 2 of Working Groups, held September 2021 through July 2022, GIFCT commissioned Dr. Jazz Rowa to take this conversation and analysis further. GIFCT Working Groups had sub questions related to algorithms and the nexus with extremism in 3 of our 5 groups and asked Dr. Rowa to sit across these groups to develop this extensive paper. She has provided an analytical framework through the lens of human security to better understand the relation between algorithms and processes of radicalization. Dr. Rowa participated in the Transparency, Technical Approaches, and Legal Frameworks Working Groups to gain insight into the real and perceived threat from algorithmic amplification. This participation was supplemented with empirical research and a range of first-person interviews. This research looks at the contextuality of algorithms, the current public policy environment, and human rights as a cross-cutting issue. In reviewing technical and human processes, she also looks at the potential agency played by algorithms, governments, users, and platforms more broadly to better understand causality.

Findings

While this paper presents a myriad of findings and poses further questions, identifying gaps for further research, there are some key takeaways that stuck out to our teams at GIFCT, which we will be processing and looking to build further work around in the future. The first takes us back to the age-old questions of definitions. In group discussions and interviews it remains clear that **there is no overarching agreement between different sectors or geographies on what online terrorist content is, what violent extremism is, what algorithms are, and what "extremist" or "borderline" content is.** If it can't be well defined, or if legislative language is vague on these points, we are still left with too much ambiguity to apply technical solutions or to ensure rigorous oversight or accountability mechanisms. Specifically for online spaces, the better you can define harm parameters the more you can measure, evaluate and risk mitigate. Vague or ambiguous terminology can lead to over censorship, under censorship, or the inability to measure and understand the nature of the problem in the first place.

While pressure escalates for tech companies to "do more", the analysis notes that **the current guidance on human rights in national, regional, and international legal frameworks is technologically suboptimal.** The pressure to expand technical solution-building is not equally matched with practical guidance of what human rights applications for technological ecosystems should look like. The paper also found that even some government representatives were wary that the term "algorithm" had become the latest buzzword and hot topic in the international debates on preventing and countering terrorism and violent extremism online, without enough clarity on the concept or the scope.

Dr. Rowa addresses the multiple reasons **why understanding algorithms, and attempts to provide meaningful algorithmic transparency, remains difficult. There is a notable difference between algorithmic explicability, interpretability, and auditability.** However, approaching algorithmic systems and its "black box" effect for analyzing input and output variables is compounded for a number of reasons; very few people understand the technical side of digital technologies, there remains a system of self-regulation for the technical evolution and review of technologies, there are methodological limitations for external researchers reviewing algorithmic systems, all combined with a trend of reactionary government regulation. The disclosure of an algorithmic formula or source code is viewed by some as useful and many as irrelevant in understanding a program's predictive behavior. Meanwhile there is a multi-dimensional and ever-changing landscape for both terrorist and violent extremist actors online and technical dynamism of platforms themselves. This conceptualisation of audits and the design of mechanisms for algorithmic oversight must therefore acknowledge the complexity of such an undertaking. To work towards greater algorithmic transparency, more work will need to be done to fully understand what "meaningful" data and algorithmic transparency means to policy makers and relevant stakeholders. Data and information sharing from tech companies can take many forms and alignment on understanding what data is useful and meaningful is crucial.

The current discourse on the role of algorithms in (non)violent extremism has for the most part

created a false dichotomy between the online and offline spaces. The discussion around user, platform, and government furthers the complexity in trying to interpret causality in processes of radicalization and agency. User agency and lived experiences particularize contextual phenomena and inform the integration of the online and offline dynamics of extremism. Dr. Rowa points out that the interplay between the user and how an algorithm operates is intrinsically tied. Algorithmic systems are representations of human decisions and worldviews. What happens in the online realm cannot be detached from real life actions. This interplay needs to also inform legislative thinking.

Related to the discourse around user and platform accountability and responsibility, the interviews highlighted the **continued discomfort with non-violent and non-violating extremist content in what might be determined "gray area" content, and what, if anything, tech companies should do about it.** If users create legal, non-violating content and other users actively search and engage with the content, should private technology companies exert absolute control over the curation and restriction of legal but 'extreme' content? The concerns over borderline content are tied to the overarching debate on the definition of extremist content, liability for content creation, and the dispersal of content across digital publics (within hybridized or algorithmically amplified systems).

While some algorithm/user interplay could potentially amplify extremist content, there remain many spaces online that are beacons for violent extremist and terrorist sympathizers, yet have no algorithmic optimization associated with content surfacing or group recommendation features. These platforms remain a beacon to hate-based groups simply because they lack proactive moderation of content. The analysis notes that the recent lone actor attack in Buffalo, New York is seen as a case of "radicalization on 4chan" by other users giving social constructive information, documents, and social feedback. The attacker was also previously known to police, meaning there were offline signals that could have been used to provide support or have led to PVE/CVE interventions.

The overall research creates many avenues for further dialogues and multistakeholder work. However, it is important to recognize where positive opportunities for future work lie. **The research concludes that algorithmic processes, while being the core scrutiny of this paper, are equally where solutions can be found.** Despite the initial research question for the paper, Dr. Rowa points out that, paradoxically, algorithmic systems are conceived as automated problem solvers. In concert with other agencies, algorithms can act as conduits for the reconciliation, remediation, and reconstitution of an increasingly dysfunctional cyber-physical order. Whereas algorithms pose (un) known challenges for extremism, the opportunities they present in the mitigation and resolution of this and other societal challenges is equally consequential.

We at GIFCT hope that this research is of utility to the broadest range of stakeholders working to counter terrorism and violent extremism online and are grateful to Dr. Jazz Rowa for the time and energy she put into this extensive research over the last year.

Dr. Erin Saltman
Director of Programming
GIFCT

Acknowledgments

I would like to thank the Global Internet Forum to Counter Terrorism (GIFCT) for supporting this research. My special gratitude goes to Dr. Erin Saltman and Dr. Nayanka Perdigao who supported the original research idea, proposed the integration of three GIFCT working groups (Technical Approaches, Transparency, and Legal Frameworks) into the research design, and provided other forms of assistance. I really appreciate Mr. Tom Thorley's patience, insights, and overall support throughout the research. I am extremely grateful to Dr. David Matthews for granting me sufficient time to meet the research deadline, and for his wise counsel during a very challenging period. My deep gratitude goes to all interviewees whose insights greatly enriched the study. I also appreciate the useful feedback from Dr. Nagham El Karhili, Ms. Sarah Pollack, and the rest of the GIFCT team. I would also like to thank Mr. Jacob Lindelow Berntsson for his responsiveness and resourcefulness during the research period. Finally, a big thanks to Dr. Joe Whittaker for reviewing the initial research questions and Mr. Scott Johnson for an interesting exchange at the start of the study.

Table of Contents

RESEARCH BACKGROUND

Section 1: Introduction

1.1 A Snapshot of the Study

This research is an offshoot of two articles previously published on the Global Network on Extremism and Technology (GNET) platform that focused on algorithmic agency in online extremism.[1] This particular study examines the role of algorithms in extremism, and in particular how they operate in context. The research posits that a singular focus on algorithms limits effective response to the complex and multifaceted problem of "online" extremism, of which algorithms are a part. Consequently, the study positions the agency, protection, and empowerment of people at the center of inquiry. An examination of how algorithms operate at the micro-context and the interplay with user agency and broader social contexts can therefore enhance understanding of the role of algorithms in extremism. It is within this milieu that the research interrogates the popular concept of "algorithmic amplification" and dissects the artificial dichotomy between online and offline extremism. The analysis of algorithmic, user, and contextual interactivities, though exploratory at this stage, similarly aims to increase understanding of the intersections of online-offline extremism. Overall, the study seeks to broaden the discourse beyond assertions that are largely or purely structured around algorithms, and inform the design of more holistic, integrated, and sustainable interventions.

1.2 The Paradox of Lone Wolf Algorithms

The role of algorithms in online extremism has garnered both support and opposition in popular and academic discourse. The existing scholarship is punctuated by a dearth of theoretical and empirical studies on the contextuality of algorithms. Moreover, the skewed focus on the visible manifestations of social media harms has diverted attention from the systemic and interconnected drivers of online-offline extremism. This research underscores the centrality of context in the ongoing debate and argues that the visible manifestations of online (non)violent extremism are rooted in the interplay between dynamic contextual factors and human-algorithmic relational agency. As such, a better understanding of extremism on social media demands a deconstruction of the broader ecosystem and the examination of the relationship between its constitutive elements.

While there is a paucity of academic literature examining the role of algorithms in extremism, critical

1 Yvonne Jazz Rowa, "Part 1: Algorithmic Deconstruction in the Context of Online Extremism," GNET (blog), September 15, 2020, https://gnet-research.org/2020/09/15/part-1-algorithmic-deconstruction-in-the-context-of-online-extremism/; Yvonne Jazz Rowa, "Part 2: Algorithmic Agency in Online Extremism: The Bigger Picture," GNET (blog), September 21, 2021, https://gnet-research.org/2020/09/21/part-2-algorithmic-agency-in-online-extremism-the-bigger-picture/.

algorithm studies have begun to produce some seminal works in this area.[2] So far, there are several lacunae in the existing body of work on algorithmic extremism that this study aims to address.[3] The plethora of research examining extremism on social media has been particularly instructive on the problem but has for the most part addressed the visible manifestations of "online" extremism.[4] The compelling array of studies that delve deeper into the architecture of algorithms and related structures is increasingly narrowing that gap.[5] Nevertheless, emerging from this work are algorithm-centric analyses that typically overlook or diminish the corresponding interactive human agency. Although some scholarship has examined algorithmic agency, the focus on singular or discrete contextual factors has significantly engineered the groundwork for further inquiry. Of the studies that address contextual factors, none has so far formulated an integrated or comprehensive framework on the contextuality of algorithms.[6]

That said, three studies with a specific focus on algorithmic contextuality stand out. Munger and Phillips have examined and proposed a supply and demand framework for analyzing politics on YouTube. Their study identifies societal parameters to assess the interplay between politics and the platform's technological affordances.[7] While the context addressed is predominantly technopolitical, underlining the agency of the audience on the demand side, the study briefly describes but does not comprehensively address the role of algorithms in context. Valentini, Lorusso, and Stephan similarly scrutinize some contextual issues and make a compelling case for the fusion of digital and physical spaces in their examination of the role of algorithms in radicalization.[8] Overall, the paper provides a basis for novel approaches to the subject matter. While it represents an informative case study on the Islamic State, the paper does not comprehensively particularize the interplay between specific contexts and algorithms. This research goes further to distill specific constitutive elements within the domain of what has been conceptualized as the cyber-physical space.

Finally, Magalhães advances a tripartite framework for algorithmic ethical subjectivation that

2 Tarleton Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media (New Haven, CT: Yale University Press, 2018); Taina Bucher, If...Then: Algorithmic Power and Politics, Oxford Studies in Digital Politics (New York: Oxford University Press, 2018), https://doi.org/10.1093/oso/9780190493028.001.0001; Davide Panagia, "The Algorithm Dispositif (Notes towards an Investigation)," UCLA School of Law: Program on Understanding Law, Science, & Evidence, 2019, https://escholarship.org/uc/item/154618gr.

3 Derek O'Callaghan et al., "Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems," Social Science Computer Review 33, no. 4 (2015): 459–78; Josephine B. Schmitt et al., "Counter-Messages as Prevention or Promotion of Extremism?!: The Potential Role of Youtube Recommendation Algorithms," Journal of Communication 68, no. 4 (2018): 780–808; Mark Ledwich and Anna Zaitsev, "Algorithmic Extremism: Examining YouTube's Rabbit Hole of Radicalization," ArXiv, 2019, http://arxiv.org/abs/1912.11211; Tiana Gaudette et al., "Upvoting Extremism: Collective Identity Formation and the Extreme Right on Reddit," New Media & Society 23, no. 12 (2020): 3491–3508.

4 Robin Thompson, "Radicalization and the Use of Social Media," Journal of Strategic Security 4, no. 4 (2011): 167–90; David C. Benson, "Why the Internet Is Not Increasing Terrorism," Security Studies 23, no. 2 (April 3, 2014): 293–328, https://doi.org/10.1080/09636412.2014.905353; Cristina Archetti, "Terrorism, Communication and New Media: Explaining Radicalization in the Digital Age," Perspectives on Terrorism 9, no. 1 (2015): 49–59.

5 Elena Esposito, "Artificial Communication? The Production of Contingency by Algorithms," Zeitschrift Für Soziologie 46, no. 4 (2017): 249–65; Ferenc Huszár et al., "Algorithmic Amplification of Politics on Twitter," Proceedings of the National Academy of Sciences 119, no. 1 (2022): e2025334119; Kevin Munger and Joseph Phillips, "A Supply and Demand Framework for YouTube Politics," (unpublished paper, 2019), https://Osf. Io/73jys; João Carlos Magalhães, "Do Algorithms Shape Character? Considering Algorithmic Ethical Subjectivation," Social Media+Society 4, no. 2 (2018): 1–10.

6 Magalhães, "Do Algorithms Shape Character?"; Esposito, "Artificial Communication?"; Munger and Phillips, "Supply and Demand"; Daniele Valentini, Anna Maria Lorusso, and Achim Stephan, "Onlife Extremism: Dynamic Integration of Digital and Physical Spaces in Radicalization," Frontiers in Psychology 11 (2020): 524.

7 Munger and Phillips, "Supply and Demand."

8 Valentini, Lorusso, and Stephan, "Onlife Extremism."

includes plurality, contextuality, and potential harmfulness. He conceptualizes contextuality as enabling conditions, which while not dependent on end-users to subsist facilitate algorithmic processes. The study addresses three domains that encapsulate epistemological, ethical, and socio-material contexts. In particular, socio-materiality constitutes the decisions that algorithms make on behalf of end-users as subjects transform themselves in response to their speculations about the logic behind algorithmic decisions.[9] This research expands these margins to encapsulate other macro-contextual dynamics of extremism that have been under-researched in critical algorithm studies. A good starting point, therefore, is the definition of extremism as applied in this study. Whereas there is no universally accepted definition of extremism, the concept essentially describes:

> Religious, social, or political belief systems that exist substantially outside of belief systems more broadly accepted in society (i.e., "mainstream" beliefs). Extreme ideologies often seek radical changes in the nature of government, religion, or society. Extremism can also be used to refer to the radical wings of broader movements, such as the anti-abortion movement or the environmental movement. Not every extremist movement is "bad"... but most extremist movements exist outside of the mainstream because many of their views or tactics are objectionable. The difference between violent and nonviolent extremism is not often clear but the main distinctive feature is the act of causing physical harm. Violent extremism is defined as "a belief system that advocates the use of violence for the furtherance of an ideological cause."[10]

In considering the foregoing analysis, this study argues that online extremism is in part symptomatic of the interaction between latent algorithmic design and capabilities, human agency, and broader societal structures. The online-offline interface within which these influences coalesce represents a hybridized domain of extremism that this study conceives as "the cyber-physical space."[11] Though cyber-physical space is the dominant concept in this study, the term is applied interchangeably with "phygital"[12] and "online-offline" as deemed fit. That said, an understanding of Cyber-Physical Systems (CPS) foregrounds the understanding of the cyber-physical space. CPS are "integrations of computation, networking, and physical processes. Embedded computers and networks monitor and control the physical processes, with feedback loops where physical processes affect computations and vice versa."[13]

9 Magalhães, "Do Algorithms Shape Character?"

10 Anti Defamation League, "Defining Extremism: A Glossary of White Supremacist Terms, Movements and Philosophies | ADL," n.d., https://www.adl.org/resources/glossary-term/defining-extremism-glossary-white-supremacist-terms-movements-and. Jason-Leigh Striegher, "Violent-Extremism: An Examination of a Definitional Dilemma," in 8th Australian Security and Intelligence Conference, Held from the 30 November – 2 December, 2015 (Edith Cowan University Joondalup Campus, Perth, Australia: SRI Security Research Institute, 2015), 75–86, https://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1046&context=asi.

11 At the time of writing, it was discovered that cyber-physical system (CPS) is indeed an existing relevant concept whose description closely relates with the notion of cyber-physical space.

12 Phygital is a term used in marketing to denote the integration of experiences in the physical and digital domains. See Mike Welsh, "The Future is Phygital: Physical and Digital," Mobiquity, April 19, 2021, https://www.mobiquity.com/insights/the-future-is-phygital.

13 "Cyber-Physical Systems - a Concept Map," Berkeley CPS Publications, n.d., https://ptolemy.berkeley.edu/projects/cps/; see also "What Is the Difference Between CPS and IoT?," Vanderbilt School of Engineering, February 28, 2022, https://blog.engineering.vanderbilt.edu/what-is-the-difference-between-cps-and-iot.

Central to the functionality of CPS is the intertwinement and coordination between the physical and computational or software components. Moreover, CPS are often viewed as distinct from but complementary to the Internet of Things.[14] Cyber-physical space as applied in this research therefore denotes the online-offline interface. It is an environment in which computing devices and programs interact with human agency and the social, political, economic, cultural, and other contextual conditions in the (im)material world. This is a space that is perceptible and semi-tactile. Subsequently, this research is premised on the communicative capabilities of algorithms and submits that "the problem is not that the machine is able to think but that it is able to communicate. The reference to communication and social context is the central issue."[15] Overall, algorithms that are seemingly autonomous agents are for the most part responsive, reciprocal, and adaptive to prevailing contexts and interactive human agency.

1.3 Aims of the Study and Research Questions

This research aims to increase understanding of the agency and contextuality of algorithms in extremist cyber-physical spaces. Ultimately, the study aims to develop recommendations on algorithmic governance and inform the design of more holistic, integrated, and sustainable interventions. While the research does not generate all the answers to existing questions, it highlights current gaps, complexities, and dilemmas, and sets the context for prospective subject-specific inquiry. In interrogating the autonomy of algorithms in "online" extremism, the research sets out to answer the following questions:

1. What role do algorithms play in extremism in an integrated (or in the integration of) cyber-physical space? (How does the interaction between algorithms, user agency, and other contextual factors in the cyber-physical space foment extremism?)
2. What is understood, misunderstood, and not yet understood about the role of algorithms in extremism? (What are the implications of scholarship on the role of algorithms in extremism for human rights, transparency, policy, and other domains?)
3. What factors should policymakers and practitioners consider in algorithmic governance and the design and operationalization of interventions to tackle extremism in hybrid cyber-physical spaces?

1.4 Methodology

The human security (HS) approach is the overarching analytical framework for this research. In as much as the model is comprehensive, people-centered, and encourages an integrated response to complex issues, the oversight of technology is noteworthy. The research therefore contributes to the development of the HS framework. The research is also embedded in three GIFCT working groups, spanning the Technical Approaches, Transparency, and Legal Frameworks workstreams. The analysis and findings are in part based on the researcher's participation and observations in the three assemblies. Within this ambit, the study further examines the technical aspects of algorithms around the interlocking themes of

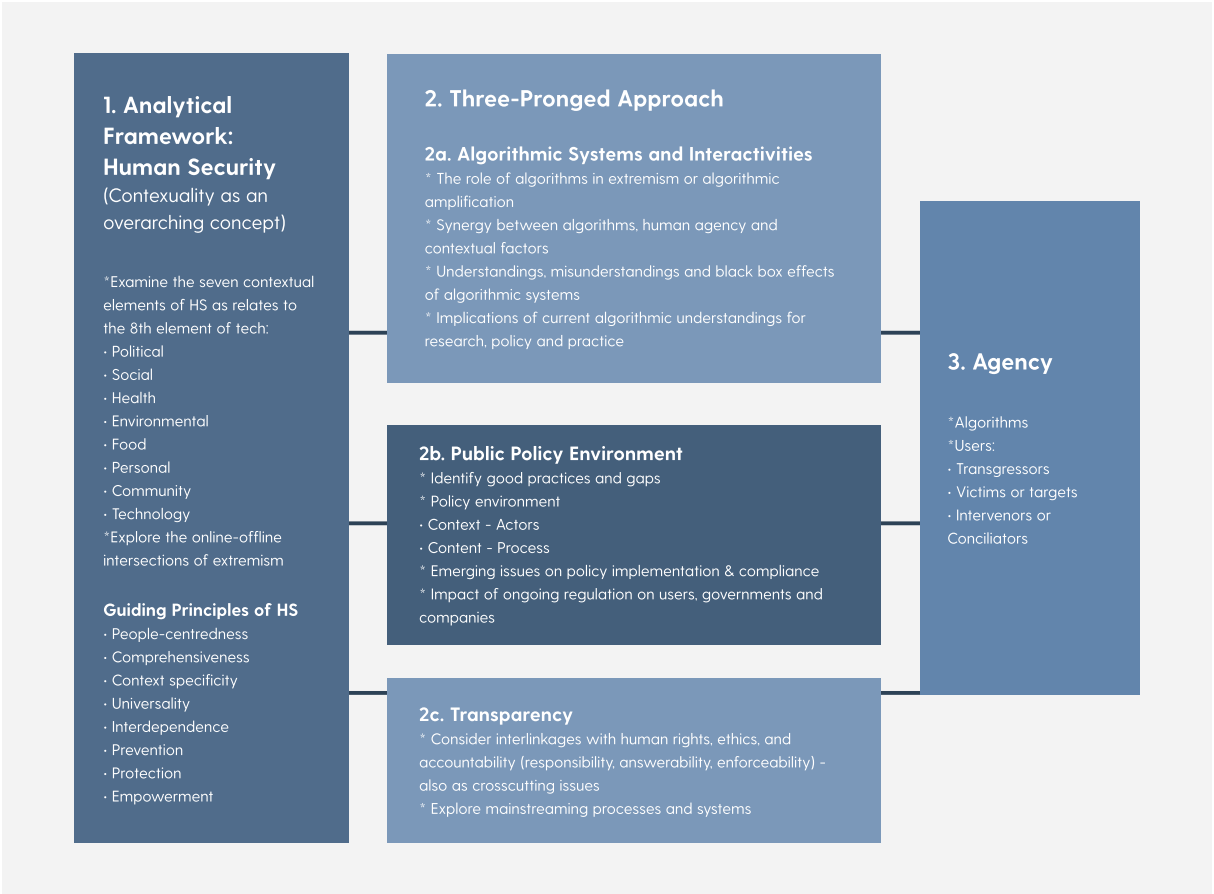
14 "What Is the Difference," Vanderbilt School of Engineering.

15 Esposito, "Artificial Communication?"

human rights and ethics. The three core areas of inquiry, represented by the trio of closely intertwined working groups, are examined under the overarching concept of contextuality. More specifically, the three-pronged approach addresses:

1. **Algorithmic systems and interactivities:** This component examines the technical properties of algorithms in context. It explores the synergy between algorithms, human agency, and other contextual factors, and casts a spotlight on what is known, misunderstood, and not yet understood. This strand additionally highlights the dilemmas and complexities that potentially limit understanding of the role of algorithms in extremism and corresponding interventions.
2. **The public policy environment:** This dimension explores the good practices and gaps in the existing tech legislation processes while establishing relevant linkages to algorithms. It examines the implications of tech legislation for industry and digital publics and the impact of algorithmic complexities on legislation.
3. **Transparency:** The research examines human rights as a crosscutting issue and draws attention to the debates around transparency and ethics in relation to human rights. Against this backdrop, the study explores institutional mainstreaming processes including the strengths and gaps in current practices.

The nature and structure of agency cuts across the three main areas of inquiry above. The study has dissected the constitution and constitutiveness of agency in an attempt to highlight the importance and relationship between actors and algorithms in online-offline extremism. In so doing, the research implicates piecemeal and algorithm-centric approaches in the production of fragmented and unsustainable results.



This study employed qualitative research methods with four streams of data sources. The first stream involved the critical review and synthesis of relevant secondary data sources. The second stream consists of primary data collected through semi-structured interviews with a total of 21 key informants. The interviews were conducted with multilateral and government officials, researchers, practitioners, technology company representatives, and legal experts. The third stream employed elements of cyber ethnography through an iterative process of observation, contextualization of online data, and digital reflexivity. The Buffalo terrorist attack and the researcher's online experiences in particular generated insights on user behavior, neglected channels in the online extremist ecosystem such as audio streaming platforms, and more broadly discursive patterns and other online dynamics that terrorist attacks induce in the online space. The final stream of research constitutes the analysis of exploratory case studies and anecdotes that complement other data sources and advance the overarching arguments of this study.

1.5 Justification, Expected Application of Results, and Limitations of the Study

In terms of practical application, the research provides policymakers and analysts with a formative framework for the analysis and design of policies and programs that address extremism in the cyber-physical space. Overall, the study advances a holistic, integrated, and sustainable approach to policymaking and interventions on extremism. Whereas the research exposes the shortcomings of a purist approach to countering "online" extremism, it is not without its own limitations. To begin with, the research acknowledges the hybridity of online threats but narrows the scope to extremism. It does not address other harms that may intersect with extremism or terrorism except when necessary. In addition, all the interviewees were open, enthusiastic, and generous with their perspectives. However, one actor who did not participate in the study expressed concern over the study's "deviation" from algorithmic systems and by extension, the circumvention of "online" extremism. Beyond face value, these concerns unmasked some undertones of frustration and distrust between technology companies and governments that should be acknowledged and addressed. The findings suggest that these dynamics not only complicate the research terrain but also impede effective government-tech collaboration.

The study did not collect the insights of "regular" internet users. This represents a gap in the analysis of findings that could have been enriched with more diverse perspectives. That being the case, the research relied on the contributions of other respondents who are essentially internet users. In some instances, due to the complexities presented in the findings, the study does not advance straightforward or conclusive recommendations but instead distills the dilemmas and complications around analytical and operational postures, including tensions and divergences in opinions and scholarship. An important constraint that underpins these challenges is the limited duration of the research. In the long run, acknowledging the complexities of extremism in the cyber-physical space and the elusiveness of compact solutions is perhaps the first step towards transforming and expanding the thinking and approaches to the algorithm question.

Section 2: Analytical Framework

2.1 A Human Security Approach: Setting the Context

The HS approach constitutes the overarching analytical framework for the study. HS is a dynamic, multi-dimensional analytical framework that guides the assessment, development, and implementation of integrated responses to complex issues requiring multi-sectoral and multistakeholder interventions that enhance policy coherence and greater impact. The framework's core vision to achieve freedom from fear, want, and indignity aims to address the challenges that emanate from and result in chronic conflicts and other concerns. The mutually reinforcing principles underpinning the approach include people-centredness, comprehensiveness, context specificity, prevention, and protection and empowerment.[16] An earlier iteration of the model additionally highlights the merits of universality and interdependence.[17] The framework's appreciation of the interconnectedness and complexity of root causes of insecurity at local, national, regional, and international levels fosters greater attention to contextualized analyses and action. The focus on prevention minimizes the impacts of threats while the protection and empowerment of people engender greater resilience and sustainability. A comprehensive and integrated approach facilitates the appraisal of existing institutional policies and practices, and ultimately establishes the relevant parameters for impact assessment and evaluation.[18]

The concept of HS challenges traditional state-centric (military) notions of security that centralize the agents and instruments of the state. It instead positions individuals as the dominant security referents and further concerns itself with both the protection and empowerment of people.[19] The approach recognizes inherent responsibilities within societies, places them at the center of action and analysis, and empowers people to respond to their needs. That said, the most practical applications of this paradigm recognize the interface between HS and the stability of the state.[20]

HS advances the comprehensive assessment of threats by shifting focus from micro to systemic inquiry. The model does not address the entirety of human life but can be adapted to emerging and neglected forms of protection. While the concept of HS has evolved, the promotion of a new framework by the United Nations Development Programme (UNDP) in 1994 was particularly monumental.[21] Another significant milestone was set in 2012 when the UN General Assembly adopted resolution 66/290 on human security.[22] Throughout its evolution, the framework has maintained seven

16 "Human Security Handbook: An Integrated Approach for the Realization of the Sustainable Development Goals and the Priority Areas of the International Community and the United Nations System," United Nations Human Security Unit (New York: United Nations, 2016), https://www.un.org/humansecurity/wp-content/uploads/2017/10/h2.pdf.

17 "Human Development Report 1994," United Nations Development Programme (UNDP), (New York: Oxford University Press, 1994).

18 "Human Security Handbook."

19 Richard Jolly and Deepayan Basu Ray, "The Human Security Framework and National Human Development Reports: A Review of Experiences and Current Debates," NHDR Occasional Paper 5 (2006).

20 "Human Development Report 1994."

21 "Human Security Handbook."

22 General Assembly Resolution 66/290 (2012, 10 September) Follow-up to Paragraph 143 on Human Security of the 2005 World Summit Outcome, A/RES/66/290,. United Nations General Assembly, 2012,. https://digitallibrary.un.org/record/737105.

foundational elements as outlined below.

- Political security
- Economic security
- Health security
- Environmental security
- Food security
- Personal security
- Community security

Whereas the value of the HS approach is widely recognized, there is a compelling array of criticisms of the approach.[23] The critique on scope is particularly relevant to this research. Jolly and Ray interrogate the amorphousness of the conceptual boundaries of human security, which they argue give considerable latitude to the inclusion of anything and everything.[24] Yet it is this degree of versatility that ironically lends the approach one of its key strengths when viewed in the context of emerging threats. This study therefore acknowledges and demonstrates the utility of the HS approach by underlining the dynamism of agency and contextual factors, of which technological opportunities and challenges are a part. The interplay between the elements of HS is additionally advanced as constitutional to the integration of online and offline environments or the cyber-physical space. Another important limitation of the HS paradigm pertains to the omission of psychosocial vulnerabilities and resilience,[25] a widely studied and thorny subject in relation to social media with a conflicting evidence base.[26] At the same time, advances (for example) in the use of machine learning algorithms in mental health diagnostics are widely recognized as a boon to the health sector.[27]

Despite the temper of the limitations that imbues the approach, HS as a policy tool has gained traction since 1994. The framework is pragmatic in as far as it addresses the interlinkages between political, social, economic, and other factors that contribute to extremism and associated harms in the cyber-physical space. Secondly, the critique on indefinite boundaries highlights the rigidity of

23 Roland Paris, "Human Security: Paradigm Shift or Hot Air?," International Security 26, no. 2 (2001): 87–102; Mark Duffield and Nicholas Waddell, "Human Security and Global Danger: Exploring a Governmental Assemblage," University of Lancaster, ESRC New Security Challenges programme, (2004) https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.116.141&rep=rep1&type=pdf; Keith Krause, "The Key to a Powerful Agenda, If Properly Delimited," Security Dialogue 35, no. 3 (September 1, 2004): 367–68, https://doi.org/10.1177/0967010604035003324; Tara McCormack, "Power and Agency in the Human Security Framework," Cambridge Review of International Affairs 21, no. 1 (2008): 113–28; Mary Martin and Taylor Owen, "The Second Generation of Human Security: Lessons from the UN and EU Experience," International Affairs 86, no. 1 (2010): 211–24.

24 Jolly and Ray, "The Human Security Framework."

25 Jennifer Leaning and Sam Arie, Human Security: A Framework for Assessment in Conflict and Transition (Cambridge, MA: Harvard Center for Population and Development Studies, 2000).

26 Roy H. Perlis et al., "Association between Social Media Use and Self-Reported Symptoms of Depression in US Adults," JAMA Network Open 4, no. 11 (2021): e2136113–e2136113; Patti M. Valkenburg, "Social Media Use and Well-Being: What We Know and What We Need to Know," Current Opinion in Psychology 45:101294 (2022) https://doi.org/10.1016/j.copsyc.2021.12.006

27 Gyeongcheol Cho et al., "Review of Machine Learning Algorithms for Diagnosing Mental Illness," Psychiatry Investigation 16, no. 4 (April 2019): 262–69, https://doi.org/10.30773/pi.2018.12.21.2; Andrew G. Reece and Christopher M. Danforth, "Instagram Photos Reveal Predictive Markers of Depression," EPJ Data Science 6, no. 1 (2016): 15.

the concept of security in international relations. This is particularly so when the state-centric model largely confines itself to the territorial boundaries, state interests, and other domains that advance the goals of those in power and relegate individuals and their relationship with the state to the margins.[28] The ongoing tech regulation processes that facilitate the constant renegotiation of the interests and needs of the people, corporates, and the state, resulting in the reconfiguration of power (a) symmetries, is particularly noteworthy. Such actions may in some instances engender flawed policies and generate broader impacts for democratization and autocratization. Beyond their conception as instruments for state-building with the potential to promote stability or fragility, digital technologies are additionally modeled and designated as regulatory targets of concern. State interventions aimed at the resolution of harms emanating from digital artifacts concomitantly serve as levers that reconfigure shifting power imbalances between dominant players. Meanwhile, the critique of the framework's indeterminate boundaries, while valid, also presents the dilemma of the degree to which the concept can maintain a realistic scope. Clearly defined boundaries can mitigate against conceptual overload, target the most critical universal systemic threats, and maintain conceptual clarity.

Against this backdrop, an emerging gap in the literature is the orientation of emerging technologies within the HS framework. While scholarship in the field is expanding, there has been limited engagement with technology as an essential component of the HS framework.[29] In view of the ongoing challenge of terrorists' exploitation of digital platforms and other online harms, does technology warrant a place in the HS framework? In 1999, the Canadian government's conception of HS in safety-oriented terms led it to declare that "the litmus test for determining if it is useful to frame an issue in HS terms is the degree to which the safety of people is at risk."[30] Digital technologies have progressively transformed social, political, and economic institutions with varying degrees of benefits, harms, and risks.[31] At the same time, the harms and benefits that accrue from AI not only emerge from the technologies themselves but are also rooted in the societal structures[32] within which technological (im)materiality[33] (including algorithmic design, development, and deployment) evolve. The exclusion of technology from the existing HS framework therefore amounts to a glaring omission of an important societal structure. In acknowledging the role of algorithms in facilitating extremism, the study integrates technological security into the existing model and examines technology as a constitutive element of the framework. It additionally examines its interconnectedness with the foundational elements of the UNDP framework.

28 Jolly and Ray, "The Human Security Framework."

29 Funda Onbaşi Gençoğlu, "Social Media and the Kurdish Issue in Turkey: Hate Speech, Free Speech and Human Security," Turkish Studies 16, no. 1 (2015): 115–30; Heather M. Roff, Advancing Human Security through Artificial Intelligence (London: Chatham House, 2017); Benjamin K. Sovacool et al., "Social Media and Disasters: Human Security, Environmental Racism, and Crisis Communication in Hurricane Irma Response," Environmental Sociology 6, no. 3 (2020): 291–306.

30 Human Security: Safety for People in a Changing World (Ottawa: Department of Home Affairs and Trade, 1999).

31 Sakiko Fukuda-Parr and Elizabeth Gibbons, "Emerging Consensus on 'Ethical AI': Human Rights Critique of Stakeholder Guidelines," Global Policy 12 (2021): 32–44.

32 Kadija Ferryman and Mikaela Pitcan, "Fairness in Precision Medicine,"Data & Society Research Institute, 2018, https://datasociety.net/wp-content/uploads/2018/02/DataSociety_Fairness_In_Precision_Medicine_Feb2018.pdf

33 Technological (im)materiality represents the physicality of technological artifacts as well as the latency of developmental and symbolic properties – for example, the values and goals – that they embody.

2.2 Technological Determinism or Not?: The Structure of Automated Agency

The long-running debate on structure and agency originally revolved around their primacy in shaping human behavior and society. Structural approaches emphasize the role of social structures[34] or socialization over the capacity of agency to transform structure. In contrast, individualist approaches advance the notion of human capacity to act independently and exercise free will but relegate the structural determinants of agency.[35] The complementary approach, which this study adopts, acknowledges the dual influence of structure on an individual's autonomy, and similarly views structure as maintained and transformed through the exercise of agency.[36] Social structures such as family, class, gender, religion, law, and economy invariably shape political, social, economic, technological, environmental, and other systems.[37] It is within this orbit that human security, its key building blocks (including political, economic, and environmental concerns), and its emphasis on the protection and empowerment of people similarly reside. The mutually constitutive relationship between structure and agency, as "geo-historically reproduced," further underlines the currency of time and historical contexts, in particular, as they relate to human behavior and social systems.[38] Historical contexts detract from the incident-driven character of contemporary extremism research and promote the analysis and interpretation of events, persons, and works over time.[39] The inquiry on online-offline extremism is as such enriched when the agency of the plurality of actors and drivers of extremism at the micro and systemic levels flexibly incorporate these parameters.

There are as many diverse representations of the intrinsic formation of agency as there are typologies. Agency broadly refers to the capacity of an individual to act and can additionally indicate "the exercise or manifestation of this capacity."[40] Agency, advanced as either individual, collective, or proxy, denotes the capacity to desire or conceive, form intentions or goals, and act.[41] Agency conceived as both individual and collective takes into account intrinsic agency (power within) and instrumental agency (power to).[42] Needs, values, and roles play an important role in agency. The mechanics by which agency is co-constituted and is additionally (re)productive of structure are essential analytical considerations. In considering these assertions, an important issue relating

34 Social structure refers to "patterned social arrangements which form the society as a whole, and which determine, to some varying degree, the actions of the individuals socialised into that structure" (Olanike F. Deji, Gender and Rural Development: Introduction, vol. 1 (Berlin: LIT Verlag Münster, 2011), 71). Social structures as institutional are both formal and informal "stable, valued, recurring patterns of behavior" (Samuel P. Huntington, Political Order in Changing Societies (New Haven: Yale University Press, 2006), 9) or "integrated systems of rules that structure social interactions" (Geoffrey M. Hodgson, "On Defining Institutions: Rules versus Equilibria," Journal of Institutional Economics 11, no. 3 (2015), 501). They broadly include structure, function, and culture, or relate to the shared rules, norms, habits, roles, values, and other socially approved behavior that inform the interactions within the social structure (see Seumas Miller, "Social Institutions," The Stanford Encyclopedia of Philosophy, 2019, https://plato.stanford.edu/archives/sum2019/entries/social-institutions/)

35 Christian Kimmich and Ferdinand Wenzlaff, "The Structure–Agency Relation of Growth Imperative Hypotheses in a Credit Economy," New Political Economy 27, no. 2 (2022): 277–95.

36 Anthony Giddens, The Constitution of Society: Outline of the Theory of Structuration (University of California Press, 1984).

37 Deji, Gender and Rural Development.

38 Roy Bhaskar, The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences, (New York: Routledge, 1998).

39 Bart Schuurman, "Topics in Terrorism Research: Reviewing Trends and Gaps, 2007-2016," Critical Studies on Terrorism 12, no. 3 (2019): 463–80.

40 Markus Schlosser, "Agency," The Stanford Encyclopedia of Philosophy, n.d., https://plato.stanford.edu/archives/win2019/entries/agency/.

41 Albert Mills, Durepos Gabrielle, and Wiebe Elden, Encyclopedia of Case Study Research (Los Angeles, CA: SAGE Publications, 2011).

42 William H. Sewell Jr., "A Theory of Structure: Duality, Agency, and Transformation," American Journal of Sociology 98, no. 1 (1992): 1–29.

to pre-(machine) selection or user selection of online content foregrounds the wider debate on technological determinism.[43] A pertinent question subsequently becomes the extent to which humans and algorithms determine the design, development, deployment, and utilization of technology and its attendant impacts. Is it one or the other, or is the process synergistic?

As the research grapples with the complexity of these questions, the relative "antiquity" of technological determinism is in itself noteworthy, as it predates current and emerging technological challenges. In fact, the notion of technological determinism has gradually evolved since circa the 1890s/1900s.[44] Early critical studies further influenced the thinking on technological determinism, foundational to the conception of agency.[45] The scope of thought has gradually expanded from the uncontrollability and potentially malign influence of technology to the deterministic weight of social and technological influences. Social shaping theory and technological determinism are somewhat antithetical in their claims on the dominance of social systems and technical effects along the techno-social spectrum.[46]

Technological determinism interrogates the extent to which technological factors influence human thought or action. The theory pre-positions technology as the principal driver of social and cultural transformation, and technology as such determines the course of history. Technology is likely to produce unexpected or unintended effects in periods of "technological drift." Humans are modeled as subservient to technology and consequently forced to adapt to prevailing technological changes.[47] The extreme version of this position is "autonomous technological determinism," which denotes the complete shift in control from human direction to the realm of sovereign technologies.

While the claim of technological determinism holds some merit, it overlooks the role of human agency in technological processes. How compelling is this proposition when for example, users funnel data for the design and refinement of evolving algorithmic systems engineered by goal-oriented humans? In considering this question, Lévy posits that negative technological outcomes stem not from the inherent nature of technology but people's poor utilization of these artifacts.[48] This assertion however disregards the complex interplay between social, economic, and other back-end factors that inform the conceptualization, design, and development of technologies.

In contrast, social determinism advances the importance of social conditions in influencing technological progress. More importantly, this notion rejects the idea of mutual exclusivity and underlines the inter-mutuality of social and technological modeling. Technology is the work of

43 Joe Whittaker et al., "Recommender Systems and the Amplification of Extremist Content," Internet Policy Review 10, no. 2 (2021): 1–29.

44 Charles A. Beard, "Time, Technology, and the Creative Spirit in Political Science," The American Political Science Review 21, no. 1 (1927): 1–11; Thorstein Veblen, The Instinct of Workmanship and the State of the Industrial Arts (New York: Augustus Kelley, 1914/1990); Thorstein Veblen, The Engineers and the Price System (New York: Kelley, 1965); Thorstein Veblen, Imperial Germany and the Industrial Revolution (Livingston, NJ: Transaction Publishers, 1990).

45 Lewis Mumford, Technics and Civilization (New York: Harcourt, Brace & World, 1963); Lewis Mumford, The Myth of the Machine: Technics and Human Development, vol. 1 (London: Secker & Warburg, 1967).

46 Kimmich and Wenzlaff, "The Structure–Agency Relation."

47 Langdon Winner, Autonomous Technology: Technics-Out-of-Control as a Theme in Political Thought (Cambridge, MA: MIT Press, 1977).

48 Pierre Lévy, Becoming Virtual: Reality in the Digital Age (New York: Plenum Trade, 1998).

humanity and is overlooked only if humans permit it. Technology is therefore one among other social processes explicated in the HS approach that society creates and develops while preferencing the use of some forms over others.[49] This philosophy firmly grounds contextuality as the overarching concept for examining the role of algorithms in extremism. In departing from these extreme positions, centrist views regard technology as mediatory. Whereas technology drives social and cultural changes, humans unequivocally control technology.

There have been several attempts to conceptualize and typify the competing doctrines of techno-determinism. The two dominant schools of thought represent radical (hard) and moderate (soft) positions. The hard proponents conceive technology as an imperative for social transformation while the soft cohort designates technology as a key factor that may or may not necessitate change. To illustrate, Bimber advances three strands of techno-deterministic thought that include normative, nomological, and theses on unintended consequences. Normative representations assert that society has ceded control to technology, with the technological goals of productivity and efficiency deemed to have supplanted political and ethical norms. Nomological interpretations impute the laws of nature in the autonomous development of technology. This internal logic of technology subsequently compels predetermined social change. Finally, the tenets of unintended consequences essentially challenge determinism. This school of thought reinforces the notion of the disruptive nature of technology in its production of unpredictable social outcomes.[50]

Overall, algorithmic agency may be conceived as conferred human agency (in algorithmic design and development), innate or automated (algorithmic) agency (in machine learning), or the convergence of both. Conferred agency represents human agency by proxy and is additionally constituted through the continuous modification of algorithm designs aimed at achieving pre-defined outcomes.[51] The transplantation of worldviews, values, goals, and other personal or institutional inclinations into the structure of algorithms is an important aspect of algorithm design. The continuous loop of interaction between human and machine denotes that there is more likely than not an element of human mediation that presents another dimension of reciprocal agency. Upon deployment, certain goal-oriented users devise strategies to optimize or manipulate algorithmic systems, resulting in the compounding of agency. The assemblage of back-end and front-end mechanics is therefore a representation of collective agency. It can also be argued that if the starting point or the activation of algorithmic agency occurs through human intervention, is universal algorithmic autonomy attainable? In considering the fragmentations of agency alongside its collectivizations, what should policymakers focus on? In addition, how can the legal system fairly and justly establish and apportion liability for extremism in the "online" space? The paper will delve into these complexities in greater detail in subsequent sections.

49 Thomas Hauer, "Technological Determinism and New Media," International Journal of English Literature and Social Sciences 2, no. 2 (2017): 1–4.

50 Bruce Bimber, "Three Faces of Technological Determinism," in Does Technology Drive History, eds. Merrit Roe Smith and Leo Marx (Cambridge, MA: MIT Press, 1994), 79–100.

51 Rowa, "Part 1."

2.3 Reciprocal Agency and the Algorithm Conundrum

The current discourse on the role of algorithms in (non)violent extremism has for the most part constructed an illusory divide between online and offline spaces. It has also led to the "silofication" of actors who in reality are highly interconnected. These conceptions, embedded in broader discourses on social media, have cast the spotlight on how technology companies should be regulated and the actions they should take to tackle extremists' exploitation of their platforms. In considering the multiplicity, interactivity, and complexity of emerging issues in upcoming sections, to what extent is this approach responsive to the interconnected dynamics of extremism in the cyber-physical space? In highlighting the complexities of this question and the elusiveness of "magic bullet" solutions, the relationship between the state and digital actors demands a brief revisitation.

In the 1980s and 1990s, the internet was conceived as a decentralized network without a central command, an attribute that was elemental to its resilience and complicated its governance. These characteristics, analysts argued, portrayed the internet as a "post-national situation" and led to speculation about the decline of the state. These representations have since evolved into misconceptions because the state never quite left the scene. In fact, the state refrained from running the internet, opting instead for more regulatory functions. The state's limited involvement in the information domain subsequently empowered the "invisible hand" of market powers. Some of the new actors that emerged in the information industry – including online service providers, search engines, application developers, content producers, and internet service providers – became increasingly powerful.[52]

In the 2000s, the state made a stunning "comeback." It began deploying ready-made and often-centralized private technologies of power. The co-option of these technologies to serve the state, in partnership with powerful private actors in some instances, has led to the convergence of public-private interests. This dynamic of reciprocity is evident in the state's investment in sophisticated surveillance technologies and the patronization of consumer databanks that private companies administrate.[53] Meanwhile, the interests of online businesses to combat extremism, fraud, and other cybercrimes incentivize collaboration with the state. It therefore turns out that the "invisible hand" has since morphed into an "invisible handshake" that is particularly beneficial to the state.[54]

Overall, the attempt to understand extremism in the cyber-physical space and the agency of algorithms in this domain should take into consideration the complex convergence of drivers and agencies highlighted in the HS framework. The complex structure of agency, as the findings will show, encapsulates actors beyond the state and industry. The findings will further expose the misconceptions of lone wolf algorithms and demonstrate that piecemeal and algo-centric approaches to (non)violent extremism are likely to (re)produce piecemeal results. Consequently, policymakers should consider the potential for existing and emerging policies to strengthen people's

52 Michael Birnhack and Niva Elkin-Koren, "The Invisible Handshake: The Reemergence of the State in the Digital Environment," SSRN Scholarly Paper, April 10, 2003, https://doi.org/10.2139/ssrn.381020.

53 Birnhack and Elkin-Koren, "The Invisible Handshake"; Nicholas Vincent et al., "Data Leverage: A Framework for Empowering the Public in Its Relationship with Technology Companies," in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (2021), 215–27.

54 Birnhack and Elkin-Koren, "The Invisible Handshake."

capacity, including the ability to make and account for their choices and address systemic issues.[55] Such an approach should additionally aim to establish linkages between algorithms and broader contextual conditions.

HS represents a strong entry point for the appreciation and examination of the complexities, the multi-dimensional character, and interlinkages among issues and actors. The focus on prevention addresses such pertinent issues as safety by design that provokes further interrogations of the broad mosaic of extremism in the cyber-physical space. Technology as integrated into the framework of this study therefore constitutes a stand-alone as well as an overarching component of HS due to its influence on every sector of society.

A PRESENTATION, ANALYSIS, AND DISCUSSION OF FINDINGS

Section 3: Contextual Considerations

3.1 The Value of Context in Analytical, Policy, and Operational Postures

The agency of internet users is closely intertwined with policy content, process, and context in the grand scheme of policy coherence. In considering the notion of discursive security,[56] specific attention should be placed on existing tech public policy designs, related outputs, and corresponding public discourse. It is within this space that the existing concepts and policies aimed at protecting cyberspace have in some instances (re)produced insecurity through the creation of new vulnerabilities.[57] For example, the policies on content moderation highlight some of the challenges relating to terrorist designations and the management of associated online content.[58] Relatedly, the discourse on tech legislation is predominantly elite-driven with little to no engagement of users who can significantly enrich contextual understandings.[59] It is against this backdrop that Moat, Lavis, and Abelson extensively engage with contextual factors. The researchers conducted a critical review of the literature on the importance of policy issues and the political, social, economic, and other contexts in which policymaking takes place. The study found that contextual factors, notably institutions such as government structures, societal values, or ideas, and the interests of relevant actors have a significant influence on policy. The study recommended that policymakers consider the utility of relevant context and issue-related factors in the policymaking process.[60]

55 Jolly and Ray, "The Human Security Framework."

56 Philippe Fournier, "The Neoliberal/Neurotic Citizen and Security as Discourse," Critical Studies on Security 2, no. 3 (2014): 309–22.

57 Myriam Dunn Cavelty, "Breaking the Cyber-Security Dilemma: Aligning Security Needs and Removing Vulnerabilities," Science and Engineering Ethics 20, no. 3 (September 1, 2014): 701–15, https://doi.org/10.1007/s11948-014-9551-y.

58 "Content Personalisation and the Online Dissemination of Terrorist and Violent Extremist Content," Tech Against Terrorism, 2021, https://www.techagainstterrorism.org/wp-content/uploads/2021/02/TAT-Position-Paper-content-personalisation-and-online-dissemination-of-terrorist-content1.pdf.

59 Asya Cooley et al., "Influencing Public Behavior: Takeaways From Public Communication Scholarship," The Media Ecology and Strategic Analysis Group, October, 2020, https://apps.dtic.mil/sti/citations/AD1118281.

60 Kaelan A. Moat, John N. Lavis, and Julia Abelson, "How Contexts and Issues Influence the Use of Policy-Relevant Research Syntheses: A Critical Interpretive Synthesis," The Milbank Quarterly 91, no. 3 (2013): 604–48.

Context, it turns out, is not only valuable in policymaking but is also the realm in which policy and practice interact and conflict. A particular concern emerging from interviews with practitioners was the dislodging of context from mainstream Countering Violent Extremism (CVE) and Counter Terrorism (CT) interventions. The historical as well as other contextual drivers of extremism that have been widely debated in the academic literature were cited as forming the cornerstone not only for a comprehensive understanding of the threat of extremism but also as facilitating or informing policies on CT and CVE. These views are consistent with fairly recent research findings on the incident-driven character of contemporary extremism research.[61] For example, in acknowledging the changing extremist context, the French government recognized the growing threat of right-wing extremism alongside Islamism. The government has established online extremist content as falling within the purview of religious, political, and social extremism. In addition, "contextual extremist content" (or what could be referred to as cyclical extremist content) typically spikes during every French presidential election. This is because (according to a French government official) federal elections are intrinsically polarizing.

It therefore follows that user agency and lived experiences not only particularize contextual phenomena but also inform the integration of the online and offline dynamics of extremism. A section of the scholarship indeed highlights the primacy of online-offline interactions[62] and the potential for offline networks to outpace the internet in fostering extremism.[63] These fundamentals essentially signal the import of offline antecedents, the interactive agency between algorithms and users, and the value of according both "actors" equal scrutiny. They cast a spotlight on the body of literature that has extensively engaged with the contextual questions on extremism[64] that complement the limited corpus on online-offline linkages.[65] Besides increasing the potential for generating a more comprehensive understanding of online-offline extremism, these complexities have broader implications for empirical studies. The conceptualization of the multifaceted character and functionality of algorithms and the design of empirical studies that can enhance understanding of algorithmic decision-making in context are critical subjects of ongoing scholarly exertions. Overall, the upcoming sections will demonstrate greater resonance with the perspectives that have so far laid the basis for a better appreciation of the contextuality of lone wolf algorithms.

3.2 Existing Global Tech Legislation: The Online-Offline Context of Algorithmism

Tech Against Terrorism has evaluated and compiled a comprehensive living handbook on existing global tech legislation in five regions. The report identifies, evaluates, and provides brief synopses of country-specific regulatory frameworks. The document highlights relevant national bodies and

61 Schuurman, "Topics in Terrorism Research."

62 Paul Gill et al., "Terrorist Use of the Internet by the Numbers: Quantifying Behaviors, Patterns, and Processes," Criminology & Public Policy 16, no. 1 (2017): 99–117. Joe Whittaker, "The Online Behaviors of Islamic State Terrorists in the United States," Criminology & Public Policy 20, no. 1 (2021): 177–203.

63 Sean C. Reynolds and Mohammed M. Hafez, "Social Network Analysis of German Foreign Fighters in Syria and Iraq," Terrorism and Political Violence 31, no. 4 (2017): 661–86.

64 Martha Crenshaw, Terrorism in Context (University Park, PA: Penn State Press, 1995); Gary LaFree, Laura Dugan, and Erin Miller, Putting Terrorism in Context: Lessons from the Global Terrorism Database (New York: Routledge, 2015).

65 Gill et al., "Terrorist Use of the Internet"; Reynolds and Hafez, "Social Network Analysis."

additionally outlines key concerns, important takeaways, and policy recommendations.[66] It is not the aim of this study to replicate the content in the toolkit but to draw attention to this important resource for further reference. This section will highlight relevant legislation that advances the analysis of the three working group themes and emerging subthemes on algorithmic systems, legal frameworks, and transparency as they relate to the online-offline context.

The current discourse on the role of algorithms in (non)violent extremism has for the most part created a false dichotomy between the online and offline spaces. This discourse has generated greater attention on the modalities of effective tech regulation and the design of appropriate interventions to tackle extremists' exploitation of digital platforms. An emerging question, therefore, is the extent to which an algorithm-centric approach is holistic and responsive to the complex dynamics of "online" extremism. Relatedly, what interlinkages potentially exist between the online and offline domains when the broader extremist ecosystem is examined? Secondly, the discourse on "online" extremism is largely structured around lone wolf algorithms and the concept of "algorithmic amplification."

An emerging pertinent question relates to whether algorithms operate in a vacuum or embody some important interactivities that are currently overlooked or understudied. The skepticism around online-offline linkages in some quarters similarly provokes further reflection on the conceptualization of the online space in relation to algorithms. These propositions uncover additional areas of scrutiny that span the following questions: what constitutes the online environment? Is this domain confined to the front-end interface visible to users or does it subsume the underlying environment of algorithmic mechanics? What are the implications of any possible response to the notion of "algorithmic amplification" of extremist content? A nuanced exploration of these questions has important implications for policy and practice.

The U.S. government has set the context for the discussion on online-offline extremism and the import of holistic and integrated approaches. The government reported that the policy direction for algorithmic operations and platform governance is not yet determined. Both the current administration and Congress are currently considering a wide range of legislative and non-legislative tools in their attempts to chart a policy roadmap. Any changes to legislation will be instituted within the framework of the U.S. Constitution and other laws. More specifically, legislative changes are expected to align with U.S. constitutional protections for freedom of speech, including international obligations and commitments to human rights (such as the freedom of expression). So far, proposed laws that target algorithmic amplification include the Justice Against Malicious Algorithms Act of 2021 and the DISCOURSE Act.

A U.S. government official identified several knowledge gaps that limit understanding of the links between online and offline extremism with implications for policy. To begin with, the migration of violent extremists to less regulated, closed, or encrypted applications hampers the ability of government authorities to collect information and effectively respond to threats. Additional intelligence gaps relate to the interlinkages among online hate speech, online terrorist and violent

66 "Tech Against Terrorism Annual Report 2020–2021," Tech Against Terrorism, August 2021, https://www.techagainstterrorism.org/wp-content/uploads/2021/09/TAT-ANNUAL-REPORT_2020-21%E2%80%93FINAL.pdf.

extremist content, and domestic violent extremism. These gaps for the most part pertain to the correlation between online content and offline violence. Thirdly, there is limited knowledge on the relationship between online conspiracy theories and offline acts of violence. Subsequently, a better understanding of the nature and role of actors (whether states or other entities) that commit violence has the potential to enrich ongoing interventions.

The U.S. government views certain algorithmic processes as potentially significant in the amplification of terrorist and violent extremist content within the online and arguably offline domains. This similarly applies to misinformation, disinformation, malinformation, and hateful online content that technology companies have the mandate to address in partnership with other stakeholders. Overall, these issues are partial representations of a complex and challenging coalition of threats. Notably, the government supports equal scrutiny of the interlinkages and differences between the online and offline extremist ecosystems. These observations, besides highlighting the limited evidence base, additionally draw attention to the potential interlinkages between various forms of harms. The assessment further represents an appreciation of complexities in knowledge and practice and underscores a more nuanced understanding of actors.

In some cases, the texts of recent policies and legislation on terrorism and extremism integrate the offline and online contexts. The French government (for example) confers equal treatment on both online and offline hate. A government official drew attention to the "Endorsement of Respect for the Principles of the Republic and Counter Separatism" as one law that was amended to incorporate online content. This law was originally designed to address political and religious separatism but was amended following the "Islamist" terrorist attack on Samuel Paty in 2017. Among other goals, the amendment was aimed at granting judicial authority sufficient means and funds to prosecute harmful online behavior (including doxing). The French official additionally acknowledged that "What happens in the online realm is not detached from real-life actions... Digital regulation is a complex issue with interlinked effects."

The appreciation of the complexity and interlinkages between online-offline extremism and other harms as reflected in the French position should inform regulatory thinking and practice. This example is additionally indicative of the synergy and dynamism of contextual influences and regulation. Notably, the killing of Samuel Paty represents a situation in which a social media campaign, including the accumulation and online distribution of private information, led to violence. It is also a case in which an incident originating offline diffused into the online environment, then circled back to the offline domain, leading to the killing of Paty in the real world. As discussed in greater detail below, the discourse is heavily skewed toward the consequences of online behavior in the offline world. Yet the reverse scenario, and more importantly the interactivities between the online and offline spheres, demonstrate that these environments are mutually generative. It is this circular feedback loop in which a multiplicity of actors and factors interact that moderates algorithm-centric contentions on extremism.

Whereas some governments have integrated the phenomenon of "online" extremism into official government documents, deliberate attempts to establish clear linkages between the online and offline domains remain nominal. The same applies to the integration of the offline effects of "online"

extremism, as demonstrated (for example) in the Australian eSafety Regulatory Guidance on Abhorrent Violent Conduct Powers, which states,

> eSafety may issue a removal notice requiring certain online service providers to remove illegal and restricted material, including material that advocates the doing of a terrorist act and material that promotes, instructs or incites in matters of crime and violence.[67]

While the offline-online links are not explicitly acknowledged, they are somewhat implied in the texts of relevant government documents, particularly as they relate to the offline effects of online extremism.

The Australian government reported the formation of the Digital Platform Regulators Forum that convenes the eSafety Commissioner, the Australian Information Commissioner, the Australian Communications and Media Authority, and the Australian Competition and Consumer Commission. The forum aims to enhance cooperation, information sharing, and the harmonization of approaches to the regulation of digital platforms. In particular, the Online Safety Act 2021 that came into effect in 2022 grants the eSafety Commissioner, an independent government agency, the mandate to regulate online safety. The agency acts on four reporting schemes that encompass adult cyber abuse, the cyberbullying of children, image-based abuse, and an online content scheme for illegal and restricted content. The online content scheme guides the removal and remedial notices to social media, electronic, or designated internet services that host illegal and harmful content.[68] The Australian government further reported that the basic online safety expectations is an important component of the Online Safety Act that outlines the government's expectations for social media and other e-services. The issuance of notices is expected to begin in mid-2022.

As part of the overall scope of the program, Australia's eSafety Commissioner will assess the fulfillment of basic online safety expectations by companies in the context of algorithms. The government does not expect a splintering of the algorithm black box. It instead aims to understand the safety protocols integrated into the design of company systems, processes, and procedures, and mechanisms for problem-solving. The eSafety Commissioner has developed two sets of publicly available risk assessment tools that small start-ups and large businesses can access on its website for free. The agency recommends a cost-effective model that involves the use of these tools in the assessment of tech design and operating systems, and in embedding safety protocols from the outset. The agency similarly recognizes the capacity constraints of small companies. In this regard, significant discretion is built into the legislation that takes into account company size and other dimensions of institutional capacity. The agency has so far developed best practices on safety and prevention as part of its broader capacity-building strategy. Finally, in considering the scale of online crime, the government observed that the strategies and vulnerabilities that are commonly exploited for terrorism can be similarly exploited in other types of crimes. The collaboration between

67 eSafety Commissioner Australia, "Abhorrent Violent Conduct Powers: Regulatory Guidance," 2021, https://www.esafety.gov.au/sites/default/files/2022-03/Abhorrent%20Violent%20Conduct%20Powers%20Regulatory%20Guidance.pdf.

68 eSafety Commissioner Australia, "An Overview of ESafety's Role and Functions," 2021, https://www.esafety.gov.au/sites/default/files/2021-07/Overview%20of%20role%20and%20functions_0.pdf.

government and industry in the development of safety by design tools is indeed remarkable.

The European Union is the only region in the world that has taken a collective approach to tech legislation. The "Regulation 2021/784 on addressing the dissemination of terrorist content online" or TCO will come into force in 2022. According to a French government official, the law has prompted the EU to reflect on the dissemination of strategic information through multiple layers of governance. It demands the reorganization of the European protocol through Europol and at national levels. An EU representative added that terrorist content is narrowly defined in this law and has possibly omitted other types of content that could lead to radicalization or violent extremism. This is in part due to the tension between online safety and freedom of expression. Propaganda material, for example, does not fit the current definition of terrorism as stipulated in the text and has been omitted as a result. Nevertheless, as with other laws, the text has indeterminately established the link between online terrorist content and consequences in the offline space. Some of the shortcomings of the TCO are addressed in the EU "Directive 2017/451." Commonly known as the Terrorism Directive, it is the key guiding document for EU regional counterterrorism efforts. It supplants previous texts and was partly introduced to address terrorist use of the internet. The directive outlines the definitions of terminology, including terrorist groups, offenses, and propaganda.

The Digital Services Act (DSA) is the EU's landmark legislation in this area. Widely regarded as more balanced in comparison to other existing laws, it aims to introduce new regimes that address harmful and illegal online content.[69] A French government official particularly highlighted the framework for transparency that the DSA aims to create. This framework could promote accountability through proportionate obligations, incremental penalties, and other measures as is the case with GDPR. The EU representative noted that this act represents a horizontal legislative tool. It is for this reason that it is laden with articles on transparency in order to scrutinize (for instance) how platforms promote or recommend certain types of content. Risk management tools have been incorporated to ensure companies understand the risks of platforming harmful and illegal content. It is possible that this act will have a significant impact on transparency and accountability. That said, some challenges related to achieving meaningful transparency will be covered in later sections.

Besides the regulatory and policy frameworks, the EU has designed voluntary tools to support the tech community to address online harms. These instruments include the "EU code of conduct on countering illegal hate speech online" that companies have ratified and the "EU action plan against disinformation." The EU Internet Forum has worked with various technology companies on online terrorist content, child sexual abuse, and other issues pertinent to online safety. An EU official identified other possible areas of cooperation and mutual concern (such as drug and human trafficking). She concluded by stating that regulation is just one among several prescriptions that technology companies will progressively adapt to over the next two to three years.

In considering the state of the regulations examined so far, does the linkage between "online" and offline extremism deserve closer scrutiny? Two government representatives expressed reservations over the broad remits in the field of inquiry that overstepped the domain of algorithmic systems. This

69 "Tech Against Terrorism Annual Report 2020–2021."

study's secondary focus on policy issues and the interlinkages between online and offline extremism was deemed as diversionary, insulating technology companies and algorithms from liability and shifting responsibility to governments. These representatives advanced their priority as tackling the prevailing threat of online extremism and related harms. These insights are particularly valuable in the examination of the techno-mediative character of algorithmic systems in the online-offline space.

3.3 The Techno-Mediative Character of Algorithms in the Cyber-Physical Space

Algorithms are adaptable systems that personalize content based on data collected through online and offline user behavior. They typically evolve in synchrony with users' continuous modification of personal online information. Notably, the algorithm's predictive capabilities are heavily reliant on users' offline data.[70] Tailored and personalized content generally reflect users' geographic location, devices used, travels, actual purchases, personal interests, tastes, beliefs, private messages, engagements with online content (likes, comments, shares, ratings, views), past behavior such as search history, online session behavior, and other digital data.[71] This process occurs through a user-algorithm feedback loop that effectively instructs the integration of online-offline spaces.[72] Any allusion to online-offline interlinkages should therefore be predicated on the invisible and often unacknowledged integrative mechanics of algorithms. The techno-mediative character of algorithms is therefore foundational to the link between the online and offline realms. Ultimately, these mechanisms underline the importance of an integrated and holistic approach to tackling extremism in these highly interactive spaces.

The events surrounding and postdating the Christchurch attack can offer further insights on the merits of scrutinizing the two interlocking spheres and the limitations of a reductionist approach. The actions of the Christchurch terrorist, Brenton Tarrant, are more often than not examined within the context of the real-world consequences of "online" extremism. Tarrant's manifesto was widely distributed online prior to its proscription. While there were calls to delete and report social media posts, links, and websites in New Zealand, the extremist ideology underpinning the manifesto seems to be endemic in both online and offline extremist circles. The phrase "once you see it, you can't unsee it" (rather apt in this context), reflects the limitations of damage control (however appropriate or well-intentioned). This scenario is additionally analogous to the closure of the stable door after the horse has bolted. Even then, minimal effort is often applied in re-stabling the horse and probing and addressing the underlying drivers of flight that are simultaneously internal and external to the stable. As such, the ideals of Tarrant and his predecessors represent a highly adaptive, self-propelling doctrine that is propagated and perpetuated in a highly interactive cyber-physical environment. Yet the circulation of the manifesto in the offline domain, as a manifestation of the interplay between the online and offline, is a subject that is for the most part tangential within the academic, policy, and broader public

70 James N. Cohen, "Exploring Echo-Systems: How Algorithms Shape Immersive Media Environments," Journal of Media Literacy Education 10, no. 2 (2018): 139–51.

71 Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information (Cambridge, MA: Harvard University Press, 2015); Tarleton Gillespie, "The Relevance of Algorithms," Media Technologies: Essays on Communication, Materiality, and Society 167, no. 2014 (2012): 167; Aliza Vigderman and Gabe Turner, "The Data Big Tech Companies Have On You," Security.Org (blog), July 22, 2022, https://www.security.org/resources/data-tech-companies-have/.

72 Cohen, "Exploring Echo-Systems."

spaces.

To provide further context on online-offline linkages, Ukrainian and Russian examples highlight some of the intersections that characterize an integrated cyber-physical space. In 2019, a Ukrainian enthusiast of the Tarrant doctrine with a channel on Telegram claimed that he had found a publisher for the paperback version of the Christchurch attack manifesto and had secured the first box of bound copies. Meanwhile, Wotanjugend, a Russian group based in Ukraine, at the time established a neo-Nazi record label and shop that retailed in anti-Semitic and racist lyrics as well as Nazi paraphernalia.[73] This demonstrates that the trade and circulation of extremist material are pervasive in both the online and offline spaces, rendering the two domains as inextricably intertwined. Subsequently, requiring technology companies to remove extremist material from their platforms or modify their algorithms, while warranted, constitutes a piecemeal approach when concerted online efforts are not matched by offline interventions and culpable users are extricated from the equation. For technology companies, this additionally underscores the importance of internal codification, operationalization, and regular reviews of safety by design programs that mitigate against "bolting horse" situations. The following case studies on the Jasmine Revolution and the Buffalo attack attempt to broaden and clarify the online-offline linkages.

3.4 Case Study: The Jasmine Revolution in Tunisia

The Jasmine Revolution in Tunisia is a compelling representation of the dynamics of online-offline extremism. The current scholarship cites economic hardship, corruption, inequality, and social media usage as among the key triggers of the Tunisian revolution in 2010.[74] This uprising inspired the Arab Spring (popularly framed as the Twitter or Facebook revolution). At the outset, the confiscation of business paraphernalia (at the micro level) and systemic corruption in government provoked the self-immolation of a Tunisian street vendor, Mohamed Bouazizi. While initially an individualized form of dissent, Bouazizi's actions sparked a chain of events that mobilized collective grievances and led to mass protests that culminated in the ouster of President Ben Ali.[75]

The clashing trends of increased democratic gains alongside the rise in violent extremism since the toppling of President Ali in 2011 has significantly unsettled prevailing academic thought. These developments contradicted dominant assertions on the links between democratization and a decline in violent extremism.[76] At the same time, other scholars have advanced the likelihood of political

73 "The Russians and Ukrainians Translating the Christchurch Shooter's Manifesto," Bellingcat, August 14, 2019, https://www.bellingcat.com/news/uk-and-europe/2019/08/14/the-russians-and-ukrainians-translating-the-christchurch-shooters-manifesto/.

74 Zahraa Barakat and Ali Fakih, "Determinants of the Arab Spring Protests in Tunisia, Egypt, and Libya: What Have We Learned?," Social Sciences 10, no. 8 (2021): 282; Marion G. Müller and Celina Hübner, "How Facebook Facilitated the Jasmine Revolution. Conceptualizing the Functions of Online Social Network Communication," Journal of Social Media Studies 1, no. 1 (2014): 17–33.

75 Philip N. Howard and Muzammil M. Hussain, Democracy's Fourth Wave?: Digital Media and the Arab Spring (Oxford, UK: Oxford University Press, 2013).

76 Geoffrey Macdonald and Luke Waggoner, "Dashed Hopes and Extremism in Tunisia," Journal of Democracy 29, no. 1 (2018): 126–40; Amichai Magen, "Fighting Terrorism: The Democracy Advantage," Journal of Democracy 29, no. 1 (2018): 111–25.

and democratic transitions precipitating violent extremism, terrorism, or civil war.[77] Interestingly, the Jasmine Revolution period witnessed significant growth in incidents of domestic terrorism. In addition, the failure of democratically elected officials to provide an effective roadmap for democratic transition intensified residual grievances.[78]

The decline in state institutional capacity and unmanaged expectations fuelled increased mistrust and disillusionment in democratic governance. As a result, Islamist groups emerged, claimed legitimacy, and embarked on systematic recruitments of marginalized Tunisians to the Syrian jihad. Extremists exploited the historical tensions between the secular state and Islamist opposition and cast themselves as progressive as a means to enhance their recruitment campaigns.[79] It must be remembered that Islamism preceded both the Arab Spring and the democratization process in Tunisia.[80] Overall, these developments have generated mixed views and research findings on the role of social media in the revolution.

The debate surrounding the role of digital technology in the Arab Spring, while largely subdued, has been primarily cyber-centric. Some accounts suggest that "the internet, mobile phones, and social media" played a pivotal role in the mobilization of protesters.[81] A study on the determinants of the Arab Spring protests in Tunisia, Egypt, and Libya showed significant results on the frequency of social media use (among other factors) that increased the probability of participating in protests.[82] The conclusions are consistent with findings in other studies on social media and anti-government movements.[83] Other more nuanced anecdotes have emerged around algorithms as playing a pivotal role in the Arab Spring.[84]

The position of cyber-skeptics on the Jasmine Revolution is equally consequential. The cyber critics downplay the primacy of social media as a causal mechanism in popular resistance.[85] Other studies on the Arab Spring and countries in the Gulf have shown marginal to no correlation between

77 Katerina Dalacoura, "Terrorism, Democracy and Islamist Terrorism," in Islamist Terrorism and Democracy in the Middle East (Cambridge University Press, 2011), 21–39; Michel Wieviorka, "ETA and Basque Political Violence," in The Legitimization of Violence, ed. David E. Apter (Springer, 1997), 292–349.

78 Macdonald and Waggoner, "Dashed Hopes."

79 Macdonald and Waggoner, "Dashed Hopes"; Mariam Abdelaty, "Democratization and Extremism: The Case of Tunisia," Theses and Dissertations, June 15, 2021, https://fount.aucegypt.edu/etds/1671.

80 Abdelaty, "Democratization and Extremism."

81 Howard and Hussain, Democracy's Fourth Wave?, 35

82 Barakat and Fakih, "Determinants of the Arab Spring."

83 Davit Chokoshvili, "The Role of the Internet in Democratic Transition: Case Study of the Arab Spring" (Budapest, Central European University: Master of Arts in Public Policy 51, 2011); Alistair McKay, "The Arab Spring of Discontent," E-International Relations, 2011, http://www.e-ir.info/wp-content/uploads/arab-spring-collection-e-IR.pdf.

84 Julia Kassem, "Ten Years After 'Arab Spring,'" Al Mayadeen, October 10, 2021, https://english.almayadeen.net/articles/blog/ten-years-after-arab-spring; "Obama Tells Letterman How Algorithms Undermined Political Promise of Social Media," MarketWatch, January 18, 2018, https://www.marketwatch.com/story/obama-tells-letterman-how-algorithms-undermined-political-promise-of-social-media-2018-01-17.

85 Sean Aday et al., "Blogs and Bullets II: New Media and Conflict after the Arab Spring," United States Institute of Peace, July 10, 2012, https://www.usip.org/publications/2012/07/blogs-and-bullets-ii-new-media-and-conflict-after-arab-spring; Nadav Samin, "Saudi Arabia, Egypt, and the Social Media Moment," Arab Media & Society 15, no. 1 (2012): 46–65; Lisa Anderson, "Demystifying the Arab Spring," May/June 2011, https://www.foreignaffairs.com/articles/libya/2011-04-03/demystifying-arab-spring.

levels of unrest and high levels of internet connectivity.[86] Gladwell contends that the Arab Spring revolutions were primarily galvanized by traditional forms of political organization and inspired by collective grievances – as has been the case for centuries in the absence of real-time social media updates.[87] Youmans and York further submit that social media policies and other tools can simultaneously enable and constrain organized dissent and collective action.[88] On their part, Byun and Hollander concluded that there is no significant correlation between popular unrest and digitality. They recommend the examination of case-specific causes of civil unrest, as their analysis of the Tunisian example attempts to do, in highlighting the value of context specificity as advanced in the HS framework.[89]

The inconsistencies in research findings underscore the significance of methodological approaches that are examined in greater detail below. These appraisals additionally draw attention to important considerations in the examination of algorithmic amplification or the role of digital platforms in extremism. Some key parameters for future assessment could include the rate of internet penetration at the time, the credibility of information, and the differentials in social media usage, whether for political, social, economic, and/or other goals.[90] Ultimately, social media platforms that leverage bit. ly linkages appeared not to have played a role in either national collective action or regional diffusion in the Tunisian case. These forms of media "are more likely to spread information outside the region than inside it, acting like a megaphone more than a rallying cry." However, this does not invalidate the role of other forms of digital media. Moreover, the complex inter-relationship between social, traditional, and alternative media renders the segregation and apportionment of responsibility a herculean undertaking.[91]

To the outside world, digital technology exposed pre-existing deep-seated structural issues in Tunisia. At the micro level, technology facilitated mass mobilization but did not constitute a dominant driver of extremism during the Jasmine Revolution. It is for these reasons that algorithmic amplification in the context of extremism presupposes a baseline or the pre-existence of permissive conditions with the potential for online diffusion. The interactions between diverse actors and influences are important elements in this complex tapestry. Additionally, the context of limited internet penetration or greater access further complicates the role of the internet, casting a spotlight on the relationship between the online and offline environments and the attendant dynamics and impact of information diffusion. As such, the conception of algorithmic amplification should consider the synergy between algorithmic decision-making, user agency, and wide-ranging contextual and interconnected drivers of extremism.

86 Chonghyun Christie Byun and Ethan J. Hollander, "Explaining the Intensity of the Arab Spring," Digest of Middle East Studies 24, no. 1 (2015): 26–46, https://doi.org/10.1111/dome.12057; Howard and Hussain, Democracy's Fourth Wave?

87 Malcolm Gladwell, "Small Change," The New Yorker, October 4, 2010, https://www.newyorker.com/magazine/2010/10/04/small-change-malcolm-gladwell.

88 William Lafi Youmans and Jillian C. York, "Social Media and the Activist Toolkit: User Agreements, Corporate Interests, and the Information Infrastructure of Modern Social Movements," Journal of Communication 62, no. 2 (2012): 315–29.

89 Byun and Hollander, "Explaining the Intensity."

90 Heather Brown, Emily Guskin, and Amy Mitchell, "The Role of Social Media in the Arab Uprisings," Pew Research Center's Journalism Project (blog), November 28, 2012, https://www.pewresearch.org/journalism/2012/11/28/role-social-media-arab-uprisings/.

91 Aday et al., "Blogs and Bullets II"; Rowa, "Part 1."

3.5 Case Study: The Buffalo Terrorist Attack in the U.S.

In 2022, Payton Gendron drove approximately 200 miles and killed ten people and wounded three others at the Tops Grocery Store in Buffalo, New York. This racially motivated violent extremist attack was aimed at Black people. Gendron's manifesto was originally uploaded on Google Docs, circulated on 4chan, and briefly livestreamed on Twitch.[92] In preparation for the attack, Gendron uploaded a to-do list of items on an account he maintained on Discord.[93] A 180-page document posted online integrates sections of Tarrant's and other forebearers' creeds – in particular the Great Replacement theory. The theory is rooted in early 20th century French nationalism, featuring prominently in the works of Maurice Barrès who lived between 1862–1923.[94] These thoughts were later popularized by Renaud Camus in his book, Le Grand Replacement.[95]

Gendron, who describes himself in his manifesto as "populist, fascist, white supremacist, antisemite, and racist" directly cites Brenton Tarrant as an inspiration for the Buffalo attack. The manifesto in particular delves into immigration, a dominant theme in Tarrant's script.[96] Both terrorists reference this offline phenomenon as among other drivers that animated their extremist discourses in the online environment, culminating in attacks in the offline domain. The online space as such facilitates the convergence of real-life grievances, online conspiracy theories, and other narratives that either emerge online or predate the age of social media. According to a researcher interviewed for this study, this space bolsters greater engagement between like-minded people and enhances content visibility and virality than would have otherwise been the case.[97] That said, the assignment of the degree of human action and algorithmic culpability subsequently becomes a necessary yet unresolved question.

This scenario induces a causality dilemma and is additionally indicative of the transplantation and interplay of pre-existing offline grievances or structural drivers of extremism with the online sphere. As such, the attacks are remarkably instructive of an online-offline feedback loop that models the cyber-physical space. A point made earlier that merits restating relates to the prevailing discourse regarding "online" extremism that is heavily skewed towards the offline consequences of online extremism. There is little to no regard for the links between online-offline causal factors of extremism. The paper engages with this dimension in more depth below. Online-offline linkages are additionally evident in the hybrid mobilization and engagement of actors and the design of strategies and interventions that configure the cyber-physical space. These include the hybrid programs of CVE and peacebuilding practitioners, techno-mediative initiatives such as YouTube's preservation of content created by human rights defenders, and other phygital initiatives by governments, researchers, and

92 Ben Collins, "The Buffalo Shooting Suspect Apparently Posted a Manifesto Citing 'Great Replacement' Theory," NBC News, May 15, 2022, https://www.nbcnews.com/news/us-news/buffalo-supermarket-shooting-suspect-posted-apparent-manifesto-repeate-rcna28889.

93 Chas Danner, "What We Know About the Racist Attack at a Buffalo Supermarket," Intelligencer, May 17, 2022, https://nymag.com/intelligencer/2022/05/ten-dead-after-gunman-attacks-buffalo-supermarket-updates.html.

94 Robert Soucy, "Barrès and Fascism," French Historical Studies 5, no. 1 (1967): 67–97.

95 Renaud Camus, Le grand remplacement (Plieux: Renaud Camus, 2012).

96 Bridget Johnson, "10 Killed in Buffalo Supermarket Attack Allegedly Inspired by Christchurch Terrorist," HS Today.US, May 15, 2022, https://www.hstoday.us/featured/10-killed-in-buffalo-supermarket-attack-allegedly-inspired-by-christchurch-terrorist/.

97 See also Danner, "What We Know."

other actors.

Closing some information gaps could provide a clearer picture of Gendron's radicalization trajectory. These include information on what preceded or led to his discovery of 4chan, and what particularly drew him to extremist and other harmful content. Was this circumstantial or premeditated or informed by other factors preceding or surrounding his exposure to extremist material? It can be argued that online research may have facilitated his offline reconnaissance trips. In considering that possibility, what other online tools and websites did he utilize? It was during one of his offline scoping visits that Gendron mapped the layout of the store, enumerated the proportion of Black shoppers, and surveilled the movement of security personnel. Thereafter, Gendron converted the data gathered from offline reconnaissance trips into an online toolkit that could inspire and guide future attacks. These online-offline linkages are particularly instrumental in shedding light on the cyber-physicality of extremism.

Assuming the exploration of the broader historical context will not provide sufficient information to fully analyze Gendron's actions in context, two arguments can be advanced. In a Discord chat log, Gendron noted, "I only really turned racist when 4chan started giving me facts." In this instance, it is important to outline clear delimitations and associations between the role of a platform and its users. Users with variable goals voluntarily opted to join 4chan. Why was 4chan a platform of choice? Potential reasons could relate (among other facilitative factors) to platform structure and permissiveness. Besides providing the space for discourse, 4chan additionally hosted information that was created and uploaded by users, which Gendron and other users consumed. All things considered, was he radicalized by the platform or the content that other users created that was uploaded and hosted on the platform? Both entities likely bear responsibility. These sticking points and more within the user-platform value chain warrant further exploration. Gendron's assertion that he only turned racist upon exposure to 4chan is therefore highly contentious.

Gendron has also been described as a lonely "teenager" who was allegedly radicalized at the height of the COVID pandemic. Classmates have described Gendron as a "quiet, studious boy who got high marks but seemed out of place in recent years, turning to online streaming games, a fascination with guns and ways to grab attention from peers... Most people didn't associate with him. They didn't want to be known as friends with a kid who was socially awkward and nerdy." When Gendron did speak, "it was about isolation, rejection, and desperation" in the offline world.[98] While there is a dearth of information on his past, the concept of "cognitive opening"[99] in relation to the search for identity, meaning, and belonging has enhanced researchers' understanding on possible radicalization pathways.[100] In Gendron's case, these offline personal deprivations appear to have diffused and consequently mediated his online interactions, thereby establishing further links between online-offline extremism. These dynamics, including the widely cited role of COVID-19, additionally centralize the notion of contextuality.

98 Bernard Condon and Michael Hill, "Buffalo Mass Shooting Suspect: 'Lonely,' 'Nerdy' Teenager Showed Signs of Trouble," Global News, May 17, 2022, https://globalnews.ca/news/8842287/buffalo-mass-shooting-suspect-payton-gendron/.

99 Quintan Wiktorowicz, Radical Islam Rising: Muslim Extremism in the West (Oxford, UK: Rowman & Littlefield Publishers, 2005).

100 Yvonne Rowa, "Liminal Boundaries and Vulnerabilities to Radicalisation in the Context of Securitisation of Migration" (PhD Thesis, University of Adelaide, 2019); Wiktorowicz, Radical Islam Rising.

The online activities of both Gendron and Tarrant similarly highlight the significance of early detection and corresponding deployment of online and offline preventative measures. Yet while propositions for preventative measures are often popularly touted, research shows the complexities around the assessment and interpretations of online risk factors and the management of imminent threats.[101] Further reports suggest that Gendron was previously known to the police as a result of a threat he had made to commit an attack at a graduation ceremony. He was identified and referred for mental health evaluation and counseling.[102] The fact that he posed a threat in the real world at a certain point effectively invalidates any suggestions of the attack as predominantly underpinned by online influences or confined to the digital space. Additionally, the gun used in the Buffalo attack was legally purchased after Gendron's health evaluation, raising questions regarding oversight by the state, the regulation of firearms, and other online-offline policy concerns.[103] Gendron's online presence and history in the real world should be understood within the broader context in which he inhabits.

Even more confounding in this grand architecture of extremism is the role that algorithms could potentially have played. Gendron's paraphernalia circulated across platforms that are powered by algorithms. At the same time, it is important to make distinctions between different platform designs. While the majority of social media platforms recommend content based on user engagement, 4chan exemplifies one of the outliers. To put this into context, users typically convened on message boards prior to the invention of websites with live feeds. Around 2002, 4chan was one of several platforms that attempted to enliven their boards in a manner that was technically unprecedented with the aim to increase user engagement.[104] The platform currently represents an ephemeral social media with two distinctive features: it rapidly disperses content regardless of popularity and permits absolute anonymity. Similar to Reddit, the platform is a forum-based system with characteristically harmful content that is subjected to little to no moderation.[105]

Notably, 4chan's toxic culture not only engenders the circulation of harmful content but also sustains a culture of impunity that users exploit. 4chan and similar models such as Futaba Channel or 2chan, Dogolachan, and Ilbe Storehouse have inspired terrorist violence in various contexts. In the past, the rhetoric on 4chan has been amplified on Breitbart (an alternative media platform) and optimized on platforms such as Meta. The right-wing media and Republican Party are similarly implicated in the exploitation of anonymity and speed on sites such as 4chan and the appropriation of algorithms on other major platforms. The amplification of harmful content is likely to happen through the diffusion of information between fringe sites such as 4chan and automated mainstream platforms. This

101 Ryan Scrivens, "Examining Online Indicators of Extremism among Violent and Non-Violent Right-Wing Extremists," Terrorism and Political Violence, 2022, 1–21.

102 Johnson, "10 Killed in Buffalo."

103 Shimon Prokupecz et al., "Payton Gendron: What We Know about the Buffalo Supermarket Shooting Suspect - CNN," CNN, May 15, 2022, https://www.cnn.com/2022/05/15/us/payton-gendron-buffalo-shooting-suspect-what-we-know/index.html.

104 Ryan Broderick, "You Can't Always Blame Algorithms," Garbage Day, May 17, 2022, https://www.garbageday.email/p/you-cant-always-blame-algorithms.

105 Benjamin Fedoruk et al., "The Plebeian Algorithm: A Democratic Approach to Censorship and Moderation," JMIR Formative Research 5, no. 12 (2021): e32427.

accordingly prompts the question of whether the internet's worst websites are algorithmic.[106]

3.6 A Postscript on Online-Offline Extremism

Several other parallels can be drawn between the online and offline environments, for instance, references to the online distribution of extreme and violent material in government policy and legislative instruments. This is the case with regard to the physical distribution of such material in the offline environment in various government documents. A legal expert additionally reported that the disproportionate focus on Muslim communities informed the design of CVE programs that were predominantly modeled around the "Muslim experience." This complexion of programming and strategic comportment was until recently largely reflected in online counterterrorism interventions.

Meanwhile, the synergy between the offline and online is continuously guided by common and evolving trends as well. By 2020, far-right extremism had considerably outpaced Islamism and far-left extremism, prompting a reorientation of both technology and government-led efforts to additionally tackle the far-right threat.[107] At the same time, the challenge about the conceptualization of extremism, radicalization, and terrorism has been replicated in the online space and has in some instances reproduced incoherent and more nuanced complications traversing online-offline CVE interventions. Another parallel relates to the role that online and offline communities or networks play in the radicalization process. Social networks, in a similar way to extremism, predate the internet and exist offline.

The relationship between online and offline extremism, while implicitly acknowledged in certain government policies and legislation, remains formally equivocal. The perception among some interviewees that "online" extremism originated in the online sphere and should as such be the primary concern of technology companies is similarly not widely shared. This view appears to have been largely provoked by government frustrations over goodwill deficits and the limited results that their engagements with technology companies have produced. In mediating these perspectives, an official from the Australian government maintains that a threat is realized when a capability and intent align with a vulnerability. The capabilities and vulnerabilities fall within the remits of technology companies while intent rests with users who seek to use platforms to perpetrate harm. Besides highlighting the inadequate efforts of technology companies, the government official's evaluation effectively establishes the link between real-world users and digital platforms.

Overall, frustrations with the performance of technology companies and the conviction that technology companies can do better were common across all groups. Notwithstanding these concerns, there are well-documented successes in programming outcomes in both offline and online

106 Broderick, "You Can't Always Blame Algorithms."

107 Seth G. Jones, Catrina Doxsee, and Nicholas Harrington, "The Escalating Terrorism Problem in the United States," Centre for Strategic and International Studies, 2020, https://www.csis.org/analysis/escalating-terrorismproblem-united-states; Elizabeth Culliford, "Facebook and Tech Giants to Target Attacker Manifestos, Far-Right Militias in Database," Reuters, July 26, 2021, https://www.reuters.com/technology/exclusive-facebook-tech-giants-target-manifestos-militias-database-2021-07-26/; Greg Barton, "ASIO's Language Shift on Terrorism Is a Welcome Acknowledgment of the Power of Words," The Conversation, March 21, 2021, http://theconversation.com/asios-language-shift-on-terrorism-is-a-welcome-acknowledgment-of-the-power-of-words-157400.

domains.[108]

All things considered, the convergence of extremist rhetoric from state and non-state actors, similarly viewed as mutually reinforcing political discourses, is consequential to policymaking and the recalibration of the social and political order.[109] The disjointed efforts between digital platforms and law enforcement agencies with both online and offline mandates therefore compel a more integrated approach to the threat of extremism.

This section has demonstrated that the dynamics of extremism are emblematic of the online environment as not only reflecting the offline space but also in many ways intersecting with it. These intersections and divergences, besides representing potential subjects for future research, also make a strong case for more holistic and integrated programming and policymaking.

Section 4: User and Algorithmic Agency in Context

4.1. Conceptualizations of Algorithmic Amplification

Interviewee representations of the potential role that algorithms play in "online" extremism or the popular notion of "algorithmic amplification" unlocked some insightful perspectives. There was almost universal consensus on the increasing prominence of algorithms in current discourse. In setting the context, a government official duly noted, "The term 'algorithms' has become the latest buzzword and hot topic in the international debates on preventing and countering terrorism and violent extremism online without enough clarity on the concept or the scope." Consequently, some governments expressed an interest in the potential causal relationships between online content and offline actions. Some interviewees appeared to support this proposition by highlighting the incident-driven character and the relegation of context in extremism studies. Moreover, a majority of the interviewees acknowledged the diverse conceptualizations and widespread limited grasp of the concept of algorithmic amplification.

The interpretation of "algorithmic amplification" varied across respondents, ranging from open acknowledgments of lack of understanding to sidestepping the question altogether. Some interviewees were pragmatic and either provided anecdotes or conceived the phrase based on their perceptions of changes observed in the social media environment in recent years. One practitioner linked her conclusions to a pattern of increased visibility of extremist content over time. In validating her experience, she drew insights from the accounts of Frances Haugen and noted that politicians have found the dissemination of online political messaging challenging. As a result, they have had to resort to more divisive speech to increase mass appeal. This prompts a deeper engagement with the interactive layers of causes of conflict or extremism, namely manifestations, triggers, and proximate and structural causes. Causes and actors, discussed in greater depth below, are closely intertwined. The behavior of actors can influence inter-group perceptions and inform group structure, grievances, strategy, goals, and action.

108 "Tech Against Terrorism Annual Report 2020–2021."

109 Rowa, "Liminal Boundaries and Vulnerabilities to Radicalisation in the Context of Securitisation of Migration."

For conflict to spread across geographies and (in this case) for the interplay between online and offline extremism to materialize, it must be underpinned by the convergence of permissive conditions. The primary conditions for conflict diffusion are long-standing grievances and pre-existing inter-group divisions.[110] Permissive conditions or structural causes, according to the conflict transformation framework, congregate personal, relational, and structural factors spanning the political, social, economic, and cultural realms.[111] Digital technologies (among other drivers of extremism) can facilitate the online mobilization of like-minded individuals. They are channels that facilitate intra and inter-group communication and can replicate dynamics in the real world, accelerate the mobilization of issues and actors, and also constitute a precondition.

With respect to the discussion on political actors, should the domain of inquiry encompass the algorithms that potentially amplify divisive political speech, the political actors who appropriate these algorithms, or both? Permissive conditions are the bedrock for catalyzing conflict or extremism into manifest violence. As such, the onset of extremism must be linked to its spatial and temporal evolution and the relevant actors. Put differently, "online" extremism is for the most part the progeny of protracted conflict with roots in the offline world. An appreciation of the intricacies in this space is essential, although the disaggregation of agencies and contextual interactivities complicate the notion of algorithmic amplification. Those dimensions and more nuanced distinctions between online-offline extremism are a matter of future research.

While some interviewees articulated coherent personal understandings, other actors that employed the phrase appeared not to have formally institutionalized the concept. It was therefore difficult to reconcile a commonly used phrase alongside its limited understandings, diverse interpretations, and potential implications for policy. Interestingly, some technology companies similarly reported complexities surrounding their understanding of "algorithmic amplification" and questioned the utility of the concept. They argued that algorithmic amplification is conceptually deficient and could be substituted with more appropriate terminology. In sum, the landmine that is algorithmic amplification brims with divergent connotations and embodies the contradictions of conceptual richness, deficiencies, and ambiguity. Representative of this was one respondent's attempt to describe her interpretation of algorithmic amplification:

> Rather than explain my definition, I guess what I observed is around 2018, we started to notice a very big difference in the way the content that we were producing was getting consumed. We were trying to develop very localized content in communities that had been targeted by ISIS and other groups. ...We don't have any data to back that up, but we sort of did notice that it was a shift that occurred. What we were noticing was that speech that was very toxic was being amplified a lot more. There was a lot more showing up in the feeds of people we were trying to reach in a way that was making it more difficult for us to do the work that we were doing.

110 Michael E. Brown, The International Dimensions of Internal Conflict (Cambridge, MA: MIT Press, 1996).

111 John Lederach, The Little Book of Conflict Transformation (New York: Good Books, 2003).

In recent years, both authoritarian and democratic governments have paid greater attention to social media algorithmic recommendation systems. The focus of proposed interventions has revolved around the algorithmic amplification of extremist, polarizing, misleading, and other forms of harmful online content. While such interventions are necessary, platform regulation can systematically and unwittingly censor unpopular views, undermine independent media, control information flows, influence public opinion, subvert the rule of law, and threaten other institutions of democracy. In 2022, China published the "Internet Information Service Algorithmic Recommendation Management Provisions." The law addresses a range of issues, including algorithm-induced "addiction or excessive consumption," a clampdown on "algorithmic fake news," the prevention of "algorithmic monopoly behavior," and the promotion of "specific protections for the elderly." The law additionally unveils information service norms and stipulates that service providers "vigorously disseminate positive energy." The new regulation has affected big technology companies in China that are powered by algorithms such as ByteDance, WeChat, and Weibo.[112]

Besides China, other governments aiming to address the prevalence of harms in the digital space are gradually training their attention on algorithmic systems as important subjects of legislation. Some proposed laws that target algorithmic amplification include the Justice Against Malicious Algorithms Act of 2021 and the DISCOURSE Act in the U.S. and the DSA in the EU.[113] Meanwhile, the U.K. Online Safety Bill exemplifies proposed laws that refer to algorithms. For its part, the Australian government reported that algorithms will form the scope of its "Basic Online Safety Expectations" with specific determinations on how companies are meeting their safety obligations in the context of algorithms. In addition, the Digital Technology Taskforce under the Department of Prime Minister and Cabinet has been holding public consultations on the regulation of algorithms and automated decision-making.

Due to its gradual diffusion into the legislative domain, algorithmic systems and in particular algorithmic amplification as a concept deserves closer scrutiny. Thorburn, Stray, and Bengani caution against policy maneuvers due to the complex questions that this ambiguous concept presents.[114] Besides algorithmic amplification, the authors contend that human-to-human amplification of content is advanced through such messaging apps as Telegram and WhatsApp. As such, algorithms are not a pre-requisite for amplification (a position is largely consistent with the findings and central argument of this research). While some laws such as the DSA, Justice Against Malicious Algorithms, and the DISCOURSE Act reference the term amplification, none have so far developed a precise definition. The skewed focus on social media additionally downgrades other recommender-driven systems such as news aggregators and audio streaming platforms.

A review of existing definitions reveals three common interpretations of algorithmic amplification. First, amplification as counterfactual presents the question, "amplified relative to what?" To understand amplification in this context, researchers discuss counterfactual scenarios that compare

112 Sara Bundtzen, "What China's Sweeping Algorithm Regulation Means for Digital Governance Globally," Institute for Strategic Dialogue, May 31, 2022, https://www.isdglobal.org/digital_dispatches/chinas-sweeping-algorithm-regulation-and-global-digital-governance/.

113 Luke Thorburn, Jonathan Stray, and Priyanjana Bengani, "What Will 'Amplification' Mean in Court?," Tech Policy Press, May 19, 2022, https://techpolicy.press/what-will-amplification-mean-in-court/.

114 Thorburn, Stray, and Bengani, "What Will 'Amplification' Mean."

the real and imagined spread of information. This interpretation raises further complications relating to methodological limitations, the generalizability of findings, and the aggregation of outcomes (evaluations supported by some of the findings of this research). The implications of algorithm-free social media are additionally tackled in the review. Notably, while purely reverse chronological rankings have been touted as an alternative, they are not necessarily neutral. These rankings typically reward more active accounts, provide no guardrails against spam, and can increase the scale of borderline content. The researchers acknowledge the complexities that stem from the wide latitude for the interpretation of amplification that inhibit the accurate estimation of "what might have been" and actual amplification figures.[115]

The second interpretation conceives algorithmic amplification as (mere) distribution of content. In this conception, the difference between content that is amplified and content that is shown makes the distinction between amplification and distribution difficult. Laws that conceive amplification as distribution and seek to restrict amplification are similarly likely to infringe on freedom of speech and other human rights.

The third interpretation views amplification as lack of agency. It represents the loss of user control or agency over content to which they are exposed and "didn't ask for," or content whose rationale for recommendation is opaque. Ultimately, the question reduces to one of how does the law distinguish between the content that was "asked for" and one that was not?[116]

Further complicating the notion of algorithmic amplification in this study is how the conceptualization process integrated the perspectives of interviewees. The framework additionally incorporated the literature on the Five V's of big data (volume, velocity, variety, veracity, and value) that appear in some sections of this paper.[117] Upon further assessment of the model and related complexities, it is evident that proposing the parameters or an alternative to algorithmic amplification constitutes an important dilemma. This should be the mandate of an interdisciplinary forum, particularly because it is also an area that policy staff in certain technology companies admittedly find complicated. If eventually defined or replaced with another terminology, this concept may still present operational challenges. The table below collates the main interrelated conceptions of algorithmic amplification. In the event the term "algorithmic amplification" loses its currency or is replaced with another concept, the framework below remains pertinent as a representation of the structure of online interactions:

115 Thorburn, Stray, and Bengani, "What Will 'Amplification' Mean."

116 Thorburn, Stray, and Bengani, "What Will 'Amplification' Mean."

117 Hiba Jasim Hadi et al., "Big Data and Five V's Characteristics," International Journal of Advances in Electronics and Computer Science 2, no. 1 (2015), https://www.researchgate.net/profile/Ammar-Hameed-Shnain/publication/332230305_BIG_DATA_AND_FIVE_V%27S_CHARACTERISTICS/links/5ca76bbca6fdcca26d011d6a/BIG-DATA-AND-FIVE-VS-CHARACTERISTICS.pdf?origin=publication_detail; Sepideh Bazzaz Abkenar et al., "Big Data Analytics Meets Social Media: A Systematic Review of Techniques, Open Issues, and Future Directions," Telematics and Informatics 57 (March 1, 2021): 101517, https://doi.org/10.1016/j.tele.2020.101517; Lucia Lushi Chen, Walid Magdy, and Maria K. Wolters, "The Effect of User Psychology on the Content of Social Media Posts: Originality and Transitions Matter," Frontiers in Psychology 11 (2020), https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00526.

Conceptualizations of Algorithmic Amplification

Properties of Algorithmic Amplification	Possible Connotations
Contextual	Amplification as the reflection or enhancement of pre-existing conditions. These conditions can either be algorithm or user-related, or stem from permissive offline contextual conditions. The amplification of specific content can be episodic (for example based on election or other cycles).
Integrative	Potential to facilitate online-offline linkages.
Distributive	Visibility: Raising the prominence of content for more interaction. The reward structure of algorithms, user creation of engaging content, and the gaming of algorithmic systems can increase visibility.
	Volume: The vast amount of data that can be produced every second. The co-creation of content between primary and secondary content creators (for example, influencers and commentors) can increase the depth and breadth of original content.
	Velocity: The rapid generation of data and speed of circulation. Algorithmic systems in concert with users enhance the acceleration of content.
Deterministic	Variety: Various types of data, including structured, unstructured, and semi-structured data like images, videos, and texts. Humans and algorithms determine the variety and visibility of content (or lack thereof). Human input informs algorithmic design and development.
(Dis)Empowering	The conferral and demonstration of agency among users who interact with both constructive and harmful content that are commonly in circulation. Whether or not users exercise their free will in the process, or if the algorithm exercises its own free will, remains an open debate.
Extractive	Value: Valuable information extracted for business and real values of the data. This is in conjunction with content that users find engaging through likes, ratings, shares, and other forms of engagement.
Instrumental	Goal-directed and targets certain outcomes.
Evaluative	Rewards most engaging content, whether subjectively high- or low-quality material.
Attributive	Veracity: Truthfulness of the data analyzed and the accuracy behind any information. Content can be amplified based on subjective attributions of accuracy, significance, relevance, and other values.
Affective	Affective transition: Changes between affective states. Trending extremist or harmful content can elicit or intensify different types of emotions (the same can be said of non-trending content).
Synergistic	Amplification as a collaborative and integrative process aggregates the agency of algorithmic systems, users, content, and background contextual factors.
Productive	The co-creation of online content or productive consumption and the potential for action.

4.2. User – Algorithm Synergies

In setting the context for what follows, a discussion on input and output is apt, calling on earlier discussions on the techno-mediative character of algorithms in the online-offline space. Among other functions, algorithms help to solve computational or real-world problems. The input data for algorithms includes information that is drawn from real contexts. As such, the real-world context is an important component of the algorithmic system. In addition, algorithmic systems are designed, continuously developed, and deployed by humans who similarly assume the role of users. Upon front-end deployment, humans also leverage the affordances of algorithmic decision-making and are at times harmed by it. Humans additionally "game" or exploit algorithmic systems in their advancement of personal, collective, or institutional goals in varying contexts. As a result, it is clear that algorithms are deeply dependent on and impacted by human interactions. Consequently, the analysis of impacts and potential harms should integrate human-machine systems in context. The sections below position and examine user and algorithmic agency within the extremist context.

In order to make determinations on the content to recommend to users, content recommender systems take into account certain signals. In the case of YouTube, these include search and watch histories and patterns (if previously user-enabled). The country and time of day additionally represent the spatial and temporal contexts that model user profiles. Other signals comprise channel subscriptions and the duration of engagement with content. Recommendation systems also administer pop-up surveys and use the feedback to fine tune and improve recommendations for users (these signals are generally overruled by efforts to reduce extremist and borderline content). Some companies consider user empowerment an important aspect of user experience. Users can therefore disable auto-play functions or clear and manipulate their search history, switch to incognito mode, and manipulate other platform functions or tools.[118] In these instances, the continuous interaction between algorithms, users, and context-specific signals is evident.

4.3. An Evolving Compilation of Actors

This section examines the convergence of users and agency in context. The omission, misrepresentation, underestimation, or overestimation of the roles and influence of certain actors can be problematic. Conversely, examining the positions, values, interests, resources, and relationships among actors can significantly enrich analytical and programmatic outcomes. To promote a greater understanding of the extremist ecosystem, an examination of the interface between digital publics and algorithmic systems as communicative digital artifacts is essential. While this section is not exhaustive, it attempts to broaden the analytical lens and enhance the understanding of algorithms in context, highlighting the interactivities among algorithms, actors, and the drivers of extremism as elements that also undergird the broader analytical framework. It is within this context that the concept of users is re-examined.

The attempts to illuminate the role of users in extremism and other online harms are widely regarded

118 "Continuing Our Work to Improve Recommendations on YouTube," YouTube (blog), January 25, 2019, https://blog.youtube/news-and-events/continuing-our-work-to-improve/.

as diversionary, yet often overlooked is the fact that users are not monolithic. Terrorists, authoritarian governments, child abusers, polarizing political figures, tech personnel, and elite cyberbullies are all users. So are the victims of online harms and other well-intentioned internet users. The failure to make these distinctions is not simply a misrepresentation of the versatility of digital actors. It is also the unwitting denigration of the rights of the victims of online harms; the tacit extrication of user agency from the claws of accountability is a manifestation of the digitization of impunity. For example, the U.S. government raised concerns over ongoing regulations that target the wholesale removal of online "terrorist" content, including "terrorist-related," "extremist," or "borderline" content that might lead to terrorism. As a result, laws within its jurisdictions are predisposed toward the litigation of technological tools rather than the nefarious actors who exploit them.

An exploration of the nature and roles of prominent and neglected actors in the "online" extremist ecosystem delimited three overlapping typologies of actors. These include those perceived as transgressors, victims or targets, and intervenors or conciliators. These groups can be further differentiated (among other distinctions) into primary and secondary actors, state and non-state actors, and variations of stakeholders with specific interests in the problem. The role of the three categories of actors is transversal and inhabits the spectrum of beneficence and harm. That being the case, those interviewed did not attribute extremism and other online harms to children while others perceived them as victims or targets.

In dissecting the character of human agency, there seems to be a false dichotomy between good and bad actors as the actions of incidental actors can precipitate unintended consequences.[119] For example, the dissemination of extremist and other harmful online content can spawn conflicting results, with the potential for the sensitization of digital publics or the reproduction and amplification of harmful content. Meanwhile, other users habitually engage in speculative online discourse immediately following violent attacks before the facts surrounding an incident have been established. An internet user can therefore assume varied identities based on specific motivations. Agency in the context of extremism should therefore be assessed and characterized along a spectrum of subjectivity. In some instances, biases in public discourse could (among other corollaries) generate blind spots that increase the threatification of technology in a dynamic that amplifies its risks and diminishes its benefits. As such, a balanced and holistic approach to understanding the role of algorithms in the grand architecture of digital technologies is prudent.

Interestingly, while a majority of the interviewees identified a variety of actors, there was limited engagement with their specific roles. All interviewees provided extensive insights on technology companies at different points in their interviews. However, when specifically asked to identify the relevant or neglected actors in the "online" extremist ecosystem, only two interviewees identified technology companies. This could be interpreted in three ways. One way is to say that actor perceptions are likely confined to the "bad guys" silo of extremists who are embedded at the epicenter of the online extremist ecosystem and must be tackled (the same can be said of algorithms). A second interpretation is that both extremists and algorithms likely dominate analytical and programmatic postures. This may additionally imply that actor mapping and analysis, while not

119 See Bert Jenkins, D. B. Subedi, and Kathy Jenkins, Reconciliation in Conflict-Affected Communities (Singapore: Springer Nature Publication, 2018).

necessarily a strategic institutional orthodoxy, may be integrated into other forms of analysis. Finally, it is possible that the appreciation of the need for comprehensive actor mappings and analyses currently falls short and ought to be addressed. Any potential oversights can create blind spots and unwittingly sustain a cycle of fractured efforts that merely scratch the surface of an otherwise complex and multi-dimensional problem. As a result, a reductivist approach minimizes the sphere of analysis to what is perceptible yet elementary.

The table below outlines the main typologies of visible and invisible or neglected actors that upon further analysis could enhance understanding of extremism in the cyber-physical space.

Actor Mapping

Typology of Actor	Identifications
Transgressors	Extremists: far-right extremists and/or identitarians such as white nationalists and Hindutva ideologues; tribalists such as extremist elements in the Rwandan genocide; religious extremists such as Islamists and Buddhist extremists; anarchists; special-interest extremists such as eco-extremists; hybridized extremists Technology companies: algorithms; digital platform investors and shareholders; tech company executives; tech personnel States and political actors Media: traditional media; alternative media Digital publics: individual internet users; internet groups or communities; New Age adherents; sovereign citizens; the manosphere such as MGTOW; RedPillWomen; WGTOW; cultural intermediaries or influencers (merchants of extremism/borderline opportunists/conspiracy theorists)
Victims or targets	Direct victims: the deceased; survivors of harms from extremist and terrorist attacks who have experienced physical, psychological, and economic trauma Indirect victims: Close relations; witnesses; proximate and far-flung communities At-risk and neglected groups: children; minority groups such as LGBTQ communities, women, and ethnic minorities; religious minorities; people living with a disability; populations with low digital literacy or individuals that are vulnerable to online manipulation; people living in environments prone to extremism or divided societies; New Age adherents; the Global South or non-English-speaking countries; state and non-state humanitarian actors; the broad constituency of digital publics
Intervenors / Conciliators	Governments Technology companies: algorithms; tech personnel; whistle blowers Mainstream media Range of civil society actors Digital publics (concerned)

4.4. Institutional Agency: Algorithms in Corporate Culture

The analyses and interventions on extremism and other online harms have in some instances overlooked the diverse characteristics of technology companies. It is therefore important to consider the particularities in the nature, size, values, products, policies, capacity, governance models, and other company-specific operational architectures. In examining the role of digital platforms in extremism, this section casts a spotlight on the cross-functional influence of technology companies on society at large. Whereas digital platforms have previously advanced themselves as neutral and technologically benign, the impact of their influence on almost every aspect of human life is palpable. Their role in the proliferation of extremist and other harmful content, their implication in election scandals, and complicity in human rights violations are all significations of corporate agency. The paradox of digital instrumentation has involved positive technological interventions in sectors such as health, security, economic, and environmental. Technology companies are therefore versatile cross-functional actors whose influence, through the adoption or preclusion of algorithms, have generated both positive and harmful impacts in the context of extremism. (In a similar fashion, algorithms are increasingly implicated in "algorithmic extremism" yet also provide automated solutions for countering extremism on digital platforms.)[120]

The dominant discourse on the impact of business models on extremism and other harms has largely revolved around big tech. Yet the issue transcends that domain and is better examined within the broader context of corporate culture and governance. The agency of corporate values and principles has the capacity to inspire, activate, and inform corporate practices, or interact with the agencies of other actors and digital artifacts. According to Virginia Dignum, "accountability in AI requires both the function of guiding action (by forming beliefs and making decisions) and the function of explanation (by placing decisions in a broader context and by classifying them along moral values)."[121] An emerging question therefore concerns the extent to which corporate belief systems, goals, and overall organizational cultures constrain or promote accountability. The accountability gap (encapsulating the ostensibly passive stature of corporate ethos) deserves closer scrutiny for its potential to be obscured by the "causality, justice, and compensation" of common contextual factors.[122] Corporate ideology as an important aspect of organizational culture is therefore an active ingredient in algorithmic design and operations. The formation of algorithmic ideology is thus both an intrinsic aspect and product of latent institutional processes that typically interact with other forms of agency.

A majority of the interviewees outside the tech community expressed strong views on the opacity of algorithms and the relationship with company bottom lines. Specifically, some government representatives identified a major knowledge gap regarding the role of algorithms in the amplification of extremist content. They contend that the lack of transparency around user data and algorithmic operations are critical levers for technology companies' business models. A government

120 Ledwich and Zaitsev, "Algorithmic Extremism."

121 Virginia Dignum, "The ART of AI – Accountability, Responsibility, Transparency," Medium (blog), March 4, 2018, https://medium.com/@virginiadignum/the-art-of-ai-accountability-responsibility-transparency-48666ec92ea5.

122 Matt Bartlett, "Solving the AI Accountability Gap: Hold Developers Responsible for Their Creations," Medium (blog), April 5, 2019, https://towardsdatascience.com/solving-the-ai-accountability-gap-dd35698249fe; Rowa, "Part 1."

official bemoaned the increased entrapment of Europeans within technology companies' terms of reference. The promotion of user engagement has influenced both online and offline behavior and is closely tied to the profitability of digital platforms. Many hold that the commodification of user data in conjunction with the dark patterns that have fostered the culture of "buyology" have continued to compromise the free will of technology users. Vigderman and Turner have inventoried the type of user data that big technology companies collect that can be referenced for more information.[123] The governments in question equate the exploitation of user data to the commodification of users, or in the words of a practitioner, "the relegation of users to labels." A few exceptions were noted among technology companies that are already operationalizing safety by design principles (recognized as the cornerstone of any sustainable business model). As one government official remarked, "If it's not safe, people won't stay on their platforms, and over time they will lose business."

For their part, researchers noted that business models and culture vary considerably across platforms. Nevertheless, algorithms were identified as crucial to the business models of big platforms. Any platform that is predicated on or profits from advertising, and therefore relies on the number of users, time spent on the platform, and the nature of engaging content, is more likely to orient its business model around algorithms. Suffice it to say that the broad consensus among researchers and practitioners centered around the AdTech business model or engagement-driven algorithms as more likely to propel business growth and success. Interestingly, an interviewee advanced that the circulation of extremist and borderline content on digital platforms is most likely not a strategic business orthodoxy as is popularly touted. Put differently, the spread of extremist content in the online space could be the result of runaway algorithms. It is a serendipitous development that has become constitutional to some business models. This "windfall" further incentivizes or complicates the arenas of legislative, economic, and algorithmic design. For instance, some algorithms will recommend "awful but lawful" content because the legal system has preconditioned or legitimated these mechanisms.

The researchers further observed that the deployment of algorithms as an operational element in the business models of certain technology companies impacts the overall corporate structure and culture. Power is typically centered around engineering teams because oftentimes companies are founded by computer scientists or people in related disciplines. As companies grow, other imperatives emerge that elicit the addition of sales and marketing, public relations, trust and safety, legal, and other teams. The governance structure therefore tends to incline towards the engineering domain due to the recency of these new functions. This has further implications for the coherence of corporate policies (for instance between the engineering and programming departments). One interviewee remarked that while corporate "ideology is probably too strong a word," it represents the type of approach that computer scientists are likely to adopt. This approach, largely abstracted from impact, is additionally data driven. The preoccupation with scale often minimizes the micro level and external impacts of company operations. This practice has shifted the focus to growth at the exclusion of risks and other important considerations.

Overall, these dynamics underline the necessity for a corporate culture shift. Some legislation has

123 Vigderman and Turner, "The Data Big Tech Companies Have."

attempted to influence this shift through the issuance of safety by design guidelines. At the same time, an interviewee acknowledged the promising but sluggish infusion of human rights values as an important step towards transforming the culture of technology companies. He additionally reflected on recent media reports and suggested that in certain companies, personnel who value and engage with these issues are more often than not marginalized or banished from the system. Meanwhile, the response to whistleblowing sometimes conveys the impression of the prioritization of business over safety concerns. A better understanding of business models therefore demands a deeper engagement with corporate culture and governance. For these reasons, it is possible that a shift in certain aspects of corporate culture and governance could foreground a meaningful reorientation of business models. This realignment may subsequently increase online trust and safety dividends.

4.4.1. Typology of Actors in the Technology Sector

An invisible sub-group within the technology sector is the institution of investors and shareholders that intersect with the internal governance structures of technology companies. In some instances, investors and policymakers can pressure technology companies to grow faster and bigger. Enhanced safety measures are therefore required for such aggressive innovation targets. Other important actors in this arena include company executives and tech personnel who influence company decisions and policies in diverse ways.

A French government official conceived actors as the producers and recipients of extremist and other harmful content. The government representative posited that online content commonly evokes both positive and negative emotional responses that lead to more interactions. The algorithmic systems on some platforms track user engagement with online content to determine the prioritization or visibility of content and other parameters based on user behavior. The link between platforms, their algorithmic systems, and users is particularly important as it denotes reciprocal agency (a topic covered in greater detail in other sections).

The official noted that social media platforms rely on engagement metrics to hook users to their platforms and subsequently increase the likelihood of users interacting with advertisements. In drawing connections between the digital and offline spaces, he contends that (for example) the Will Smith slap was not acceptable in the real world, and therefore the rule of law should apply proportionately in both domains. He emphasized a stronger focus on the effects of online extremism on children, populations with low digital literacy, people who are amenable to online persuasions that are not reflective of the real world, and people living in environments prone to extremism or divided societies.

4.5. User Tactics in the Exploitation of Algorithms

Regular internet users, operating individually or within online communities, emerged as important actors within the extremist ecosystem. While this group embodies a broad array of actors, more specific differentiations have been applied accordingly. The popular discourse on internet users is predominantly techno-determinist and accentuates the subservience of humans to technology.

However, user engagement constitutes one of the driving forces behind algorithmic amplification. The research on user agency, though sparse, has demonstrated the exploitation of algorithms by a subset of internet users.[124] Different types of users leverage algorithms in a variety of ways that include the gaming and circumvention of algorithms to achieve certain outcomes. There are certain differences and similarities in the malicious exploitation of algorithms or AI between violent extremists or terrorists and regular users. The table below identifies tactics that are likely to be appropriated by violent extremists or terrorists.[125]

AI Tactics Likely to be Employed by Violent Extremists or Terrorists

Enhancing Cyber Capabilities	Denial-of-Service Attacks Malware Ransomware Password guessing CAPTCHA breaking Encryption and decryption
Enabling Physical Attacks	Autonomous vehicles Drones with facial recognition Genetically targeted bio-weapons
Providing Means for Financing	Audio deepfakes Crypto trading
Spreading Propaganda and Disinformation	Deepfakes and other manipulated content
Other operational tactics	Surveillance Fake online identities and impersonation Counterfeit documentation Online social engineering

Collated from UNCCT and UNICRI: Algorithms and Terrorism (2021)

In certain instances, users employ a range of linguistic and social strategies to outwit algorithmic systems in order to mitigate against censorship and interpersonal accountability, protect privacy, and provide authentic feedback that can enhance online decision-making.[126] Big technology companies have designed several interventions to tackle terrorist exploitation of their platforms. These efforts have in turn inspired a patchwork of counter-strategies that has enabled maleficent actors to adapt to the evolving online environment. Some terrorist networks have migrated to encrypted and decentralized platforms and opted for self-operated websites, alternative platforms,

124 Samuel C. Woolley and Philip N. Howard, Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media (Oxford, UK: Oxford University Press, 2018); Soomin Kim et al., "Trkic G00gle: Why and How Users Game Translation Algorithms," Proceedings of the ACM on Human-Computer Interaction 5, no. CSCW2 (2021): 1–24.

125 "Algorithms and Terrorism: The Malicious Use of Artificial Intelligence for Terrorist Purposes," United Nations Counter-Terrorism Centre (UNCCT) and United Nations Interregional Crime and Justice Research Institute (UNICRI), 2021, https://www.un.org/counterterrorism/sites/www.un.org.counterterrorism/files/malicious-use-of-ai-uncct-unicri-report-hd.pdf.

126 Kim et al., "Trkic G00gle."

and infrastructure providers.[127] Others have increasingly exploited simple online tools such as pasting, archiving, and file-mirroring sites to sustain their messaging on the web.[128] In addition, the organization of fake networks as a strategy tricks both users' cognitive biases and platform algorithms. These bots flood networks with content and alter the structure of social networks to create an illusion of the dominance of content, perspectives, and political and other figures.[129] In recent years, search engines have developed sophisticated techniques in order to counter the gaming of content popularity signals such as link, troll, or content farms.

Additional tactics that users employ include evasion as exemplified in "for the record" and private browsing sessions that aim to maintain privacy. Strategic crafting includes engaging in online activities such as liking, following, unfollowing, and posting from multiple devices and browsers with the intention to outpace seemingly sluggish algorithmic decision-making. Online self-misrepresentations through engagement with irrelevant content (whether by default or design) are part of data contamination. The use of multiple online identities to "evade" detection is another form of crafting. Tactics such as questioning ad recommendations and pranking intrusive ads and bots are aimed at exposing the vulnerability and limitations of algorithms.[130] A recent study has proposed a user empowerment framework that includes data strikes, data poisoning, and conscious data distribution. The use of ad blockers and applications such as "Go Rando" and other web extensions that can modify search results exemplify data strikes and data poisoning.[131] While useful, such data strikes are double-barrelled arsenals that can be exploited by well-meaning and maleficent actors alike. Taking this into account, how have various actors instrumentalized algorithms in pursuit of nefarious goals within the "online" extremism architecture?

4.5.1. The Co-creation of Content as Productive Consumption

In setting the context for user agency, online content creation constitutes a meaningful entry point. Content creation somewhat delimits the production and influence of online material to the primary creators of content. Yet the role of users who engage with primary content in various ways is just as important. Ancillary content creators (i.e., users), through likes, ratings, shares, comments, and other forms of engagement, can shape a piece of content substantively or reproductively. Their actions are not discrete but are facilitated by online technical tools or functions which they deploy. This mutuality in co-creation can (for instance) amplify or increase the visibility of content or expand the substance of the original content through the comment sections.

Ultimately, these seemingly disconnected information synapses serve to shape and influence online

127 "Content Personalisation."

128 Ali Fisher, "Swarmcast: How Jihadist Networks Maintain a Persistent Online Presence," Perspectives on Terrorism 9, no. 3 (2015): 3–20.

129 Filippo Menczer, "How 'Engagement' Makes You Vulnerable to Manipulation and Misinformation on Social Media," The Conversation, September 20, 2021, http://theconversation.com/how-engagement-makes-you-vulnerable-to-manipulation-and-misinformation-on-social-media-145375.

130 Pamela Pavliscak, "How We Game the Algorithm to Tame the Algorithm," Medium (blog), May 19, 2016, https://medium.com/@paminthelab/how-we-game-the-algorithm-to-tame-the-algorithm-99f287d81a3b.

131 Vincent et al., "Data Leverage."

discourse. The congregation of primary and ancillary actors alongside facilitative technical artifacts is essential to the process of co-creation as productive consumption. The primary and ancillary creators of extremist content are therefore inextricably intertwined. That being the case, some interviewees suggested that counter-extremist interventions should target the primary actors. One researcher additionally submitted that while comment sections may seem extremist, instrumental effects are more constrained by the fact that the majority of people are generally more discursively inclined and less action-oriented. However, offline contextual factors have the potential to change the calculus and trigger action.

4.5.2. Cultural Intermediaries

Of significance particularly within the extremist category is an offshoot of online actors who instrumentalize extremism for economic gains. The paper will engage more with the notion of mercantile extremism in the next section. That said, the commercialization of extremism and other harms has thrived in part due to the unresolved questions around borderline content. One group that overlaps with the online merchants of extremism are cultural intermediaries who aim to influence public opinion, reclaim eroding values, or transform various aspects of society. An EU official stated that online terrorist content is easily identifiable, which is why the EU Internet Forum is more closely monitoring the dominant transmitters of borderline content. This typology of content (exemplified by conspiracy theories) has successfully infiltrated mainstream discourse and stoked public emotions with the potential to foment radicalization and polarization. Opportunistic intermediaries employ skillful content framings that are relatable to their target audiences. They cast themselves as the "Picassos of knowingness" with a theory for everything. In this regard, the EU official reports that they are highly adaptive in as far as they are adept at shifting between subjects. The official remarked on how some COVID conspiracists have recently dabbled in the Russia-Ukraine conflict:

> All these people that spread these messages are somewhat opportunistic – look at what is happening in the current news environment, for example, like Russia and Ukraine. They will take the opportunity to spin some of those conspiracies toward what's happening there. They constantly update these types of messages – for example, we see that there are some channels or online environments that used to spread conspiracies on COVID and now they spread it on Russia. They just switch and their whole objective seems to be to radicalize people, to polarize society... What we are worried about is when it's a concerted effort, when it is really people pushing others to be divided on certain topics. And we see that mainly through conspiracies and disinformation about sensitive matters.
>
> So, it's difficult to define as that's what we're trying to do. You know, what is really harmful? What is something that is borderline or not illegal but harmful? I think trying to classify or categorize that is very challenging, but it's something we need to have a conversation about. Theories such as the Great Replacement, if received online, anywhere, in any shape or form – how can we really justify that it is speed and not speech when we know how it is being used?

In further dissecting the ecology of the diverse expressions of cultural intermediation, borderline opportunism is particularly noteworthy. The borderline actors may display (among other imprints) characteristics of mercantile extremism. This group occupies and exploits the middle ground between terrorists whose online activity has been curtailed and individuals perceived as "normies." In a similar way to borderline content, borderline actors are not accounted for in existing laws. If borderline content is flourishing, these borderline opportunists are evidently thriving off it. A researcher describes one particular group as ideological travelers who act as publishers, networkers, or community builders. They are neither terrorists nor do they promote or conduct terrorist activity. These individuals are for the most part ideological middlemen subsisting between the extremism of the right or left wings of the political spectrum.

Borderline opportunists promote content as a strategy to gain visibility and can be sympathetic to extremists or terrorists without maintaining any formal or direct ties. Spaces such as The Daily Stormer and Red Ice TV are possible representations of the amplified version of Fox News. Supporters often subscribe to these websites and other channels likely to expose them to borderline and extremist content. They typically do not engage in illegal activity but skilfully tinker with legal but harmful content. These actors push the envelope just enough to maintain the remits of legality. They understand the consequences of breaking the law, including the risk of proscription in countries with strict laws on extremism. This group exploits and is unwittingly insulated by the banner of free speech and other freedoms that complicate the policy and regulatory space. Whereas they are accommodated on certain platforms, some technology companies have applied various containment strategies such as quarantine, demonetization, and demotion.

The growing attention on cultural intermediaries and their discursive patterns is necessary. If strong scientific evidence has invalidated the inaccuracies of flat earth, COVID-related, and other conspiracy theories, what then is the threshold for proscription? Should borderline content be labeled, demoted, reframed, or countered? Alternatively, how can the management of such content conform to an environment that upholds fundamental human rights? These are some of the questions that plagued a public official who implicated digital disinhibitionism for the perversion of facts. The online environment boosts information cascade, content visibility, and the mobilization of audiences. This is in contrast to the offline domain where disinformation was previously more likely to be ridiculed and discredited with alacrity. In as much as online cultural intermediaries have demonstrated remarkable agility for knowledge contortionism, some interviewees maintained that public ridicule is not a viable solution to the problem.

4.5.3. Political Actors

Political actors who foment extremism in the online environment were reported as capable of generating greater impact due to their status and reach. There is growing distrust in public institutions whose adoption of double standards has diminished people's trust in these systems. If left unchecked, extremist rhetoric within the political arena can progressively lead to the institutionalization of extremism and ingrain a state of normative strife. Some interviewees expressed that the focus should shift from regular "moms and dads" whose engagement and "virtual reaction" to extremist content is the result of (dis)information diffusion from primary actors. Continued exposure to such material

can lead to desensitization and the normalization of psychological, discursive, and physical violence. Put differently, regular users' engagement with extremist content is merely symptomatic of deep-seated issues that should be tackled at the root. Consequently, measures such as censoring can be counterproductive as is evidenced in the clamor for freedom of speech and the incentivization of online pilgrimages on truth-seeking. Central to this configuration is the role of algorithmic systems as the invisible artifacts that shape the digital platforms that serve as vehicles for the articulation of extremist views, grievances, and injustices., As exemplified in content moderation, algorithmic systems are paradoxically conceived as automated problem solvers within the same sphere. In concert with other agencies, algorithms act as conduits for the reconciliation, remediation, and reconstitution of an increasingly dysfunctional cyber-physical order.

In the political arena, the manipulation of algorithms can happen through computational propaganda, a likely precipitant of extremism when considered alongside other drivers. Computational propaganda is "the use of algorithms, automation, and human curation to purposefully distribute misleading information over social media networks."[132] The structure of algorithmic systems in conjunction with front-end digital platform dynamics complicates the predictability of unintended algorithmic effects for designers and digital publics alike. Political actors typically employ autonomous programs to disseminate political messages ranging from issue-based content to propaganda and (mis/dis)information with the aim of influencing public opinion. The deployment of obscure algorithms and bots that mimic human behavior can generate voluminous data. The outcomes of the mobilization of political bots encompass the amplification of follower numbers, likes, and retweets, the production of fake reviews, the escalation of attacks on opponents, and the suppression of activist discourse.[133] Overall, it can be argued that the deployment of bots in synergy with user agency has implications for the amplification of extremist or borderline content.

4.5.4. Actors in the Traditional and Alternative Media Space

Besides social media platforms, the traditional channels of digital communication such as radio, print, and television that are hosted and accessed online warrant equal attention.[134] One interviewee specifically reflected on the potential for technology to aggravate destructive human instincts, citing how the radio played a marginal but important role in the incitement of the 1994 Rwandan genocide. The recency of that crisis, it seems, has not generated sufficient urgency for decisive action in an increasingly volatile cyber-physical environment. Hallmarks of the Nazi and Hutu extermination machinery such as dehumanization can amplify existing feelings of hate. The deprivation of shared human values and the relegation of human identity to animal and object status can be cataclysmic for unwanted populations when considered alongside other drivers of violence.

A separate interviewee identified dehumanizing language and narratives as a core element of not only genocides but also violent extremism. The respondent repeatedly emphasized the importance of recognizing dehumanization as a key feature of violent extremism. Suffice it to say that among

132 Woolley and Howard, Computational Propaganda, 3

133 Woolley and Howard, Computational Propaganda.

134 Rowa, "Part 1."

other interactivities, hate speech and extremism intersect at the locus of dehumanization. In the long run, the softly-softly approach to borderline content can lead to the normalization of dehumanization if not violence. The mainstreaming of the Great Replacement and other conspiracy theories on "traditional," alternative, and social media should constitute sufficient grounds for governments to address hybrid media structures in the ongoing tech legislation processes. Legislation should be inclusive and adaptive to the particularities of the prevailing media ecology. In the absence of concerted multistakeholder consultations and action, these networks will continue to hide behind the anachronism of traditional media while diverting attention to digital platforms.

4.5.5. Victims of Extremism and Other Neglected Actors

Some respondents identified populations that are neglected and are directly and indirectly impacted by extremism as important actors. They experience direct harms from shortcomings in algorithmic operations and transparency, which they may possess or lack the capacity to interrogate. Their inclusion and participation in policymaking processes are therefore paramount. Meanwhile, most users outside the English-speaking world are often locked out of tech policy discussions. The application of policies targeting adversarial actors in the Global North is largely inconsistent with implementations in the Global South. At the same time, the impact of policies on threats targeting the Global North is often overlooked in the Global South.

Children were additionally identified as neglected actors in the victim and at-risk categories. The Australian government in particular discussed the challenges of reaching young voices who spend considerable time online. Whereas children have not been granted sufficient consideration and focus, their inclusion in legislative and other processes requires the involvement of their parents. In 2021, the eSafety Commissioner in Australia established a user-oriented Youth Advisory Council to strengthen the inclusion and participation of young people in policies relating to online safety. Besides symbolizing good practice, this exemplary model additionally represents an important stakeholder that can enrich ongoing and future GIFCT programs.

4.6. The Convergence of Agencies: Mercantile Extremism in Context

A dimension that is somewhat neglected but is gaining increasing focus is the conundrum of the commercialization of extremism by certain users.[135] In the online extremist ecosystem, users typically design business models that thrive off extremist and borderline content. The strategic positioning of actors within this space accords them reputational influence, popularity, and financial and other forms of capital. The symbiosis between extremist and platform business models, whether by default or design can be conceived along three (and possibly more) intersecting representations of the business of extremism. First, the infractional or incidental business model of extremism is a by-product of lax protocols or user infringement of platform policies and regulations. The unwritten, unwitting, and invasive contract between the merchants of extremism and certain technology companies

135 Cynthia Miller-Idriss, The Extreme Gone Mainstream: Commercialization and Far Right Youth Culture in Germany (Princeton, NJ: Princeton University Press, 2018).

has implications for the diffusion of online extremist content. Secondly, the normative or partnership business model of extremism relates to the synergistic efforts between users and platforms in the normalization of extremism and other harmful cultures. It is characterized by an implicit contract that for the most part sustains a culture of harm and impunity. This dynamic is prevalent on fringe platforms such as 4chan and Gab. Finally, the transactional business model of extremism conceives extremism as an enterprise that is regulated through the use of formal contracts such as the compact between Spotify and Joe Rogan. Overall, these business models present both risks and opportunities for technology companies and the broader digital public that are outside the scope of this study.

The commercialization of extremism, while featuring in a handful of interviews, strongly conveyed a tenor of urgency as an issue that has been widely overlooked and requires intervention at the source. A legal expert specifically conceded, "There are also actors who are doing this (extremism) as a business, putting a lot of time and energy creating and running stories, keeping constant engagement. These are the people we should target and not the moms and dads." These merchants of extremism have devised innovative strategies aimed at securing and broadening their base. The strategies range from branding and marketing, sale of extremist merchandise that could be legal but harmful, networking, fundraising, algorithmic hacks, and even hosting high-voltage online debates.

To put this into context, the Global Disinformation Index and the Southern Poverty Law Center have raised alarm over extremist financing. In 2020, some extremist groups' earnings averaged $1.5 million from leveraging several fundraising tools such as cryptocurrencies that harness various algorithms referred to as hash algorithms.[136] Some groups used DLive, a video streaming platform that allows cryptocurrency donations. Other entities that opted out of bitcoin donations settled for Monero cryptocurrency, which is particularly difficult to trace. Other fundraising tools utilized include social media and the procurement of tax exemption status by registering as charitable organizations. The tax-exempt status provides automatic access to donations from digital platforms.[137] Meanwhile, QAnon was successful on social media because influencers followed the metrics that attracted new audiences and sources of income. Their online postings were determined by content that garnered more followers who powered platform algorithms.[138]

Overall, as with other mainstream actors, the realm of political actors presents technology companies with the dilemma of counter-extremism scope creep. The significant costs associated with the application of platform policies to public figures can trigger a range of unintended effects. The cost of intervention for industry may conversely have instrumental value for public figures. This can occur when public outrage insulates and deflects from the actions of perpetrators to the presumably excessive actions of technology companies. In broadening the context, the current

136 Peter Stone, "US Far-Right Extremists Making Millions via Social Media and Cryptocurrency," The Guardian, March 10, 2021, https://www.theguardian.com/world/2021/mar/10/us-far-right-extremists-millions-social-cryptocurrency; Kunal Dhariwal, "Cryptocurrency Mining Algorithms and Popular Cryptocurrencies," Medium (blog), March 3, 2018, https://medium.com/@Mr.dhariwal/cryptocurrency-mining-algorithms-and-popular-cryptocurrencies-48176d3559d6.

137 Stone, "US Far-Right Extremists."

138 Cleo Chang, "The Unlikely Connection Between Wellness Influencers and the Pro-Trump Rioters," Cosmopolitan, January 12, 2021, https://www.cosmopolitan.com/health-fitness/a35056548/wellness-fitness-influencers-qanon-conspiracy-theories/.

state of polarization, partisanship, and the politics of survival has induced the gradual erosion of the values and ethics that have often guided public service and private practice. The increasingly institutionalized strategic inaction is therefore likely to shift a huge part of the responsibility for speech moderation to technology companies. The question then becomes whether interventions that target more mainstream actors should fall within the purview of institutional affiliations.

These findings further validate popular assertions of a lack of a universal extremist or terrorist profile. In this regard, some researchers reported that their analysis has increasingly pointed towards the hybridization of digital threats. Diverse audiences often commune around the mutually reinforcing and intersecting threats of extremism, disinformation, conspiracy theories, and hate speech. In the COVID environment, conspiracy theory networks connected with other disparate communities. The Institute for Strategic Dialogue and Basit studies in particular demonstrate the symbiotic relationship and divergences between alternative health practitioners, other conspiracy theorists, and extremists.[139] In this particular context, the role of conspiracy theories in extremism should be considered in the existing PCVE architecture. More broadly, rather than silo the analyses and response to threats, a more holistic and integrated approach is essential.

Section 5: Issues Emerging from (Mis)Understandings of Algorithmic Systems

5.1. The Explicability, Interpretability, and Auditability of Black Box Models

The data on algorithmic amplification or the role of algorithms in extremism was subjected to further analysis and sorted along three streams of comprehension. These included: (1) what is currently understood, (2) what is misunderstood, and (3) what is not yet understood about the role of algorithms in extremism. It is remarkable that even as the subject of discussion was initially designed around algorithms, important linkages between algorithms and the broader context subsequently emerged during the interviews. This in itself demonstrates the difficulties of examining algorithmic amplification in a vacuum, moderating algorithm-centric assertions in favor of amplifying their contextuality. That said, the thesis should be subjected to further empirical research.

The meaningful and informed participation of internet users in the design, development, and deployment of algorithms ensures that programs are responsive to user needs, hedge against (un)intended harms, and promote accountability. Meaningful participation is contingent upon meaningful transparency, the absence of which impedes the right to information. Public education on the benefits and challenges of AI has the potential to strengthen participation and promote accountability. Restrictions on access to proprietary data, review of algorithmic operations, sources of training data, and evidence of impact (among other constraints) generate power asymmetries between users and developers and impede participation, accountability, and the formulation of

139 "Anatomy of a Disinformation Empire: Investigating NaturalNews," Institute for Strategic Dialogue, 2020, https://www.isdglobal.org/wp-content/uploads/2021/10/20211013-ISDG-NaturalNews-Briefing.pdf; Abdul Basit, "Conspiracy Theories and Violent Extremism: Similarities, Differences and the Implications," Counter Terrorist Trends and Analyses 13, no. 3 (2021): 1–9.

necessary data protections.[140] The prevailing challenges with black box algorithms therefore compel an additional layer of analysis on their role, impact, and interaction with human agency and broader societal structures.[141]

The common perception of machine learning algorithms as obscure has led to their characterization as "black box" constructs. Pasquale conceives a black box as "a system whose workings are mysterious; we can observe its inputs and outputs, but we cannot tell how one becomes the other."[142] The technical properties that designate the opacity of algorithmic systems include data privacy, changes in data over time, and the interconnectedness of processes and decisions that are learned from data. Another major aspect is complexity as relates to the interconnectedness of algorithms, iterative processing, the scale of data, and randomized tiebreaking.[143] For their part, interviewees identified a range of factors that compound the black box effect. These include the complexity of algorithmic systems in combination with limited algorithmic transparency, the technification of digital technologies, and the self-regulation of technology companies.[144] Additional factors include methodological limitations in algorithm studies, reactionary government regulation, an iceberg approach to government regulation, the multi-dimensional character of extremism, and technological dynamism.

In setting the context, a French government official emphasized the importance of making a distinction between algorithmic explicability, interpretability, and auditability. These are closely linked concepts that are prone to conflation. Explicability denotes the ability or potential to explain in human terms how an AI algorithm works, the algorithmic input and output mechanisms, or the inner workings of an algorithmic system. Interpretability is the ability or potential to understand a model and the procedural mechanisms or underlying basis for decision-making. It is the extent to which cause and effect are observable or determinable, and the degree to which outcomes are predictable when changes in input or other algorithmic parameters occur. In essence, interpretability signifies "being able to discern the mechanics without necessarily knowing why. Explainability is being able to quite literally explain what is happening."[145] Notwithstanding these distinctions, explicability and interpretability are often applied interchangeably.

Auditability is the potential or ability to collect data on algorithmic behavior in order to "ensure that the context and purpose surrounding machine learning applications directly inform evaluations

140 Fukuda-Parr and Gibbons, "Emerging Consensus."

141 "Human Rights in the Age of Artificial Intelligence" Access Now, 2018, https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf.

142 Pasquale, The Black Box Society, 3

143 Ansgar Koene et al., "A Governance Framework for Algorithmic Accountability and Transparency," European Parliamentary Research Service, 2019, https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf.

144 See also Lene Hansen and Helen Nissenbaum, "Digital Disaster, Cyber Security, and the Copenhagen School," International Studies Quarterly 53, no. 4 (2009): 1155–75.

145 Richard Gall, "Machine Learning Explainability vs Interpretability: Two Concepts That Could Help Restore Trust in AI," KDnuggets (blog), 2018, https://www.kdnuggets.com/machine-learning-explainability-vs-interpretability-two-concepts-that-could-help-restore-trust-in-ai.html/; Christian Herzog, "On the Risk of Confusing Interpretability with Explicability," AI and Ethics 2, no. 1 (2022): 219–25.

of their utility and fairness."[146] It is indeed remarkable that the context in which an algorithm is developed, and in particular the corresponding complex societal context within which an algorithm is deployed, is rarely advanced as an important aspect of algorithmic auditing. The literature on the contextuality of audits, while sparse, is additionally highly abstract with a restricted technical focus on bias, fairness, and transparency, while overlooking the role of relevant stakeholders and broader social contexts.[147] Nevertheless, this scholarship has made significant contributions in the field and recalibrated the understanding of the risks and harms of AI, alongside offering potential solutions. The conceptualization of audits and the design of mechanisms for algorithmic auditing must therefore acknowledge the complexity of such an undertaking. Meanwhile, the development of appropriate algorithmic audits remains an open question and the subject of ongoing research.[148] This particular section attempt to address the issues around explicability, interpretability, and auditability in relation to transparency and other emergent areas of inquiry.

A government official conceived algorithmic explicability as the rationalization for the promotion of specific content or advertisement. Auditability does not relate to the recommendation or promotion of content but examines how those actions are linked to wider user networks with varying tastes or profiles. It encapsulates the concept of border effects, which compels greater scrutiny of the border of one subject or element in order to understand the mechanics of algorithmic decision-making. The interviewee further observed that digital platforms are quite keen and generally capable of addressing algorithmic explicability. However, auditability draws attention to the black box conundrum and the limited capacity for platforms to comprehend the complex processes at play. This is rather problematic considering the fact that algorithm black boxes are the "bread and butter" of most platforms. It is possible that the promotion of content is largely driven by algorithms. At the same time, human or users' "psychological bias" demonstrates that platform algorithms mirror human behavior. Human psychological predilections therefore represent an important parameter that complexifies the research on "algorithmic radicalization," as one government official explained:

> If you let the YouTube algorithm play, if you tick the option for the following video to play automatically and just let videos play, you're gonna end up with conspiracy theories. It is not terrorist content but there will be conspiracy theories. What does this mean? There are two options. First, it is because YouTube is a conspiracy platform, which I don't think is the case. I don't think they believe in conspiracies. It could also be the reflection of a human trait, which is that we like to see things that engage us emotionally, we like to see content that is not bland. And we like to see Chris Rock being slapped by Will Smith because it's fun. And

146 Sara Kassir, "Algorithmic Auditing: The Key to Making Machine Learning in the Public Interest" The Business of Government, 2020, 89, https://www.businessofgovernment.org/sites/default/files/Algorithmic%20Auditing.pdf.

147 Brent Daniel Mittelstadt et al., "The Ethics of Algorithms: Mapping the Debate," Big Data & Society 3, no. 2 (2016): 2053951716679679; Andrew D. Selbst et al., "Fairness and Abstraction in Sociotechnical Systems," in Proceedings of the Conference on Fairness, Accountability, and Transparency, 2019, 59–68; Margaret Mitchell et al., "Model Cards for Model Reporting," in Proceedings of the Conference on Fairness, Accountability, and Transparency, 2019, 220–29; Brent Mittelstadt, "Principles Alone Cannot Guarantee Ethical AI," Nature Machine Intelligence 1, no. 11 (2019): 501–7; Inioluwa Deborah Raji et al., "Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing," in Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020, 33–44; Shea Brown, Jovana Davidovic, and Ali Hasan, "The Algorithm Audit: Scoring the Algorithms That Score Us," Big Data & Society, (January 2021), https://doi.org/10.1177/2053951720983865.

148 Brown, Davidovic, and Hasan, "The Algorithm Audit."

> when you push that psychological bias, you want not to trust your government, or you want to believe that there's a grand conspiracy.

A dimension likely to be overlooked is the fact that psychological mechanisms fall within the personal and relational spheres when the conflict transformation framework is applied.[149] The links to the underlying social, political, and economic structures are perhaps more aptly exemplified in the literature on far-right ideology. This ideology promotes a system of social hierarchy, exalts the preservation of racial purity and homogeneity, and opposes globalization, migration, and other liberal values. The adherents additionally amplify anti-elite and establishment sentiments that target individuals who are disillusioned with their personal lives and the political system.[150] These dynamics are further articulated in the tension between real and perceived fears, and their links to broader contextual factors. On this basis, Salmela and von Scheve have advanced compelling arguments on the psychological mechanisms of "ressentiment" and "emotional distancing," and the threat of "declassement" or "precarization."[151] Though elementary, this study provides one possible explanation for the espousal of the Great Replacement theory and the government-related grand conspiracy theories that the French government official alluded to. The dynamics additionally represent the transmutation of the ecology of offline experiences and grievances into digital phenomena.

The complexity of machine learning models may hamper the ability of humans, including the original algorithm developers, to explain the mechanisms and rationale behind the predictive capabilities of the generated models.[152] This is because black box models are directly created from data by an algorithm. The sophistication of the models makes it difficult for designers to understand how the combination of variables leads to certain predictions. The possession of input variables similarly does not increase the capacity for humans to understand the relationships between variables and the underlying mechanisms that influence certain predictions.[153] There are techniques that provide some degree of reasoning but not the reasons behind individualized predictions. They generally operate on an (algorithmic) systemic scale and can provide descriptions of the importance of input variables to algorithmic accuracy during training across multiple individuals.[154]

The complexity of machine learning algorithms renders them incomprehensible except through their inputs and outputs. Besides the initial coding done by humans, the internal processing mechanisms mediating the inputs and outputs are for the most part unexplainable. The 2018 Explainable Machine Learning Challenge represents an illuminating case study for determining the utility, trade-offs, or

149 See Lederach, The Little Book of Conflict Transformation.

150 Rodney P. Carlisle, Encyclopedia of Politics: The Left and the Right, Vol. 2, (Thousand Oaks, CA: SAGE, 2005); John Jack Malone, "Examining the Rise of Right Wing Populist Parties in Western Europe," College of Saint Benedict/Saint John's University, 2014, https://digitalcommons.csbsju.edu/cgi/viewcontent.cgi?article=1044&context=honors_theses; Blake Evan Garcia, "International Migration and Extreme-Right Terrorism" (PhD Thesis, Texas A&M University, 2015), https://oaktrust.library.tamu.edu/handle/1969.1/155240; see also Rowa, "Liminal Boundaries."

151 Mikko Salmela and Christian Von Scheve, "Emotional Roots of Right-Wing Political Populism," Social Science Information 56, no. 4 (2017): 567–95.

152 Michael L. Rich, "Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment," University of Pennsylvania Law Review 164 (2016): 871–929; Deven R. Desai and Joshua A. Kroll, "Trust but Verify: A Guide to Algorithms and the Law," Harv. JL & Tech. 31 (2017): 1–64.

153 Cynthia Rudin and Joanna Radin, "Why Are We Using Black Box Models in AI When We Don't Need to? A Lesson from an Explainable AI Competition," Harvard Data Science Review 1, no. 2 (2019), https://doi.org/10.1162/99608f92.5a8a3a3d.

David Lehr and Paul Ohm, "Playing with the Data: What Legal Scholars Should Learn about Machine Learning," UCDL Rev. 51 (2017): 653–717.

necessity of black box over interpretable models. In recent years, advances in deep learning have promoted the view that the most accurate problem-solving models are intrinsically complicated and uninterpretable. Interpretable models can deliver the technical equivalence of black box models and possibly enhance ethical standards due to their capacity to decipher the relationship between variables and corresponding predictions.

However, the promotion of accuracy over interpretability is misguided. It has sanctioned the trade in proprietary or black box models when interpretable models can perform similar tasks.[155] For example, recall is an important performance metric in the prediction of true positives. Whereas the level of accuracy can be as high as 99.9%, the recall could be as low as zero. Ultimately, algorithmic systems as representations of human decisions and worldviews, and their relationship with extra-algorithmic contextualities, possibly increase understanding but do not resolve the question of the mechanics of "algorithmic amplification." An emerging field variously advanced as Explainable AI (XAI) or Interpretable AI has garnered its own share of criticisms but represents one avenue for the exploration of the underlying inference mechanisms of AI.[156]

The disclosure of algorithm formula or source code is doubly viewed as useful and irrelevant. It is deemed useful in as far as it can enhance algorithmic transparency and improve oversight. Conversely, the inspection of source codes may provide limited information on a computer programs' predictive behavior. This is because decisional rules in machine learning models emerge automatically from the specific data under examination. As a result, the capacity for humans to explain or evaluate algorithmic decision-making could be significantly diminished. The source code may therefore reveal the machine learning method applied without necessarily enhancing understanding of the data-driven decision rule.[157] These sticking points have clear implications for algorithmic transparency.

5.2. The Challenges of Cause, Effect, and the Disaggregation of Agency and Accountability

During the course of the study, important questions emerged pertaining to whether algorithms are inherently beneficent or malign, or whether the input in algorithmic systems has the potential to generate positive or negative outputs. The examination of the character of algorithms therefore presents further questions on cause and effect and the complexities around the disaggregation of agency and accountability. Gluyas and Day submit that the multiplicity of actors engaged with an AI system, encapsulating data providers, designers, manufacturers, programmers, developers, algorithmic systems, users, and other external factors, limits the fair and just assessment and

155 Rudin and Radin, "Why Are We Using Black Box Models."

156 Giulia Vilone and Luca Longo, "Explainable Artificial Intelligence: A Systematic Review," ArXiv, 2020, https://arxiv.org/abs/2006.00093; Dang Minh et al., "Explainable Artificial Intelligence: A Comprehensive Review," Artificial Intelligence Review 55 (2022): 3503–68; Mir Riyanul Islam et al., "A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks," Applied Sciences 12, no. 3 (2022): 1353.

157 Joshua A. Kroll et al., "Accountable Algorithms," University of Pennsylvania Law Review 165 (2017): 633–705.

apportionment of liability for extremism in an integrated cyber-physical space.[158] In the same way, the right to algorithmic explanation presents its own challenges.[159]

Challenges in the disaggregation of agency are informed by the control of the algorithmic systems (by tech companies) relative to input into search or results-based algorithms (by users and tech companies) which train algorithms to generate specific outputs. The attribution of responsibility and control in the generation of a specific outcome within and outside the algorithmic system is therefore noteworthy and elusive. Vulnerability to threats or malign influences similarly confound the fair and efficient allocation of liability and restitution. In this regard, a government official additionally notes that user perceptions of digital platforms as the "Far West" or insular spaces of anonymity and unbridled online impropriety is misguided. User and algorithmic interactivities, though implied in this context, are yet another indication of the challenges of disaggregating the multiplicity of actors and drivers in the online extremist ecosystem. Another factor that limits algorithmic auditability and accountability includes continuous algorithmic modifications. Finally, the unpredictability of algorithmic decision-making and attendant impacts complicate the incentivization of precautionary conduct and other measures that can restrict liability to designers.[160]

5.3. Trust and Limited Access to Research Data

Trust emerged as a dominant theme pitting governments against tech companies, civil society against tech companies, and civil society against governments. Trust was advanced as underlying the challenge of access to platform data with implications for algorithmic transparency, algorithmic auditability, and effective policymaking. Some interviewees suggested that the selective nature of data sharing encourages the embellishment of information while glossing over the challenges that technology companies grapple with. Trust in collaborative partnerships is similarly crucial. Knowledge enclaves were reported as hampering interdisciplinary linkage and learning. Although exemplary collaborative programs exist, the divide between technology and social science disciplines, and schisms within the technology sector itself compounded the fractionation of knowledge and impact. At the same time, civil society organizations have emerged as important interlocutors to existing programs and initiatives. Outside GIFCT and other similar coalitions, the underrepresentation or marginalization of civil society further diminished opportunities for collaboration.

An important line of investigation in this study, a government official observed, is the identity of algorithms, and more specifically the determination of whether they are human or technological creations. The question of whether an algorithm, as a technological or human construct, should be anthropomorphized and conferred legal status is equally consequential. These are confounding interconnected issues that governments are grappling with that can be deciphered through the

158 Lee Gluyas and Stefanie Day, "Artificial Intelligence – Who Is Liable When AI Fails to Perform?," CMS, 2018, https://cms.law/en/gbr/publication/artificial-intelligence-who-is-liable-when-ai-fails-to-perform.

159 Nick Wallace, "EU's Right to Explanation: A Harmful Restriction on Artificial Intelligence," TechZone360, 2017, https://www.techzone360.com/topics/techzone/articles/2017/01/25/429101-eus-right-explanation-harmful-restriction-artificial-intelligence.htm.

160 Yavar Bathaee, "The Artificial Intelligence Black Box and the Failure of Intent and Causation," Harvard Journal of Law & Technology 31 (2018): 889.

testing of theories. The official further noted that one working theory of radicalization foregrounds the role of filter bubbles or spheres of information, yet that theory remains untested due to lack of access to research data. It is vital to assess the systemic risks or effects of algorithmic operations on the mundanities of human lives and broader societal structures. However, access to platform data remains a persistent challenge. It is for these reasons that the DSA has established a transparency framework to guide and enhance oversight, and institute what the government representative describes as "more effective seatbelt and highway regulations."

5.4. Methodological Limitations in Algorithm Studies

All things considered, varied models of scientific inquiry can influence divergences in the methods for querying and testing the role of algorithms in extremism and determining the particularity of results. The development and framing of research questions can greatly enhance the tests that determine the harmful or beneficial character of algorithms. For example, should a specific line of inquiry assess whether algorithmic proactivity or reactivity surfaces specific types of (non)violent extremist content based on specific types of outputs? Put differently, the effect of user input on specific output is an important area of exploration. When user input is considered, "is the input neutral but the output extreme? Or is it that extreme input leads to increased extreme output. This is an important distinction."[161] Alternatively, should the research question probe the role of algorithms in leading users to groups and individuals who are members of violent extremist movements? Then again, how many people actually moved to action as a result of algorithmic decision-making, and what are the causal links? The distinctions among these three questions inform the different approaches in testing methods.

A section of academic and popular literature suggests that algorithms are problematic while neglecting to offer a rationale behind such conclusions.[162] Based on the heterogeneity of user profiles, the same search term can yield different results over time. At the same time, the limitations of reaching certain research conclusions relating to agency, recommendations, and other externalities are often poorly documented. This is exemplified by poor reporting on the difficulties in decoding how a specific input generated a certain output, not to mention the poor acknowledgment of inconclusive research evidence and implications on findings. Additional issues that complicate the research terrain relate to the autonomy of developers in algorithm design, changes to algorithmic systems over time, and the disaggregation of input by multiple actors in algorithmic design and development processes. As such, there are as many limitations in reaching certain conclusions as there are in finding definitive answers.

A broader discussion on what is understood, misunderstood, or not yet understood takes a snapshot of findings from a literature review that is closely interlinked and complementary to this study. The review engages with a corpus of 15 empirical studies examining the role of social media

161 Reviewer, Transparency Working Group Paper, 2022.

162 See for example, Joe Whittaker, "Recommendation Algorithms and Extremist Content: A Review of Empirical Evidence," Global Internet Forum to Counter Terrorism, 2022.

recommendation algorithms and potential links to extremist content.[163] Three of the 15 studies reviewed employ a controlled intervention design. The studies utilize a range of methodologies, the majority of which deploy an "external black box" approach. For the most part, researchers access platform Application Programming Interfaces (API) to examine content that could potentially be recommended. While able to manipulate users, the studies that rely on public datasets are not as proficient in manipulating recommendation systems to observe effects on users. The analysis of input and output data is as such devoid of a clear understanding of underlying systems and operations. As a corollary, the risks related to full disclosure of algorithmic structure and operations – including unfair competitive advantage and malign influence – and the limitations of externally accessed data inhibit full algorithmic transparency.[164]

The choice of units of study, language of content under study, and the research timeline inform interpretations of available evidence on the relationship between algorithms and extremist content. While popular discourse has focused on big tech, the existing literature is predisposed towards YouTube, Twitter, Reddit, Meta, the far-right, and English-language content.[165] The predominance of select big technology companies in academic and popular discourse can generate selective perceptions and other forms of bias while also deflecting attention from and insulating smaller and other big players that do not routinely garner similar scrutiny. The slant towards English-language and Western-centric content suggests the need for greater investments in South-North collaborations. It additionally positions cultural diversity as a bedrock for healthy corporate culture, with broader impacts on human security. Whittaker further notes that some of the studies were conducted in the mid-2010s, prior to the rollout of new platform policies on content downranking, removals, and other measures aimed at preventing, mitigating, and countering extremism.[166] As such, a review of earlier literature should similarly signal the shift from a laissez-faire approach to platform policies that enhance proactivity (likely to have begun circa 2016). Notably, these limitations should be regarded alongside the widely documented factors that constrain the understanding of algorithmic operations and the implications on the conduct of research in the field of tech.[167]

The widespread inconsistencies in the application of "extremism" and related concepts in researching coding systems present another point of inflection. A majority of the studies demonstrate the potential for platforms to recommend extremist content but overlook potential interlinkages with personal and contextual factors. The precision of certain codifications on extremism that

163 See Whittaker, "Recommendation Algorithms."

164 Alistair Knott et al., "Responsible AI for Social Media Governance: A Proposed Collaborative Method for Studying the Effects of Social Media Recommender Systems on Users," The Global Partnership on Artificial Intelligence, 2021, https://gpai.ai/projects/responsible-ai/social-media-governance/responsible-ai-for-social-media-governance.pdf; Tal Zarsky, "The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making," Science, Technology, & Human Values 41, no. 1 (2016): 118–32; Whittaker, "Recommendation Algorithms."

165 Whittaker, "Recommendation Algorithms."

166 Whittaker, "Recommendation Algorithms."

167 Michael Kearns and Aaron Roth, "Ethical Algorithm Design Should Guide Technology Regulation," Brookings (blog), January 13, 2020, https://www.brookings.edu/research/ethical-algorithm-design-should-guide-technology-regulation/; Andrew Tutt, "An FDA for Algorithms," Admin. L. Rev. 69 (2016): 83; Jenna Burrell, "How the Machine Thinks: Understanding Opacity in Machine Learning Algorithms," Big Data & Society (January-June 2016): 1-12. doi: 10.1177/2053951715622512; Mittelstadt et al., "The Ethics of Algorithms"; Luciano Floridi, The Fourth Revolution: How the Infosphere Is Reshaping Human Reality (Oxford, UK: Oxford University Press, 2014).

preference the coding of online channels over content are equally noteworthy. The former system of coding could ambiguate or mischaracterize a piece of content as extremist when it is essentially thematically broad-based. The classification of online channels based on existing literature or online databases without the qualification of extreme content represents another point of concern.[168]

Besides the amorphousness and volatility of extremist groups, tensions exist between the internal perceptions of group members and their characterization by the academic and media communities.[169] The perceptions, constructions, and counter-associations with extremism between in-out groups similarly have a bearing on extremism as a concept. This degree of conceptual charitability, besides generating ambiguities, also pays inadequate attention to the malleability of constructs relative to the dynamism of contexts. In fact, the contemplation of the fluidity of liberalism and conservatism in Jost et al. and Rowa's dissection of the contextual dynamics of ideological transmutations are instructive of both the rigidity and (perhaps to a lesser extent) the fluidity of the margins of extremism.[170]

The challenge of "online" extremism has provoked important discussions that relate to the liability of digital actors and the legality of online content. There are concerns that extreme online content, though largely ill-defined, leads to (non)violent extremism. Yet the empirical evidence on the role of digital technologies in shaping information consumption and corresponding effects is inconclusive.[171] There are three points to consider in this regard. First, content-sharing algorithms do not primarily amplify terrorist and violent extremist content (TVEC) material but can disseminate content that could fuel extremism. This type of "extremist" or "dangerous" content that is the subject of intense policy debate on algorithmic amplification may not be codified as "illegal" in international, regional, or national laws.[172] That said, the legality of the content under study further complicates the existing conceptual issues. For example, variances in the statutory determination of extremism across legal jurisdictions influence digital platforms' obligations for content removal.[173]

The findings on the state of methodological rigor call for further reflections on the user question. For instance, some of the studies indeterminately account for personalization due to the anonymity of test accounts, the simulation of personalization, or because of material disregard for personalization, yet make inferences regarding radicalization.[174] Nevertheless, the studies that examine user behavior tentatively highlight some of the complexities of integrating user dynamics. It therefore seems

168 Whittaker, "Recommendation Algorithms."

169 Manoel Horta Ribeiro et al., "Auditing Radicalization Pathways on YouTube," in Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2019, 131–41.

170 John T. Jost, Christopher M. Federico, and Jaime L. Napier, "Political Ideology: Its Structure, Functions, and Elective Affinities," Annual Review of Psychology 60, no. 1 (2009): 307–37; Rowa, "Liminal Boundaries."

171 Brent Kitchens, Steven L. Johnson, and Peter Gray, "Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption," MIS Quarterly 44, no. 4 (2020): 1619–49.

172 GIFCT CAPPI Working Group, "Part 1: Content-Sharing Algorithms & Processes," Global Internet Forum to Counter Terrorism, 2021, https://gifct.org/wp-content/uploads/2021/07/GIFCT-CAPII-2021.pdf.

173 Whittaker, "Recommendation Algorithms."

174 Whittaker, "Recommendation Algorithms." The studies include Ribeiro et al., "Auditing Radicalization Pathways"; Kostantinos Papadamou et al., "'It Is Just a Flu': Assessing the Effect of Watch History on YouTube's Pseudoscientific Video Recommendations," in Proceedings of the International AAAI Conference on Web and Social Media, vol. 16, 2020, 723–34; and Gaudette et al., "Upvoting Extremism."

reasonable to presume that inadequate attention to user choices could inadvertently cast them as bereft of both agency and lived experiences. Finally, the meta-review reported mixed findings on algorithmic recommendations of extremist content and user choices. These range from positive effects and mixed outcomes to minimal-to-no effect. Studies should therefore distinguish between the examination of whether recommender systems actually cause harm vis-à-vis whether these systems direct users to potentially harmful content.[175]

5.5. The Complexity of the Extremist Ecosystem and Technological Dynamism

The complex, multi-dimensional, dynamic, and interactive character of "online" extremism was cited as compounding the algorithm black box effect. The dynamism of artificial intelligence was cited as undergirding the overlap of knowledge gaps and limiting the understanding of the impact of algorithms on society. Meanwhile, terrorists and extremists continue to evolve and adapt to changing online and offline environments. Taken together, these issues present broader ramifications for effective policymaking and the design of appropriate and sustainable interventions. Some respondents submitted that while extremism pre-dated the internet, the enabling environment that digital platforms have created amplifies extremist content in a manner that is unprecedented. One disillusioned practioner remarked:

> I think the biggest thing that's been on my mind is how do you legislate and intervene in a fairly holistic way, because extremism and the online is so big and so complicated. And if you introduce one law, and introduce one algorithmic change, you can have unintended consequences... How do you start to operationalize this?... it won't get addressed in my lifetime and I'm very conscious of that. It will probably be better incrementally.

Another layer of complexity in the extremist ecosystem relates to user agency. Whittaker asserts that "for extremist materials to be recommended (and for researchers to study this using open-source methods) it must be available on social media platforms."[176] In advancing this train of thought, for extremist content to be available on digital platforms, the content was most likely created by humans, uploaded online by humans, on platforms designed by humans, that disseminate content consumed by both humans and algorithms, in a dynamic of mutual enrichment. Users in fact employ specific tactics to game or optimize algorithmic functionality while (semi)autonomous algorithms capture data from human digital trails to enhance personalization.[177]

Following from this, the question of whether content is pre-selected or user-selected warrants further appraisal. The mechanics of pre-selection or user selection is a fairly knotty question when the nature of input and output is considered, and particularly when algorithms in part rely on users' digital trails or browsing histories for personalization and recommendation. What informs the pre-selection of content? To what extent can legal systems, the academic community, policymakers, and other actors

175 Whittaker, "Recommendation Algorithms."

176 Whittaker, "Recommendation Algorithms," 19.

177 Fisher, "Swarmcast."

establish clear demarcations between pre- and user-selected content? Product conception, design, deployment, continuous refinements, and user consumption convene a multiplicity of needs, values, and goals that potentially inform personalization. Whittaker validates this view by submitting that algorithms influence user-choice to some degree (for example, in the "recommended for you" content on YouTube and Twitter's personalized timeline).[178] Crucial to these assessments are the needs, values, goals, and other codes that intersperse with the social, political, economic, environmental, and other contextual factors that comprise human security.

Section 6: Implications of Algorithmic Grips and Gaps for Policy and Practice

6.1. Impediments to Effective Policymaking for Governments

Previous sections have highlighted more comprehensive technology and algorithm-specific bills and legislation, including implications of algorithm-centric regulations. This section examines good practices and gaps in the existing and ongoing tech legislation processes while establishing relevant linkages with algorithms. More importantly, in considering the findings on what is understood, misunderstood, and not yet understood about algorithms, what are the implications for policy and practice? In this regard, this section additionally highlights what is misunderstood and the dilemmas and complexities around algorithmic systems and online harms. What are the likely impacts of proposed and existing legislation on industry policy and practice? How do government and tech industry policies impact the user context? In addressing these questions, the study acknowledges that policymaking in the real world can be somewhat disordered. The prevalence of innovative models for effective policymaking, while notable, cannot fully reconcile the complexities of the process.

6.2. Conceptual Ambiguities

Overall, interviewee contributions gravitated heavily towards impediments to good practice in policymaking over prevailing facilitative influences. The perspectives additionally addressed the links between policy and practice within both government and industry. Conceptual ambiguity emerged as a dominant theme afflicting both government and technology companies. Definitional inconsistencies inhibit the effective design, enforcement, and industry compliance with laws on online content. The most common problematic concepts included "terrorism" and related terms such as "extremism" and "radicalization." At the meso level, the varied conceptualizations or framings of terrorism and other harms, as adaptations of bureaucratic incoherencies, dictate the translation of platform policies into practice. This occurs within a context that applies well-intentioned (and, in some instances, minimal to inconsistent) contextual discretion in the enforcement of policies that are complicated by scale.

Additional conceptual obscurities relate to algorithmic amplification and other legal terminology such as the application of motive and intent. The conflation of motive and intent and the widely debated

178 Whittaker, "Recommendation Algorithms."

relationship between the two concepts was reported as complicating proof of motive. A challenge related to intent that cuts across the online-offline domain pertains to the treatment of individuals who have expressed an intent to commit a terrorist attack and the perception of threat on the basis of an openly expressed threat. Subsequently, a judge can establish religious cause as intent and inflate the actions of a defendant. A legal expert noted that these actions may for example delineate "something that is inconsequential for a Muslim as consequential because they are Muslim."

6.3. Low Institutional Capacity

Governments' legislative muscles are significantly weakened by low institutional capacity. In this regard, the U.S. government official acknowledged the limited understanding of algorithms in policy circles. More specifically, policymakers lack a clear understanding of algorithmic operations in various business processes. The Australian government shares these views and additionally stated that algorithms are completely opaque to governments because they are commissioned in confidence and owned by private companies. The "monumental opacity" of algorithmic operations imposes limits on policymakers' understanding of the problem and hinders effective legislation. The discourse on online extremism should therefore revolve around algorithms.

The U.S. government official further identified four key areas of algorithmic decision-making that are poorly understood and as such constrain effective legislation. The first point relates to the limited understanding of the mechanisms surrounding algorithmic amplification of TVEC. Variances in platform and algorithm design and other platform-specific characteristics further widen the knowledge gap on how recommender algorithms function. Secondly, the restrictions on access to platform data limit understanding of the types of content that platform recommender systems amplify and demote. Information on the targets of these interventions is additionally anonymized. Thirdly, the official highlighted the limited knowledge of the role and extent to which machine learning and more broadly artificial intelligence can address the amplification of TVEC. This point somewhat accentuates and complements other interviewee positions that companies should engage more openly with the challenges in the field.

The U.S. government official further acknowledged the capacity for automated technologies to support in the identification of TVEC. However, they similarly held the position that these technologies lack the capabilities for the responsible and reliable removal of content in the absence of human review and oversight. A fourth capacity gap relates to the extent to which algorithmic amplification of harmful content leads to offline violence and other harms. U.S. government officials noted that recent studies have shown a correlation between "subversive online activity" and susceptibility to certain harmful propaganda. In addition, data has shown a decrease in the period within which individuals radicalize and mobilize to violence. However, more evidence is required to specifically demonstrate the mechanisms and extent to which algorithms and the algorithmic amplification of content can contribute to offline behavior. A French government official provided an illustration of the issue:

> It's the difference between, when I'm on Facebook and I like football, I'm gonna have ads on trainers basically. So that's understandable, right? And

> because I like soccer and pizza, they're going to offer me an advertisement on something completely unrelated, like going out to the cinema and the link is not understandable to humans. It's understandable to the machine because it has been tested so many times and knows that people that like soccer and also like pizza often go to the movies. This cannot be demystified by humans. And I think this is one of the core problems for the legislator is that we need to understand how an algorithm works or how it was trained so that we can design better policies.

Researchers pointed out that the algorithm capacity gap extends to experts in the field, including journalists, researchers, and other members of civil society. While these professionals are authorities in their fields, they tend to have a limited grasp of the technicalities of algorithms. In spite of this gap, they often participate in various forums that influence policy agenda and outcomes. That said, these individuals are often involved in policy discussions for wide-ranging reasons beyond their technical capabilities. This could represent a key concern for other stakeholder forums (e.g., GIFCT stakeholder format) that assemble individuals from diverse backgrounds with complementary knowledge and skills. That notwithstanding, certain forums on policy, trust, and safety are populated with tech professionals with a glaring exclusion of individuals from the fields of peacebuilding, PCVE, and subject matter experts on other online harms. Consequently, communities both within and outside specific disciplines remain siloed when an interdisciplinary and holistic approach could significantly minimize existing capacity gaps.

6.4. Inadequacies in the Evidence Base

The literature on the politics of policymaking is at best evidence-informed rather than evidence-based. Evidence-informed policymaking is guided by three factors. The first aspect relates to building consensus on the definition of evidence and the ability to identify sound and relevant research. Second, the vast body of evidence can overwhelm policymakers, driving them to rely on cognitive and organizational shortcuts to process sufficient evidence to inform decisions. And third, unpredictable policy environments impede systematized policy cycles. The limited control that policymakers command in dynamic contexts implies that respective policymaking arenas will rely on their own networks, apply distinct formal and informal rules, and formulate dominant ways to define policy problems.[179]

It is within this backdrop that policymakers' understanding of extremism in the "online" space informs unwavering policy positions. The nature, relevance, rigor, and objectivity of the evidence base that potentially informs legislation are often not clear. One government official noted the "exponential growth of online harms" that has been matched by a sluggish response from technology companies. Additional concerns from another interviewee arose over this study's excessive focus on legislative

179 Ruth Mayne et al., "Using Evidence to Influence Policy: Oxfam's Experience," Palgrave Communications 4, no. 1 (2018): 1–10; Sandra M. Nutley, Isabel Walter, and Huw T. O. Davies, Using Evidence: How Research Can Inform Public Services (Bristol, UK: Policy Press, 2007); Paul Cairney and Christopher M. Weible, "The New Policy Sciences: Combining the Cognitive Science of Choice, Multiple Theories of Context, and Basic and Applied Analysis," Policy Sciences 50, no. 4 (December 1, 2017): 619–27, https://doi.org/10.1007/s11077-017-9304-2.

gaps when the problem was "clearly" inaction from industry. The attempt to pivot responsibility from technology companies coupled with demands for more transparency from governments were viewed as disregarding the central role of algorithms in online extremism. While these views deserve serious consideration, the growing body of conflicting evidence and debates on algorithmic radicalization and amplification both complicate policy discussions and represent potential entry points for consensus building on policy sticking points. The debate between Haidt and Lewis-Kraus on the existing evidence and complexities of social media harms is particularly illuminating.[180]

6.5. Contextual Disparities and Inconsistencies in the Application of Policies

An EU representative observes that Counter-terrorism Directive 2017/451, by virtue of being merely a directive, is inconsistently applied by member states. This is in contrast to regulations that must be applied by member states as stipulated. To illustrate the problem, a 2017 article that required member states to remove or block access to content that promotes the propagation of terrorist offenses revealed discrepancies in its application. Countries such as the U.K. and Germany's extraordinary NetzDG are perceived to have exceeded their mandate whereas other states underperformed. It is for these reasons that an EU-wide regulation was introduced. These conceptual and operational inconsistencies in turn have broader ramifications for tech sector compliance.

In a similar vein, the convergence of conceptual and legislative disparities at the country level has considerable influence on collective action in global governance. For example, the concept of "violent extremism" is applicable in the United States as in other jurisdictions. However, extremism without violence is generally permissible under U.S. law. This does not invalidate the commitment to civil rights and liberties, including the freedom of expression and privacy. In general terms, the First Amendment precludes the U.S. government from prohibiting speech or language except to the extent that such speech or language is found to constitute child pornography, a threat to use force or violence, or other limited exceptions. The government further notes that this is consistent with the position that the United States has adopted in relation to international law. Committing to the Christchurch Call presents countries an opportunity to tackle the threat posed by online TVEC while remaining zealous advocates for the freedom of expression. All things considered, contextual disparities have broader implications for technology companies. The inconsistencies in legal interpretations of extremism and related terms, and the application of laws relating to extremism as equally problematic, are discussed in greater detail in other sections.

6.6. Policies as Instruments of State Repression

The paradox of digital instrumentation compels closer scrutiny of the chronic tensions between the constructive and malign use of digital artifacts by the state. Birnhack and Elkin-Koren submit that the state's involvement in the information environment became more prominent post-9/11 and is also

180 Jonathan Haidt, "Why the Past 10 Years of American Life Have Been Uniquely Stupid," The Atlantic, April 11, 2022, https://www.theatlantic.com/magazine/archive/2022/05/social-media-democracy-trust-babel/629369/; Gideon Lewis-Kraus, "How Harmful Is Social Media?," The New Yorker, June 3, 2022, https://www.newyorker.com/culture/annals-of-inquiry/we-know-less-about-social-media-than-we-think.

reflected in the legislation passed on terrorism and cyberspace.[181] The reconceptualization of the digital environment as a space that facilitates both productive engagements and terrorist activities has progressively remodeled the internet as a domain for intelligence operations. The proliferation of algorithmic techniques in the collection of intelligence data has political and ethical implications for security via algorithms in the 21st century.[182]

The capacity for various approaches to enhance security in one context while weakening conditions in another is a crucial hallmark of counterterrorism architecture.[183] The track record for good practice in tech policy design and implementation is highly variable across governments. Whereas legislation can strengthen the rule of law and curtail the abuse of executive power, laws as instruments that undermine democratic institutions are similarly conceivable in certain jurisdictions. The U.S. government expressed concern over the precedent being set for authoritarian as well as burgeoning democracies. The governments of Nigeria and India have in recent times compelled technology companies to bend to their will.[184] The failure of industry to comply with government requests could induce the threat of stiff penalties such as fines or imprisonment. The removal of content deemed as "extremist" but which in actuality constitutes free speech (typically associated with political opposition) represents a point of inflection. Some governments are already applying these justifications. These responses (including the promulgation of potentially conflicting national regulations) raise additional concerns under the constitutional protections for freedom of speech, and international obligations and commitments to human rights such as freedom of expression.

In this regard, the laws in some African countries have spurred the deployment of digital surveillance tools and the fabrication of terrorism charges against civilians deemed as threats to the state.[185] In Uganda, the novelist Kakwenza Rukirabashaija was charged with "insulting President Museveni and Commander of the Land Forces of UPDF Muhoozi Kainerugaba via his Twitter handle" and charged under "Computer Misuse Act for the use of offensive language."[186] Although digital platforms facilitate the advocacy of rights and other issues, they are inversely de facto state surveillance tools. Besides the treatment of digital platforms as rich evidence repositories for state repression, tech legislation as a tool for suppression is equally noteworthy.

Since 2014, the Indian government has made significant investments in e-governance, aimed at the

181 Birnhack and Elkin-Koren, "The Invisible Handshake."

182 Louise Amoore and Rita Raley, "Securing with Algorithms: Knowledge, Decision, Sovereignty," Security Dialogue 48, no. 1 (February 1, 2017): 3–10, https://doi.org/10.1177/0967010616680753.

183 Jolly and Ray, "The Human Security Framework."

184 Emmanuel Akinwotu, "Nigeria Suspends Twitter Access after President's Tweet Was Deleted," The Guardian, June 4, 2021, https://www.theguardian.com/world/2021/jun/04/nigeria-suspends-twitter-after-presidents-tweet-was-deleted; Rana Ayyub and Courtenay Werleman, "Govt Asks Twitter to 'take down' Freedom House's Tweets," June 29, 2022, https://www.telegraphindia.com/india/twitter-doc-shows-government-requests-for-blocking-tweets-of-some-advocacy-groups-politicians/cid/1872112.

185 Isabel Linzer, "Digital Technology Helps Governments Target Critics Across Borders," Slate, February 24, 2021, https://slate.com/technology/2021/02/paul-rusesabagina-rwanda-trial-digital-technology-critics-abroad.html; Stephanie Kirchgaessner and Jason Burke, "Rwanda Dissidents Suspect Paul Rusesabagina Was Under Surveillance," The Guardian, September 3, 2020, https://www.theguardian.com/world/2020/sep/03/rwanda-dissidents-suspect-paul-rusesabagina-was-under-surveillance.

186 "Ugandan Intelligence Confirm Kakwenza Rukirabashaija Is Rwandan Agent," Kampala Post, January 4, 2022, https://kampalapost.com/content/ugandan-intelligence-confirm-kakwenza-rukirabashaija-rwandan-agent.

improvement of service delivery, deepening civic engagement, and subsequently strengthening democratic governance.[187] Yet the Hindutva ideology, an extremely chauvinistic nationalist doctrine that has largely been shaped by social media and other cultural practices, continues to thrive with the tacit and open support of the state. This mosaic of technological appropriation is further punctuated by the convergence of diverse motivations and agents, including digital platform volunteers and commercial interests that harness social media analytics and bots.[188]

The mutually reinforcing tactics of state repression through online surveillance and the expropriation of digital policies constitute gross violations of human rights. The correlation between counterterrorism discourse and practices, systematic human rights infractions, and backlash effects on extremism has been widely studied.[189] At the same time, the capacity for online public discourse to embolden the legitimation of extraordinary government policies signals the convergence of agencies in the co-construction of counterterrorism and tech legislation architectures.[190] The digitization of state repression therefore has important ramifications for extremism, counterterrorism, and tech policies.

There are further portrayals of the interlinkages between the state and the information environment, often in concert with political minions and other actors. The deployment of algorithms in political decision-making and campaigns can influence the knowledge, attitudes, and behavior of digital publics.[191] Of note is the Cambridge Analytica interference in the Kenyan and U.S. elections with the collusion of corporate and political machines, which informed Facebook's decision to moderate its algorithms to downrank content that was rated as false.[192] It is equally important to underline the role of political figures within and outside government institutions in stoking "online" extremism, often interspersed with offline spaces.[193]

The agency of internet users represents an important element in the overall architecture of security. The positioning of individuals as the dominant referents in HS does not occlude their links to state security. In fact, the most practical applications of HS acknowledge the nexus between the HS

187 Puneet Kumar, Dharminder Kumar, and Narendra Kumar, "E-Governance in India: Definitions, Challenges and Solutions," International Journal of Computer Applications 101, no. 16 (2014), https://ssrn.com/abstract=2501127.

188 Sahana Udupa, "Enterprise Hindutva and Social Media in Urban India," Contemporary South Asia 26, no. 4 (2018): 453–67.

189 Anurug Chakma, "Does State Repression Stimulate Terrorism? A Panel Data Analysis on South Asia," Journal of Policing, Intelligence and Counter Terrorism 17, no. 2 (2021), https://www.tandfonline.com/doi/full/10.1080/18335330.2021.2022184; Javier Argomaniz and Alberto Vidal-Diez, "Examining Deterrence and Backlash Effects in Counter-Terrorism: The Case of ETA," Terrorism and Political Violence 27, no. 1 (January 1, 2015): 160–81, https://doi.org/10.1080/09546553.2014.975648; A. Carl LeVan, "Sectarian Rebellions in Post-Transition Nigeria Compared," Journal of Intervention and Statebuilding 7, no. 3 (2013): 335–52.

190 Ariane Bogain, "Understanding Public Constructions of Counter-Terrorism: An Analysis of Online Comments During the State of Emergency in France (2015-2017)," Critical Studies on Terrorism 13, no. 4 (October 1, 2020): 591–615, https://doi.org/10.1080/17539153.2020.1810976.

191 Ujué Agudo and Helena Matute, "The Influence of Algorithms on Political and Dating Decisions," PLOS ONE 16, no. 4 (April 21, 2021): e0249454, https://doi.org/10.1371/journal.pone.0249454; Alessandro Bessi and Emilio Ferrara, "Social bots distort the 2016 U.S. Presidential election online discussion," First Monday 21, no. 11 (November 7, 2016), http://dx.doi.org/10.5210/fm.v21i11.7090.

192 "Facebook's Algorithm: A Major Threat to Public Health," Avaaz, August 19, 2020, https://avaazimages.avaaz.org/facebook_threat_health.pdf.

193 Manuela Caiani and Linda Parenti, "The Dark Side of the Web: Italian Right-Wing Extremist Groups and the Internet," South European Society and Politics 14, no. 3 (September 1, 2009): 273–94, https://doi.org/10.1080/13608740903342491; Lisa Visentin, "NSW Election 2019: Labor's Michael Daley Claims Foreigners Taking Young People's Jobs," The Sydney Morning Herald, March 18, 2019, https://www.smh.com.au/nsw-election-2019/michael-daley-claims-foreigners-taking-young-people-s-jobs-20190318-p51591.html.

paradigm and the stability of the state.[194] Since human and national security is closely intertwined, threats to personal and group security are likely to threaten the security of the nation-state. As such, a people-centered approach to security is complementary to national security. This synergy further suggests that the devaluation of contextuality could potentially undermine security not just within but also between states, as is the case with transnational extremism.[195]

6.7. The Burden of Limited Public Participation and Technical Capacity on Enforceability

The GDPR (2016/679) gives individuals the right to challenge and request a review of automated decision-making that adversely affects their legal rights. However, these rights cannot be promoted and protected in the absence of robust public awareness campaigns. This being the case, public communication and engagement as a strategy for promoting digital literacy is variously cited among potentially viable solutions for influencing public behavior.[196] In fact, existing scholarship has identified participation alongside other key principles that are particularly instrumental in the implementation of human rights.[197] That said, interviewees observed that the opacity of AI or algorithmic systems limit the ability for rightsholders to challenge automated decision-making. User agency further intersects with unresolved algorithm black box challenges when the accountability of designers and the capacity to contest their decisions is diminished. It therefore follows that the capacity for enforceability or compliance and the preconditions for mounting successful challenges on algorithmic decision-making still represent important gaps in existing policy architectures.

The policy disconnect between governments and digital publics is palpable in some instances. In considering the ongoing public discourse about technology companies, the presumption that public rage will translate into greater enthusiasm for policy interventions is highly appealing. After all, users' sharing of personal data creates a digital trail that progressively refines their online profiles and influences the capability of algorithms. This puts the right to privacy and data protection at risk, rendering the latter almost impossible.[198] As part of the complexities of policies on technology, a study conducted by Pew Research reveals some variance between the public and American government's appetite to litigate social media companies over user-generated content, with 56% of Americans opposing the move.[199] Besides the disengagement of users in mainstream policymaking and poor critical digital literacy skills, the discordance between technology companies, governments, and publics on the particularities and feasibility of policy is likely to precipitate policy incoherence.

194 "Human Development Report 1994."

195 "Human Security Handbook"; Hans-Jakob Schindler, "Emerging Challenges for Combating the Financing of Terrorism in the European Union: Financing of Violent Right-Wing Extremism and Misuse of New Technologies," Global Affairs 7, no. 5 (2021): 795–812.

196 Cooley et al., "Influencing Public Behavior."

197 Fukuda-Parr and Gibbons, "Emerging Consensus."

198 "Human Rights in the Age of Artificial Intelligence."

199 Colleen McClain, "56% of Americans Oppose the Right to Sue Social Media Companies for What Users Post," Pew Research Center (blog), July 1, 2021, https://www.pewresearch.org/fact-tank/2021/07/01/56-of-americans-oppose-the-right-to-sue-social-media-companies-for-what-users-post/.

6.8. The Reductivist and Reactionary Character of Policymaking

The iceberg approach describes a situation in which interventions target the elements of a problem that are visible or obvious to the actors while overlooking the underlying mass of ice in which a problem is rooted. An interviewee reported that Canada has introduced some bills that would be disastrous by virtue of being reactionary.[200] In addition, these bills address the symptoms rather than the underlying problems of extremism and other online harms. This is in part due to the low capacity of government institutions to make sense of black box algorithms. However, in recent times, Canada has displayed some encouraging signs and intends to re-examine its widely criticized framework for online regulation. The government has clearly listened to feedback and instituted a consultative process to guide the redesign of the regulatory framework on online harms. Moreover, the process will incorporate the feedback received during the national consultation process.[201] Canada is evidently an outlier among countries that have often forged ahead with legislation even in the face of strong criticism.

6.9. The Politicization of Regulatory Processes

The interplay between the politicization of regulatory processes and hard-line public discourses, pitting the threat of extremism against fundamental freedoms, represents an important impediment to effective legislation. To begin with, political discourse was cited as an important driver of extremism yet polarizing political figures have garnered limited government scrutiny and policy intervention. Some members of civil society who were interviewed saw terrorist designations as highly politicized processes aimed at the advancement of "geo-political interests as opposed to the protection of the state." While progress was noted in recent times, some designation lists were interpreted as discriminatory due to the underrepresentation of far-right groups. In addition, a practitioner noted that international humanitarian law is not typically applied in situations in which designated terrorist organizations such as Hamas engage in service delivery. The goal of Hamas, it was argued, is to resist Israeli occupation, an act that is permissible under international humanitarian law. Consequently, the interviewee recommended that the application of human rights should be people-driven as opposed to state-centric and consider the impact of occupations on conflict-affected populations such as the Palestinians in Gaza.

6.10. The Unintended Consequences of Tech Legislation

In complicating the terrain of policy, the regulation of algorithms could present some legal challenges. Transparency laws and policies are often hailed as economical and minimalistic regulatory instruments in comparison to direct regulation. A social media company can apply both codified and ad hoc policies (the latter executed in response to crises and other unforeseen circumstances). While

200 See also Daphne Keller, "Five Big Problems with Canada's Proposed Regulatory Framework for 'Harmful Online Content,'" Tech Policy Press, August 31, 2021, https://techpolicy.press/five-big-problems-with-canadas-proposed-regulatory-framework-for-harmful-online-content/.

201 "The Government's Commitment to Address Online Safety," Government of Canada, July 8, 2022, https://www.canada.ca/en/canadian-heritage/campaigns/harmful-online-content.html.

the momentum for mandatory editorial transparency has increased, the potential for adverse effects must be assessed. Mandatory transparency can increase regulators' powers over online speech and motivate the enactment of more counterproductive speech laws.[202] Gautam Hans additionally notes that "Any algorithmic regulation is probably going to have some First Amendment questions, if not problems... I think tech companies would feel fairly confident they have a good hand to play if they end up in court."[203]

Section 7: Impediments to Effective and Sustainable Interventions for Technology Companies

7.1. Fragmented Learning and Sharing Culture

Some interviewees expressed concern over the learning culture within the technology sector. They could not determine the extent to which industry has embraced the culture of the systematic collation and public dissemination of operational challenges and lessons learned. More importantly, it was not clear to what extent the learning cycle is aimed at and contributes to the enrichment or reorientation of ongoing interventions. Most of the big technology companies do publish scientific papers on a range of subjects.[204] They additionally provide briefings and highlight illuminating success stories on their websites. While industry engagement with challenges and lessons learned appeared negligible in certain forums, it is also possible that this is a practice that is already ingrained in the DNA of internal corporate culture.

Nevertheless, limited engagement with these issues is a drawback that if applicable to a majority of technology companies could diminish the prospects for requisite collaborative programs. This particular gap constitutes a point of divergence with offline actors who have a long tradition and experience in monitoring, evaluation, and learning cycles, including public engagements on similar issues. As such, limited public engagement on operational challenges may additionally lower opportunities for capacity strengthening from government, civil society actors, and other digital publics. If gradually fostered, a robust learning and sharing culture has the potential to increase trust, the lack of which is a creeping problem among the various stakeholders.

7.2. The Reductivist and Reactionary Character of Interventions

Some civil society interviewees decried the "hit and run" approach to the threat of extremism on digital platforms. An Australian government official broadened this thinking by drawing attention to the policy gap emerging from the existing technology architecture. Technology companies are viewed as more inclined to prioritize the patching of existing vulnerabilities over the mainstreaming of safety by design protocols. Suffice it to say that the complexity of the drivers and temper of

202 Eric Goldman, "The Constitutionality of Mandating Editorial Transparency," Hastings Law Journal 73 (2022).

203 Julia Zorthian, "Washington Wants to Regulate Facebook's Algorithm. That Might Be Unconstitutional," Time, October 13, 2021, https://time.com/6106643/facebook-algorithm-regulation-legal-challenge/.

204 See for example Huszár et al., "Algorithmic Amplification."

"online" extremism has invariably fostered piecemeal strategies. This is particularly pertinent when the analytical and intervention frameworks have strongly gravitated towards content-centric approaches, particularly within the online environment.

7.2.1. The Shortcomings of Content Moderation

In broadening the analysis on content-centric approaches, it is important to incorporate the nature and role of users. Some researchers reported that platforms have increased the capacity for networking through the creation of vast webs of channels with somewhat similar themes or subjects of interest. Within these vast networks, a small number of users disseminate and complement the amplification of the bulk of extremist content. Meanwhile, the majority of users are likely to engage in extremist or hateful messaging on fewer occasions. Whether these dynamics are conceived as a pyramid, concentric circles, or another representative form, there is often a small, hardcore subset of users who drive extremist content. The smaller cohort is generally composed of influential or charismatic offline and online figures. The scope broadens with the inclusion of micro-influencers who are valued in their communities but are less well-known. The margin further widens when other users who are fairly interested in specific content but are not quite as proactive join the fray.

This three-tiered representation of the typology of users leads the researchers to interrogate the rationale behind content moderation. Digital platforms, they contend, excessively focus on content takedowns while limiting their response to networks with content that violates their policies. They further observe that a networked approach to tackling the coordinators of behavior that violates policies has not been seriously considered or applied in the realm of extremist networks. In short, the removal of extremist content has therefore not addressed the problem of the persistent networks of accounts that intermittently promote extremist content. In addition, there is limited investment in restricting the ownership of multiple accounts and the creation of new ones after a takedown under similar user names and details.

The cycles of content takedowns are indeed crucial in the short term. However, any approach that overlooks the role of smaller segments of hardcore users is incapable of generating sustainable results in the long term. In considering the perspectives of interviewees, should interventions focus on content, individual users, user networks, or consider a multi-pronged course of action? What factors are likely to limit or facilitate the capacity for technology companies to expand their sphere of impact? In a similar vein, should group recommender systems (for example) garner further attention? A provisional thesis that should be subjected to empirical research would maintain that content recommender systems and the demotion, promotion, or removal of extremist content are less impactful in comparison to group recommender systems and attendant interventions. Framing the algorithmic question as such expands the field of inquiry to subsume users as subjectivities, groups, or networks and how they relate (or not) to extremist content.

7.2.2. Poor or Restricted Responsiveness

Some interviewees reported that upon flagging extremist content, certain technology companies

typically resort to the removal of content of concern with minimal (if any) changes to their systems. One respondent said, "We do not want this to be a long-term relationship and I don't want to spend my life collecting extremist content. It is also very toxic... We need to focus on the pages that are mainly responsible for creating these communities and environment." These spaces are viewed as hot spots that are likely to configure a pathway and additionally provide internet subjectivities a sense of community or collective purpose. Gendron, the Buffalo attacker, stated that he consumed a lot of extremist content from the notorious 4chan board, /pol/ (politically incorrect). It is within this online community that he was first exposed to Brenton Tarrant's manifesto and the livestream video. He also "found other fighters, like Patrick Crucius, Anders Breivek, Dylan Roof, and John Earnest," and subsequently felt "awakened" and decided to emulate the actions of his predecessors.[205] This is again a case where the role of online communities and the policy conundrum relating to content-based versus group recommender systems comes into sharp focus.

A legal expert further reported that when lists of platform pages and users or networks were submitted for assessment and intervention, some platforms often reported that their investigations concluded that the pages or networks did not pose any threat.[206] The expert additionally remarked that further inquiry on assessment protocols typically led to dead ends, with informants duly advised that the information requested was not subject to public disclosure. Third-party exposure of problematic pages and networks with the use of the same tools developed by technology companies additionally appears to do little to galvanize any meaningful action. While these dynamics reveal some limitations of technical or tooling systems, they additionally present the question of goodwill that is more reflective of and interconnected with broader institutional structures and culture. Ultimately, these institutional cultures may inadvertently generate misconceptions about platform priorities as placing business over public safety. This mode of engagement can chip away public trust and diminish the currency of gradual gains in transparency standards.

7.2.3. The Role of Digital Territoriality in Restricting Impact

Some practitioners and government officials were particularly concerned about cross-platform user behavior and the fragmentary approach to tackling extremism within the online space. The failure to take responsibility for harms outside individual platforms, they argued, could debilitate the impact of their interventions. As a case in point, increasingly stricter content moderation standards within mainstream platforms have unwittingly generated dividends for fringe platforms. Social networks with conflicting operational ethos such as Gab, MeWe, and Telegram have adopted a decent proportion of mainstream digital orphans. While this represents an important gain for the previously affected platforms, the online displacement of extremism not only signifies the shortcomings of current approaches but also highlights the associated constraints of the optimization of impact and sustainability. Nevertheless, the platforms that are already committed to taking action cannot mitigate the problem of inaction from companies that have shunned a collective approach to tackling online harms. An Australian government official viewed the deflection of responsibility from industry to government and the fragile spirit of collective responsibility as additional impediments

205 Johnson, "10 Killed in Buffalo."

206 In this particular case, the legality of the content, and the context of reported content were not clear from the interview.

to effective policymaking. The digital territorialization of interventions therefore necessitates reemphasizing more effective, collaborative, and holistic approaches.

7.3. The Bane of Borderline Content

The content that is deemed legal but with the potential to harm represents another dimension of complexity. A surprising finding revealed that a section of technology companies and bureaucrats were largely in consensus on borderline content. They view most borderline content as illegal and expressed frustration over the lack of decisive government action. Certain technology companies have expressed a willingness to act on this content pending government direction. All things considered, borderline content could be consequential to the extremist landscape by virtue of straddling and exploiting the ambiguities and affordances of the "legal but harmful" construct.

The strategic inaction on borderline content does not appear to have fully accounted for the experiences of existing victims and risks to potential victims of extremism. Moreover, inaction by governments and platforms may not have considered the long-term impact of this type of content on societies. Nevertheless, borderline content is a knotty domain that presents some dilemmas for technology companies and governments. The subjectivity of interpretations presents the question of what constitutes the crossover threshold from "awful but lawful" to illegal harms. The treatment of borderline content as benign may overlook the interplay between borderline, extremist, and other harmful online content. The dilemma of whether the threat to freedom of (hate) speech supersedes fundamental human rights and vice versa is noteworthy.

There is content that platforms view as harmful and illegal yet it is categorized as borderline. In such situations, what is generally termed as borderline or grey zone content may not violate digital platforms' terms of service.[207] While the visibility of this typology of content is widely restricted on platforms, the question of how to respond effectively remains.[208] If users create content determined to be legal and other users actively search and engage with the content, should private technology companies exert absolute control over the curation and restriction of legal but "extremist" content? The concerns over borderline content do not eclipse the overarching debate on the definition of extremist content, liability for content creation (user or platform-generated), and the dispersal of content across digital publics (within hybridized or algorithmically amplified systems).

Among the contentious issues between policymakers and digital platforms are company policies on grey zone content. Some platforms do not recommend what they have classified as borderline content but actively reduce its visibility. This does not imply that the content cannot be discovered. A tech company representative established an important link with users by indicating that "you can still find the content if you are looking for it. If you have the exact URL or something else, you can still find it." These types of content are considered legal and therefore do not violate company policies. That notwithstanding, companies employ various counter-strategies to reduce the likelihood of

207 "Working Group on Infodemics: Policy Framework," Forum on Information & Democracy, November, 2020, https://informationdemocracy.org/wp-content/uploads/2020/11/ForumID_Report-on-infodemics_101120.pdf.

208 GIFCT CAPPI Working Group, "Part 1: Content-Sharing Algorithms & Processes."

users interacting with potentially harmful content in the borderline category. Such measures, viewed as intrinsic to the broader safety by design philosophy, include upranking authoritative sources. All things considered, policymakers typically do not understand why demoted extremist content that is simultaneously borderline is still discoverable online. For their part, some technology companies have consistently expressed the willingness to act on borderline content pending government stewardship.

Both policymakers and technology companies acknowledge the challenges of developing a taxonomy for online terrorist content. There are instances when researchers may interpret a piece of content as extremist and actually get recommendations for this type of content in the course of their experiment. However, the distinction between borderline and extremist content is a highly subjective process. The problem of conceptual inconsistencies led a technology company representative to call for clearer distinctions: "We have to draw the line. We have to draw the line somewhere." The room for technology companies to maneuver is constrained by governments' hesitancy to provide clear policy directions on the remits of borderline content. This compels companies to step in and draw their own lines, resulting in stakeholder appraisals of company actions ranging between excessive and suboptimal performance. These mixed evaluations effectively thrust companies into operational limbo when efforts to address the demands of one group generate dissatisfaction in another group.

An additional complication around borderline content relates to the global scale of industry operations and the internet. The ubiquity of the internet ensures the free circulation of harmful content, particularly in the absence of a common purpose and the atomization of efforts among technology companies. GIFCT is one among several other exemplary and evolving models of intervention that are gradually transforming this complex landscape. Additional complications of scale include the constant delicate balance between online safety and freedom of expression, responsibility, the relativity of human rights, disparities in national and regional legislation, conceptual ambiguities, stakeholder perceptions and competing interests, political partisanship on tech-related issues, and cultural or contextual heterogeneities. However disparate, the law overrides the majority of the complications that technology companies must contend with, and the fulfillment of legal obligations often takes precedence.

Section 8: Meaningful Transparency

8.1. Meaningful Transparency in the Space of Human Rights and Ethics

While the provision of security is the primary responsibility of the state, corporate entities similarly represent important security actors. Technology companies in particular apply important norms related to human rights and good governance that can positively transform security dynamics, influence other actors, or cause (un)intended harms. Industry therefore plays an important role in online security. As a result, the Human Security Business Partnerships is an important initiative that aims to apply the key principles of human rights, transparency, justice, and equity to strengthen collaborative alliances, understand the impact of business operations, and establish common goals.[209]

209 "Human Security: An Approach and Methodology for Business Contributions to Peace and Sustainable Development," London School of Economics, n.d., https://www.lse.ac.uk/ideas/Assets/Documents/project-docs/un-at-lse/LSE-IDEAS-Human-Security-Background.pdf.

Technology companies have adopted diverse approaches to guide the expression and promotion of values and moral obligations related to their operations. The dominant concepts that inform such initiatives incorporate either singular or integrated approaches to human rights, ethics, and transparency, and compel closer scrutiny. Human rights are commonly advanced as more universal and well established in comparison to ethical principles, which are in some instances "negotiable" with the potential for circumvention. Human rights therefore constitute a more robust framework due to their legitimacy as an international set of norms that parties have consented to and are enforceable through domestic and global mechanisms.[210]

Interviewees were in consensus on human rights promotion as the primary responsibility of technology companies, governments, and civil society actors. Whereas these actors are the principal duty bearers, the findings on digital rights holders place equal emphasis on user accountability for human rights. Beyond the realm of responsibility, the rights of individuals are moderated by personal, interpersonal, and institutional constraints and mediations. The prevention of violent extremism is an obligation in the Universal Declaration of Human Rights and other international instruments. At the same time, the violation of human rights represents an important driver of violent extremism while the promotion of rights is an essential component of interventions against violent extremism. The temper of online interactions in the context of rights is as such pivotal to the mitigation and promotion of extremism.

The findings have demonstrated the versatility of users across platforms and their capacity to flourish under the banner of anonymity while exploiting the protections of human rights. One technology company observed that the masking of identities facilitates the circulation of extremist content and the abuse of other internet users. That said, the risk and existing capabilities for identity exposure identify this issue as more a matter of pseudonymity as opposed to anonymity. The prevailing discourse on anonymity is skewed towards the protection of human rights defenders and other minority groups. However, this discourse neglects the victims of online abuse, some of whom have lost their lives due to the violence perpetrated by incels and other types of extremists. This dynamic additionally resonates with the dominant perspectives on companies taking responsibility for the harms on their platforms, yet anonymity significantly promotes cross-platform delinquency among users. The transparency of digital identity in the context of rights therefore remains a knotty issue.

Ethical principles are focused on moral guidelines and are neither binding nor do they constitute legal norms.[211] In some instances, human rights defenders have noted the lack of universality of ethical frameworks. While the divergence of opinion on ethical subjectivism or relativity further complexify the design of model ethical standards, similar debates on universality and cultural relativism have

210 Fukuda-Parr and Gibbons, "Emerging Consensus."

211 Amy Lehr, "A Human Rights-Based Approach to AI," Center for Strategic & International Studies (podcast), 2019, https://www.csis.org/podcasts/humanity-wired/human-rights-based-approach-ai; "Ad Hoc Expert Group (AHEG) for the Preparation of a Draft Text of a Recommendation on the Ethics of Artificial Intelligence," United Nations Educational, Scientific and Cultural Organisation (UNESCO), 2020, https://unesdoc.unesco.org/ark:/48223/pf0000373199?posInSet=2&queryId=N-EXPLORE-64abc70f-4002-4d27-a7ac-ad15d3645536.

featured in the field of human rights.[212] The universality of human rights principles and standards therefore does not preclude circumventions or variations in their application in diverse contexts.[213]

Ethical principles and practices often account for human rights while human rights can shape the benchmarks for ethics. At the same time, the unethical use of AI is likely to violate human rights.[214] Nevertheless, there are certain questions that AI raises that can be assessed within the rubric of ethical frameworks rather than the human rights system. The potential also exists for ethical principles to inform the design and implementation of regulatory and policy frameworks. They are useful in the absence of clear norms or when technological development has outpaced policy response.[215] As such, the current codification of human rights in national, regional, and international legal frameworks is technologically suboptimal. Similar to technology, the human rights framework is dynamic and subject to continuous appraisal.[216]

Suffice it to say that human rights law emerged in the pre-digital age and as a work in progress has not anticipated or addressed the full range of rights related to evolutionary AI.[217] The concepts of human rights and ethics are therefore complementary in as much as human rights law is well established, universal, operational, and consequently commands legitimacy. However, human rights account for remedy and restitution and address power imbalances. Ultimately, an important difference between a human rights and ethics approach lies in their engagement with accountability and power asymmetries.[218]

Against this background, the discourse on ethics within technology companies has revolved around "good" and "bad" AI, and understandably so. After all, ethical concepts related to transparency, accountability, fairness, and justice have significantly enriched the ongoing debates on responsible AI and guided the principles and practices of technology companies.[219] Several tech companies have rolled out voluntary programs on "ethical practices" to address concerns about the harmful and unintended effects of digital technologies. Yet a study by Fukuda-Parr and Gibbons reveals that the ethical frameworks that some companies adopt are not comprehensively grounded on international human rights law. Moreover, the doctrine of "ethical AI" predominantly focuses on "transparency" and falls short regarding accountability standards while neglecting enforcement and participation. Nevertheless, the interrelated principles of transparency and accountability, which are elemental to effective participation, can enhance the promotion of human rights. It therefore follows that access to information and user participation can generate greater accountability and other positive

212 Fukuda-Parr and Gibbons, "Emerging Consensus"; Ulf Johansson Dahre, "Searching for a Middle Ground: Anthropologists and the Debate on the Universalism and the Cultural Relativism of Human Rights," The International Journal of Human Rights 21, no. 5 (2017): 611–28.

213 Fukuda-Parr and Gibbons, "Emerging Consensus."

214 "Human Rights in the Age of Artificial Intelligence."

215 "Ad Hoc Expert Group"; Peter Asaro, "A Review of Private Sector AI Principles: A Report Prepared for UNIDIR" (Geneva: UN Institute for Disarmament Research, 2019).

216 Fukuda-Parr and Gibbons, "Emerging Consensus."

217 Lehr, "A Human Rights-Based Approach"; Fukuda-Parr and Gibbons, "Emerging Consensus."

218 Fukuda-Parr and Gibbons, "Emerging Consensus."

219 "Human Rights in the Age of Artificial Intelligence."

outcomes.[220]

In some instances, the symbolic invocation of human rights obfuscates weak enforceability and accountability standards and shifts focus on the singular notion of the right to privacy. This could potentially reinforce a piecemeal approach to human rights. In contrast, companies whose approach is grounded on international human rights law are more likely to operationalize them. Consequently, voluntary ethical guidelines have forged de facto norms and reconceptualized human rights into what would ordinarily be considered "ethical" practice. This study calls for further scrutiny of the integration of human rights standards and principles into existing ethical frameworks. Companies that fall short are essentially "'ethics branding' themselves as committed to human rights"[221] or aim to "forestall the introduction of legally binding legislation."[222]

Other reviews of ethical guidelines have identified recurrent themes on ethical frameworks that while varied in framing appear to overlap. Besides the vague articulation of some principles, these frameworks are further characterized by the absence (or weak enforcement) of mechanisms. In addition, there are important variances in the interpretation of the principles, the reasons they are considered important, the issues, context, and actors they address, and the mechanisms for operationalization.[223] Overall, corporate ethical frameworks that lack a robust commitment to accountability and the related principles of transparency and participation are problematic.[224]

8.2. Algorithmic Transparency and Beyond

The findings so far present the question of the adequacy and effectiveness of current approaches to algorithmic transparency in addressing extremism in the cyber-physical space. Another question pertains to the extent to which decision-making on algorithms (and ultimately algorithmic transparency) is linked to the overall structure and functions of institutions (including product design). The institutionalization of transparency or lack thereof can potentially influence the operational and reporting aspects of algorithmic transparency. Beyond the specifications of algorithmic and institutional transparency, this section demonstrates the import of transparency standards at the level of interagency interactions. Third-party access to user data and the integrity of data partnerships (among other sticking points) necessitate collaborative transparency on relevant benchmarks. The New Zealand government has set the pace with its recent publication of a transparency report on "Digital Violent Extremism."[225] That said, the concept of "digital violent extremism" remains

220 Fukuda-Parr and Gibbons, "Emerging Consensus."

221 Fukuda-Parr and Gibbons, "Emerging Consensus.", 32

222 Fukuda-Parr and Gibbons, "Emerging Consensus.", 42

223 Fukuda-Parr and Gibbons, "Emerging Consensus"; see also Anna Jobin, Marcello Ienca, and Effy Vayena, "The Global Landscape of AI Ethics Guidelines," Nature Machine Intelligence 1, no. 9 (2019): 389–99; Jessica Fjeld et al., "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI," SSRN Scholarly Paper, January 15, 2020, https://doi.org/10.2139/ssrn.3518482.

224 Fukuda-Parr and Gibbons, "Emerging Consensus."

225 "2021 Digital Violent Extremism Transparency Report," Department of Internal Affairs, New Zealand Government, March, 2022, https://www.dia.govt.nz/diawebsite.nsf/Files/Countering-violent-extremism-online/$file/DVE-Transparency-Report-2021-a.pdf; see also "Tech Against Terrorism's Assessment of New Zealand's Digital Violent Extremism Report," Tech Against Terrorism, May 18, 2022, https://www.techagainstterrorism.org/2022/05/18/tech-against-terrorisms-assessment-of-new-zealands-digital-violent-extremism-report/.

controversial for reasons outlined in this report. Ultimately, issues around feasibility are likely to arise and present further dilemmas and opportunities for continued adaptation.

In considering the foregoing issues, it is important to note that the increasing use of algorithms to shape, guide, and make decisions has elicited calls for universality, transparency, accountability, and equity. A better understanding of algorithmic decision-making requires information on the data used, the models guiding assumptions, and the practices that developers employ. The knowledge generated, Heather Roff argues, will improve the accuracy of the likely effects of algorithmic decision-making and more coherent policies.[226] The existing proposals aimed at promoting algorithmic transparency and accountability include both broad and specific measures: "awareness raising: education, watchdogs and whistle blowers; accountability in public-sector use of algorithmic decision-making; regulatory oversight and legal liability; global coordination for algorithmic governance… algorithmic impact assessments; an algorithmic transparency standard; counterfactual explanations; local interpretable model-agnostic explanations (LIME)," among others.[227]

The current discourse on transparency is more inclined towards technology companies. Yet the findings suggest that transparency regimes are applicable in other sectors that interact with industry. For some interviewees, meaningful transparency represents the power of users to request information from both technology companies and governments. Both entities bear responsibility for full disclosure of user data. At the same time, users should have more control over their data as well as exercise the right to own their data. Other parameters of meaningful transparency that were cited included increased third-party access to platform data, types of data used in training, how data is structured, cleaned, and screened for bias, and the incentives that guide these processes.

Data transparency on human rights encompasses the people whose rights were most violated as a result of policy changes, the proportion of false positives on human rights content, and other unintended consequences of tech legislation and enforcement. This data is valuable in its capacity to highlight lessons learned and areas of improvement. The requirements for transparency and accountability for operational systems and processes were also highly recommended. Their strong links to the highly acclaimed safety by design protocols could mitigate some of the issues related to broader institutional cultures. Interviewees further pointed out that while some technology companies have requested legislative interventions and encouraged researchers and journalists to identify vulnerabilities in their systems, industry has been fairly unreceptive to third-party insertions.

A U.S. government official acknowledged the numerous calls for digital platforms to increase transparency on human and automated content moderation decision-making and practices. In the United States, transparency efforts have so far occurred within multistakeholder forums in a voluntary capacity. A practitioner additionally suggested that multistakeholder mechanisms such as GIFCT have the potential to bolster transparency within technology companies. However, transparency reporting is largely voluntary and segregated along specific online harms that are potentially interconnected. In addition, the low capacity and sluggish response within sections of the technology sector are major

226 Roff, Advancing Human Security.

227 Roff, Advancing Human Security; Rowena Rodrigues, "Legal and Human Rights Issues of AI: Gaps, Challenges and Vulnerabilities," Journal of Responsible Technology 4 (December, 2020): 100005, 2

drawbacks to transparency efforts. The practitioner further remarked that the complexity around algorithms implies that even "very smart people don't know what to do."

The calls for increased transparency from governments did not particularly garner widespread support. Governments expect increased transparency from technology companies on the challenges they face and corresponding areas of development. They argue that increased access to information will greatly enhance policymaking processes. A government official further noted that there is ongoing debate on the existence of clear and sufficient public accounting within multistakeholder forums. The prevailing discourse raises additional questions on the output of these forums and the nature of data requested from platforms. More importantly, the meaningful impact of specific actions taken by digital platforms in response to these voluntary initiatives represents another inflection point. The discourse additionally revolves around the level of accessibility of algorithmic data and processes (particularly for independent researchers or auditors).

Technology company representatives discussed the challenges related to algorithmic transparency. The degree of feasibility and misconceptions of transparency as a broad-spectrum solution to prevailing problems are among the thorny issues. Whereas algorithmic transparency is regarded as a "simple" requirement, it is challenging to achieve, but the potential for improvement exists. Some transparency measures can generate unintended consequences such as the exploitation of algorithmic systems. Assuming this is a possibility, will technology companies gain authorization to eliminate such transparency measures? The regulatory push for transparency stems from a lack of trust in technology companies. A practitioner further advanced the notion that suboptimal transparency standards have contributed to the portrayal of algorithms as harmful. In this regard, a tech representative conceded that in as much as certain technology policy personnel aspire to have a better grasp of algorithms, it is challenging to understand the underlying mechanisms of algorithmic decision-making. The separation of issues requiring transparency and other forms of remediation is similarly noteworthy. For example, matters relating to the unauthorized collection of user information represent a privacy and surveillance problem and should be addressed as such. The sluggish and fragmented formulation of policies further complicates the extent and speed with which companies can conform and comply with new laws. Striking the balance between responsiveness to transparency demands and the perpetual cycle of policy guidelines with the potential to instigate system vulnerabilities is evidently challenging.

Transparency has broader implications for ethical, legal, and other operational domains. In this regard, a random test question gauged the level of researchers' engagement with ethical issues in the field of technology. Whereas great strides have been made, there appear to be gaps in certain aspects. A researcher highlighted the contradictions in transparency requirements with the suggestion that researchers and policymakers lack clarity on the nature and levels of transparency standards. At the same time, platforms are not entirely clear on the data they should share with researchers. Ethical, legal, and other dilemmas or limitations come into play. Researchers may at times request source codes, which are trade secrets that companies cannot disclose. In this instance there are clear legal limitations, but on other occasions researchers may make requests that are deemed unethical when they request access to back-end platform domains for the purpose of user experimentations. For their part, platforms are bound by their terms of service, which are essentially contracts with

users.

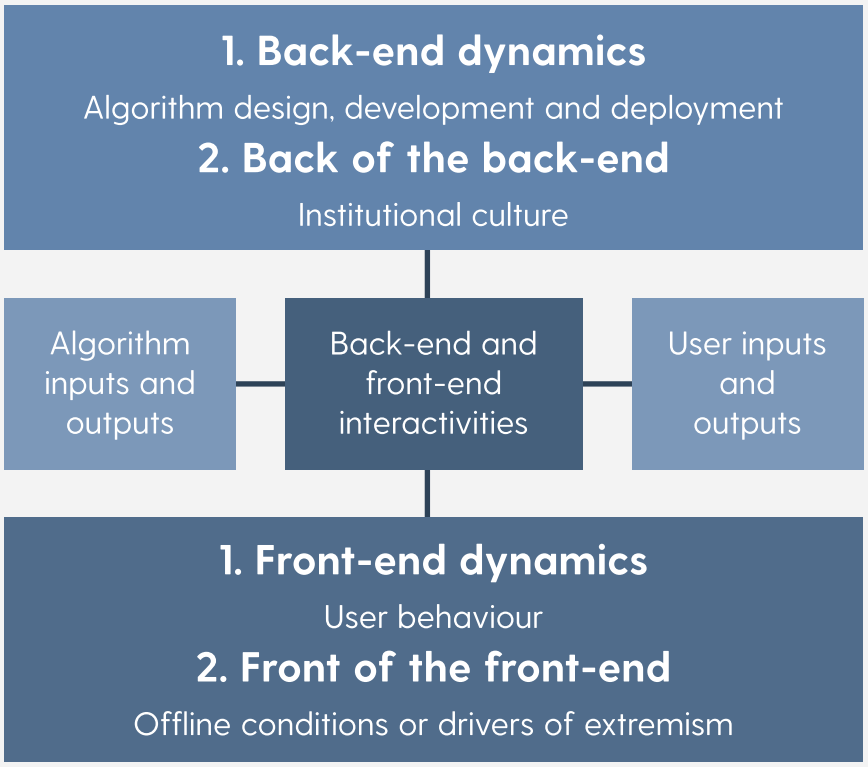
Third-party access to user data, whether by researchers or other entities, presents some ethical questions relating to informed consent, breach of confidentiality, and other ethical benchmarks. Conflicts of interest can lead to violations of privacy when (for instance) third parties unwittingly gain access to the data of research subjects with whom they are familiar or at odds. The ownership of data, third-party access to user data, and other interrelated transparency concerns are therefore important aspects of the data quagmire. These issues are further complicated by the absence of comprehensive legislation on data ownership in most jurisdictions. However, the GDPR has set the standard in this arena.[228] While the law has prescribed some rights that relate to data ownership, its impact in the context of user rights awareness and policy uptake remains to be seen. A sensible re-examination of these issues and the development of robust data ethics guardrails could significantly strengthen the rights of digital publics.

Section 9: Conclusion

9.1. The Bigger Picture of the Role of Algorithms in Extremism

The issues that punctuate the varied levels of understanding on the role of algorithms in extremism occasion a re-engagement with the context in which they emerge and evolve. Overall, as illustrated in the diagram below, this examination of extremism in the cyber-physical space has underlined the primacy of front-end and back-end dynamics and the interplay between the two domains. The back-end encapsulates the mechanisms of algorithm design, development, and deployment, while the real-world problems that algorithms aim to solve alongside the worldviews, values, goals, and other personal or institutional orientations constitute the "back of the back-end." The secluded "back of the back-end" subsumes manifestations of institutional culture that interact with the back-end and broader contextual conditions. All things considered, should companies accentuate the tweaking of algorithms or the "tweaking" of institutional culture? A healthy organizational culture should undergird and will safeguard the comprehensive mainstreaming of safety by design principles and protocols. The appraisal of business models should therefore occur within the wider scope of cultural transformation.

228 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1 S (2016).



The front-end dynamics encapsulate various dimensions of user behavior including the expression of content co-creation as productive consumption. The "front of the front-end" entails the less acknowledged offline conditions or drivers of extremism that influence or are mirrored in front-end user behavior. The "back-end and front-end interactivities" convey the symbiotic relationship between users, algorithms, and broader contextual factors. This relationship is animated in the framework through the properties of algorithmic amplification. In the long run, a better understanding of extremism calls for a better grasp of these contextual dynamics. More importantly, a better understanding and appreciation of these complexities requires greater receptivity to new and conflicting ideas and a balanced assessment of emerging evidence. It is possible that a shift in thinking could herald a gradual transition to more holistic, integrated, and sustainable interventions.

9.2. Summary of Key Findings

This research has examined the contextualized potential role of algorithms in extremism. It has assessed the varied levels of understanding on the subject and illuminated important findings alongside emerging complexities and dilemmas. More specifically, the analytical framework applied has examined the role of algorithms within the interlocking architectures of public policy environment, transparency, and the nature and structure of agency. The study has examined the role of algorithms in extremism through the HS framework. This model calls for "people-centred, comprehensive, context-specific and prevention-oriented responses that strengthen the protection and empowerment of all people and all communities."[229] While the conception of HS is for the most part threat-centric, it is important to underscore the value of capability and utility models in the sphere of technology. Such models incorporate and affirm (among other relevant benchmarks) the

229 "Human Security Handbook.", 6

existing capacities, opportunities, safeguards, and rights that may be amplified or replicated in pursuit of positive outcomes.[230] Despite these mechanisms, the prevailing discourse on the role of algorithms in extremism and other online harms does not eliminate the risk of the securitization of algorithms or AI.

The derogation of contextuality in the examination of online extremism is indicative of the incident-driven character of contemporary research. Acontextuality is linked to several layers of potential outcomes. The relegation of context has created an artificial dichotomy between online and offline extremism. Meanwhile, the compartmentalization of user and algorithmic agency, while analytically prudent, is increasingly generating a disproportionate focus on algorithms. These deficiencies have unwittingly contributed to the fragmentation of interventions with broader implications for sustainability. A more comprehensive context analysis can inform a better understanding of "online" extremism, the interlinkages with associated harms, and the hybridity of actors. The reclamation of context in analytical processes can in turn bolster policy and operational exertions on extremism in the cyber-physical space.

These considerations additionally highlight the relative autonomy of algorithms in the user-context spheres of interactivity. The cost of contextual oversight includes the design of algorithm-centric legislation as evidenced in China. Besides the subjectivity of interpretations and the challenges of enforceability, the replication and perversion of such laws in other jurisdictions could be just as costly. The case studies on the Jasmine Revolution and the Buffalo terrorist attack demonstrate the complex relationship between online and offline extremism and additionally accentuate the value of context.

The agency of algorithms and internet users in concert with other contextual factors is a critical research hot spot. The injection of more analytical capital into this domain of inquiry can enhance a deeper understanding of extremism in the cyber-physical space. It can also enrich the review of current thought and practices and increase the potential for more coherent policy and operational prescriptions. At the same time, algo-solutionism, while no panacea for the prevailing challenge of "online" extremism, can provide real-time, cost-effective, and efficient solutions that complement human capabilities.[231] While representing important points for introspection, the agentic and contextual issues raise important questions on extremism in the cyber-physical space. The first relates to the relationship between the online and offline extremist ecosystems and the importance of an integrated and holistic approach. Correspondingly, the notion of lone wolf algorithms invokes a deeper appreciation for contextuality.

Extremism in the cyber-physical space is characterized by the hybridity of actors, strategies, channels, causes, and consequences in the online and offline domains. In this ecosystem, discourse and action elicit policy while certain practices advance discourse within a network of actors in a mutually reinforcing relationship. The dynamics assessed not only represent the spillover of extremism from the offline to the online domain, but also the continuous interplay between both spheres. The harms largely perceived as digital can be conceived as iterations and adaptations of pre-existing

230 Sabina Alkire, "A Conceptual Framework for Human Security," Working Papers, CRISE (Department of International Development, University of Oxford), 2016, https://ora.ox.ac.uk/objects/uuid:d2907237-2a9f-4ce5-a403-a6254020052d.

231 Roff, Advancing Human Security.

offline harms or as reflections of the real world.[232] A deeper examination of the dominant discourse suggests that online-offline links are often established at the level of the offline consequences of online extremism. This line of thought rarely (if at all) recognizes the offline causes of online extremism. In addition, the feedback loop between the online and offline contexts in relation to the complementarity and mutually reinforcing influences in both environments is often neglected.

The diverse conceptions of algorithmic amplification are a testament to the interactivity between algorithms, users, and the phygital world. The role of algorithms in facilitating the hybridization of the cyber-physical space, with implications for extremism and other online harms, should therefore animate ongoing discourse, interventions, and future studies. Beyond theory, the knowledge generated from the three arenas can systematically inform the design of more integrated and holistic interventions that address extremism in the cyber-physical space. Secondly, the findings ought to impel policymakers and industry to revisit and reconfigure the policy environment to account for user agency, values, preferences, rights, and attendant capacity constraints.

The utility of the HS approach as people-focused has re-centralized the notion of agency in the prevailing discourse on "online" extremism. Lieberman's assertion that user (and more broadly human) interaction is key in algorithmic processes is still pertinent in the current climate.[233] The constructions and expressions of agency represent important parameters of analysis in the domain of extremism and related policies. The complexities around beneficent and adversarial influences demonstrate the potential for both ill and well-intentioned actors to transmit their goals and value systems in the design, deployment, and exploitation of algorithms. As a result, humans and algorithms that embody forms of subjectivities are important elements in the (re)constitution of security in the cyber-physical space.[234]

The formative actor mapping exercise has unmasked the sophistication of the nature and agency of actors in the extremist space. Actor mapping informs the identification of key players in the policy environment. The distinctiveness of actors in relation to roles, values, strategies, resources, and the relationships between them are important levers in the development of strategies of engagement and influencing. Overall, actor mapping contributes to a more nuanced and holistic understanding of context and the development of targeted interventions.

The findings on the role of algorithms in extremism with regard to what is understood, misunderstood, and not yet understood have wide-ranging implications for policy. The ongoing production of tech legislation is an important component of interventions aimed at countering extremism. Public policies and internal corporate policies fall within the remits of government and industry, which are complemented by civil society oversight and support. Policymaking in both spheres should therefore garner equal attention. The iterative character of policymaking in the real world is complicated by the dynamism and evolution of technology and extremism. Some of

232 Rowa, "Part 1."

233 Henry Lieberman, "Interaction Is the Key to Machine Learning Applications," in Workshop on Learning from Examples and Programming by Demonstration, International Conference on Machine Learning (Lake Tahoe, California, 1995), https://web.media.mit.edu/~lieber/Lieberary/AI/Interaction-Is/Interaction-Is.html.

234 Rowa, "Part 1"; Rowa, "Part 2."

the emerging issues on policy relate to the value of good practices in fast-evolving technological contexts that in turn compel continuous introspective and discretionary policy approaches.

The axis of context, content, actors, and process are widely considered as integral to good policymaking (in combination with more nuanced or elective considerations). The evaluation of policy impacts along the standards of implementation, enforcement, compliance, and other parameters is premature at this stage considering the currency of new and emerging laws. However, that should not invalidate the significance of short and medium-term policy monitoring and stock-taking. The extent to which these laws are untethered from the trap of algo-centrism but at the same time address prevailing algorithmic challenges calls for more comprehensive models of analysis that test and guide interventions on extremism. The effectiveness of these laws and in particular their responsiveness to the prevailing problem of extremism could positively or negatively inspire copycat laws. A systemic back and front-end approach that among other standards of good practice integrates robust strategies of stakeholder engagement is particularly pivotal.

It has emerged that the mixed or inconsistent research findings, conceptual ambiguities, and general methodological shortcomings complicate policy and programmatic efforts. The findings on algorithmic recommendation of extremist content range between positive effects, mixed outcomes, and minimal-to-no effect. The existing evidence on the potential role of algorithms in extremism is therefore particularly instructive if not compelling. The dearth of research on algorithm-user interactivities in context is a crucial gap in current literature. These limitations make it challenging to draw conclusive arguments on the role of algorithms in extremism. Existing studies provide more insight into the general mechanics of content recommendation and the extent to which content-sharing algorithms may or may not recommend extremist content. But they provide less insight on the more complicated question of how the recommendation of extremist content is likely or not to foment extremism, which in itself amplifies the importance of context. Nevertheless, these studies are foundational as far as the generation of model research designs is concerned. They additionally call into question the neutrality of platforms and algorithmic systems and provide some preliminary indications on the potential role of algorithms in extremism.

Overall, the utility of digital technologies is becoming more apparent in the domain of HS, which nevertheless faces three important constraints. These include poor threat prediction, inadequate planning of appropriate interventions, and the weak capacity of stakeholders to effectively respond. While capacity remains a central issue, the goodwill to tackle prevailing challenges with an increased appreciation of the complexities, interlinkages, comprehensiveness, and collective complementary actions can also go a long way to address capability concerns.

Transparency is also an overarching and recurrent theme as well as a cross-sectoral concern. It is therefore important to consider the character of algorithmic transparency in context within its institutional habitat while considering the systemic deficiencies and limitations around algorithmic transparency and how they can be effectively addressed. The nature of institutional relationships with reference to data governance and the functionality and access to algorithmic systems necessitates the conception of transparency as a horizontal and inter-mutual endeavor. Collaborative transparency should progressively enhance accountability across all actors engaged with industry

and promote a more integrated and comprehensive approach to human rights and ethics.

Ultimately, effective interventions on extremism in the cyber-physical space require a deeper level of introspection on current approaches, a critical and balanced assessment of existing and emerging evidence, an appreciation of the complexities, and greater receptivity to new and competing ideas.

9.3. Recommendations

Policymakers

1. Take the lead in stimulating collective action and act as intermediaries between neglected actors and existing consortiums on tech. Invite fringe platforms and other ambivalent actors to policy spaces and link them to relevant multistakeholder forums.
2. Strongly consider the experiences of victims, human rights defenders, and other at-risk groups, and prioritize and address their needs in the regulation of encryption and online identity (anonymity or pseudonymity). In addition, strengthen the current focus on children who are extremely vulnerable to extremism and other online harms. The exclusion of any group is likely to generate flawed policies that preference the needs of some actors over others. Consultation processes should additionally include secondary stakeholders with the objective to find a middle ground in the design of fair and just policies that are responsive to the primary groups of concern.
3. Identify, increase focus on, and design tailored policies for neglected but important actors and platforms such as audio streaming and other types of platforms. The disproportionate focus on podcasts relegates music, games, and other art forms that imbue unique layers of sophistication and complicate the detection of extremist content and other harms. Design targeted interventions that account for differences in the institutional character of platforms. Finally, traditional and alternative media are closely intertwined with the digital ecosystems of extremism. Review and accommodate hybrid media structures in the ongoing tech legislation processes in situations where they violate and strategically neglect their corporate codes of conduct.
4. Beyond evidence-based policymaking, consider the nature, currency, and limitations of existing research to inform legislative decisions. In addition, governments should make provisions for broad-based public consultations and submissions in tech policymaking processes.
5. Beyond classifications, future actions on borderline content should consider the severity of potential long-term impacts on victims, at-risk groups, and society in general. The ostensibly benign short-term effects of borderline content could misinform response. That said, the first port of call for managing mainstream offenders should be public service and other institutional codes of conduct and ethics regardless of user status. The intervention of technology companies, whether primary or subsidiary, should be discretionary and informed by the presence or absence of those codes. Actions that are unlawful in the offline environment should not be legitimated in the online environment. User capacity for better self-regulation in offline environments stands in stark contrast to their online disinhibitionism.

Accountability mechanisms should therefore target both industry and primary content creators, page administrators, and other primary offenders.

6. Develop a global governance framework that stipulates the rules, norms, and principles for technology companies and governments to reconcile existing inconsistencies in approaches. Adapt or translate the global framework into regional or National Action Plans (NAPs). The EU should mobilize other regions in the development of a global framework as a next step.
7. Increase engagement with emerging technologies that are under consideration for regulation. Test these technologies to understand their mechanics and conduct rigorous evaluations of tough cases. The capacity to listen and understand, receptivity to new ideas, and adaptability can bolster the confidence of industry to improve.

Technology Companies

1. Establish the infrastructure for safety by design that will build trust, promote safety principles and protocols, and anchor the development of safer digital technologies.
    a. A good starting point is the routinization of organizational introspection to inform the transformation of institutional cultures.
    b. Commit to the proactive and systematic mainstreaming of rights-based approaches in corporate policies and operations.
    c. Ensure the systematic integration of institution-wide transparency and ethical standards, including risk assessment and management.
    d. Operationalize strategies to harmonize new guidelines with existing policies and procedures. Policy alignment additionally includes deconflicting existing cross-departmental policies, and appraising competing business interests against existing safety, human rights, ethics, and transparency frameworks.
    e. Proactively design user-centric technologies with embedded safety features and mechanisms that are pre-tested prior to deployment.
    f. Devise robust data governance frameworks including monitoring, evaluation, and learning systems.
    g. Address operational challenges and other capacity gaps through the culture of information sharing and requests for technical assistance.
    h. Regularly monitor for compliance, successes, challenges, and other benchmarks.
    i. Evaluate impact over time and reorient the culture, systems, and products accordingly.
2. Safety by design cannot futureproof every conceivable risk or harm. The continuous anticipation and assessment of risks and the modification of systems in light of new risks and harms are vital. The modification of algorithms or recommender systems should take into account the potential for unintended effects. Provide clear and transparent trails of documentation on implementation, learning, and adaptation.

3. Technology companies should increase access to their personnel, data, and systems. Beyond building trust, such measures can strengthen the understanding of research problems and inform the design of appropriate studies and interventions that are more responsive to the drivers of extremism. While greater access raises its own risks and complexities, the technology and research communities should re-examine, adapt, and/or reformulate the standards of what qualifies as ethical and meaningful access, and put in place the necessary guardrails to mitigate against any risks arising from all parties.
4. Consider a multi-pronged course of action in the online space. Target extremist content and the surrounding (user) networks of amplification. Another suggestion from one of the working groups that could be tested includes maintaining digital activity logs on content takedowns that can also trace back the primary spreaders of extremist content for network takedowns. Trace-backs may identify both algorithm and user activity.
5. Technology companies should set realistic goals for online safety. An incremental approach, in consultation with government, could foster the prioritization of specific areas of intervention and adaptation to new laws. Such an approach can facilitate resource optimization and minimize patchworks and band-aids that typify "a bit of everything" or a "cocktail" approach to complex problems.
6. There are diverse models of online safety that are institution-focused, multistakeholder-focused, or unresponsive postures that can fracture or strengthen collective efforts and impact. Technology companies should reserve the right to choose their associations, influences, or preferences. Establish unitary or regional tech consortiums to enhance cooperation, information sharing, and the harmonization of approaches to extremism and other online harms.

Civil Society

1. GIFCT: A starting point for refreshed GIFCT working groups could include the annual review of actions that companies took as a result of the knowledge, tools, and guidelines generated in the previous year. Explore the tools and mechanisms that companies found useful, how they were applied in context, and what did not work well. This could be an opportunity to review pending targets, engage with challenges in the field, and access technical support from other stakeholders.
2. Integrate the perspectives and experiences of real users in studies with greater focus on victims or targets of extremism and other online harms. This will generate a more comprehensive understanding of the dynamics of extremism. It also promotes fair subject selection and compels a delicate balance between ethical risks and opportunities.
3. Researchers working on socio-technological issues who do not have an online presence are more likely to bolster their research through active digital ethnography.
4. Minimize the "silofication" of research methods and approaches. Cultivate a stronger appreciation for and implement complementary or mixed research methods. Conduct interdisciplinary research that aims to address the multifaceted nature of extremism, cross-cultural dynamics, human rights, and other crosscutting issues. Theoretical research has

its strengths and limitations. Researchers outside the field of technology should therefore increase engagement with the tech community in order to strengthen the reliability and validity of their studies.

5. Case studies deserve more consideration. Besides aggregating perpetrator and victim experiences, in-depth case studies can expose complexities, distinctions, and other aspects of a case that can enhance the understanding of the dynamics of extremism in the cyber-physical space. Case studies can also represent an initial exploratory method that can define the parameters for subsequent research methods. These studies can additionally highlight what in the onset may manifest as an outlier or deviant case but in reality portends broader impacts for society.
6. Regularize the clear documentation of study methodology and limitations to guide the review and interpretation of research findings. A robust research trail will additionally clarify what is understood, misunderstood, or not yet understood and create avenues for future research.
7. Develop good practices or benchmarks for testing algorithms.
8. Examine the impact of tech (algorithmic) legislation on both big and small companies, governments, and users, alongside the convergences and divergences of experience. Determine the extent to which impacts on governments provide incentives for improved practices in policymaking.

All Stakeholders

1. Adopt a strategic and holistic approach in the review of existing interventions on extremism in the online and offline environments. Develop multi-actor, multi-strategy, and multi-level integrated approaches that account for phygital interactivities to increase impact and sustainability.
2. Build or strengthen existing mechanisms for collaborative transparency. Every actor should formulate standards that guide their interactions and increase transparency in their engagements with industry.
3. Policymaking should be matched by the periodic review of gaps or potential legislative pitfalls, the practicability of application, and enforcement. Additionally, consider the risk of unintended consequences such as the potential for certain laws to promote certain rights while undermining the enjoyment of others.
4. Implement tailored capacity-building programs that address specific gaps across all sectors. Promote user empowerment through enhanced digital literacy and civic engagement programs. Strategies for public sensitization include tech camps and the organization of forums to strengthen the awareness of online threats, trends, safety tools, access to social services, and other offline resources. The integration of relevant curricula in formal institutions of learning is essential. Increase publicity for online offenders and defenders to educate and curb impunity.
5. Establish a robust user empowerment architecture and incorporate strategies such as

consultations and user-centric data governance that also increase user control or power over their data. Fulfill public (individual) requests for information on user data.

6. Build global North-South partnerships to address some of the challenges relating to language, culture, knowledge gaps, and marginalization.
7. Cultural transformation should occur at personal, relational, and institutional levels.
8. Shared responsibility and concerted action should be a priority for all stakeholders.

9.4. An Overview of Emerging Complexities and Dilemmas

There are several complications and dilemmas that the study has identified. The major ones are outlined below:

1. Conceptual ambiguities: Insufficient clarity and scope on extremism, algorithmic amplification, borderline content, and other relevant terminology, with implications for research, policy, and practice.
2. The algorithm black box: The technical properties that designate the opacity of algorithmic systems include data privacy, changes in data over time, and the interconnectedness of processes and decisions that are learned from data. Another major aspect is complexity as relates to the interconnectedness of algorithms, iterative processing, the scale of data, and randomized tiebreaking. Overall, the complexity of algorithmic systems in combination with limited algorithmic transparency characterize the challenges of explicability, interpretability, and auditability of black box models. These issues have broader implications for research and policy.
3. Attribution (cause and effect and the disaggregation of agency): The determination of the proportion of amplification that algorithms are responsible for is complicated. The same applies to the assignment of agency among the multiplicity of actors at the back and front-ends. A pertinent question is who or what is responsible for a specific part of an outcome. The distinction between pre-(machine) selected and user-selected content is similarly a knotty domain. More specifically, the complex online inter-relationship between social, traditional, and alternative media renders the segregation and apportionment of responsibility a complex task. These issues and more raise the question of the fair and just apportionment of accountability.
4. Borderline content: It is increasingly clear that a small group of users is thriving off borderline content for financial and other gains. Besides the subjectivity of the interpretation of borderline content, the blurred lines between mainstream and borderline designations have become exploitable. Borderline actors exploit and are unwittingly insulated by the banner of free speech and other freedoms that complicate the policy and regulatory space.
5. Counter-extremism scope creep: As with other mainstream actors, the realm of political actors presents technology companies with the dilemma of counter-extremism scope creep. The significant costs associated with the application of platform policies to public figures can trigger a range of unintended effects. Ultimately, who reserves primary

responsibility for the actions of such actors?

6. Challenges in research: Methodological limitations and inconsistencies in the existing evidence base on the role of algorithms in extremism confound research audiences. In some instances, studies do not make clear distinctions between the examination of whether recommender systems actually cause harm vis-à-vis whether these systems direct users to potentially harmful content. A second challenge relates to building consensus on the definition of evidence and the ability to identify, interpret, and apply sound and relevant research. These issues present broader ramifications for effective policymaking and the design of appropriate and sustainable interventions.
7. Models of intervention: This relates to the scope and strategies for intervention. Should interventions target content and technological tools or the nefarious individuals and networks who exploit them, or both? The management of mainstream actors exploiting and engaging with borderline content is equally important.
8. Collective action: Some platforms are committed to concerted action through such multistakeholder forums as GIFCT. Their efforts are diminished by the inaction of companies that shun a collaborative approach to tackling extremism and other harms. Consequently, the digital territorialization of interventions has a restricted impact. What strategies of inclusion are likely to draw these actors to existing programs?
9. User empowerment and accountability: The nature of users and their rights and roles call for a deeper understanding and management of user dynamics. How can policymakers reconfigure the policy environment to account for user agency, values, preferences, and rights?
10. Conflicting values: Users hold varying perspectives on the regulation of encryption and online identity or anonymity. This is based on the nature of users (for example, victims, human rights defenders, and legislators), their values, and preferences (among other divergent characteristics).
11. Contextual disparities: The subjectivity of culture, human rights, and other contextual nuances in the global arena. These variances could complicate compliance for technology companies. The convergence of conceptual, contextual, and legislative disparities at the country level has considerable influence on collective action in global governance.
12. Digitization of state repression: State exploitation of tech legislation to simultaneously democratize and autocratize.
13. The nexus between mental health and extremism: This has implications for user accountability.

Bibliography

Abdelaty, Mariam. "Democratization and Extremism: The Case of Tunisia." Theses and Dissertations, June 15, 2021. https://fount.aucegypt.edu/etds/1671.

"Ad Hoc Expert Group (AHEG) for the Preparation of a Draft Text of a Recommendation on the Ethics of Artificial Intelligence." United Nations Educational, Scientific and Cultural Organisation (UNESCO), 2020. https://unesdoc.unesco.org/ark:/48223/pf0000373199?posInSet=2&queryId=N-EXPLORE-64abc70f-4002-4d27-a7ac-ad15d3645536.

Aday, Sean, Henry Farrell, Marc Lynch, John Sides, and Dean Freelon. "Blogs and Bullets II: New Media and Conflict after the Arab Spring." United States Institute of Peace, July 10, 2012. https://www.usip.org/publications/2012/07/blogs-and-bullets-ii-new-media-and-conflict-after-arab-spring.

Agudo, Ujué, and Helena Matute. "The Influence of Algorithms on Political and Dating Decisions." PLOS ONE 16, no. 4 (April 21, 2021): e0249454. https://doi.org/10.1371/journal.pone.0249454.

Akinwotu, Emmanuel. "Nigeria Suspends Twitter Access after President's Tweet Was Deleted." The Guardian, June 4, 2021. https://www.theguardian.com/world/2021/jun/04/nigeria-suspends-twitter-after-presidents-tweet-was-deleted.

"Algorithms and Terrorism: The Malicious Use of Artificial Intelligence for Terrorist Purposes." United Nations Counter-Terrorism Centre (UNCCT) and United Nations Interregional Crime and Justice Research Institute (UNICRI), 2021. https://www.un.org/counterterrorism/sites/www.un.org.counterterrorism/files/malicious-use-of-ai-uncct-unicri-report-hd.pdf.

Alkire, Sabina. "A Conceptual Framework for Human Security," Working Papers, CRISE (Department of International Development, University of Oxford), 2016. https://ora.ox.ac.uk/objects/uuid:d2907237-2a9f-4ce5-a403-a6254020052d.

Amoore, Louise, and Rita Raley. "Securing with Algorithms: Knowledge, Decision, Sovereignty." Security Dialogue 48, no. 1 (February 1, 2017): 3–10. https://doi.org/10.1177/0967010616680753.

"Anatomy of a Disinformation Empire: Investigating NaturalNews." Institute for Strategic Dialogue, 2020. https://www.isdglobal.org/wp-content/uploads/2021/10/20211013-ISDG-NaturalNews-Briefing.pdf.

Anderson, Lisa. "Demystifying the Arab Spring," May/June 2011. https://www.foreignaffairs.com/articles/libya/2011-04-03/demystifying-arab-spring.

Archetti, Cristina. "Terrorism, Communication and New Media: Explaining Radicalization in the Digital Age." Perspectives on Terrorism 9, no. 1 (2015): 49–59.

Argomaniz, Javier, and Alberto Vidal-Diez. "Examining Deterrence and Backlash Effects in Counter-Terrorism: The Case of ETA." Terrorism and Political Violence 27, no. 1 (January 1, 2015): 160–81. https://doi.org/10.1080/09546553.2014.975648.

Asaro, Peter. "A Review of Private Sector AI Principles: A Report Prepared for UNIDIR." Geneva: UN Institute for Disarmament Research, 2019.

Ayyub, Rana, and Courtenay Werleman. "Govt Asks Twitter to 'take down' Freedom House's Tweets," June 29, 2022. https://www.telegraphindia.com/india/twitter-doc-shows-government-requests-for-blocking-tweets-of-some-advocacy-groups-politicians/cid/1872112.

Barakat, Zahraa, and Ali Fakih. "Determinants of the Arab Spring Protests in Tunisia, Egypt, and Libya: What Have We Learned?" Social Sciences 10, no. 8 (2021): 282.

Bartlett, Matt. "Solving the AI Accountability Gap: Hold Developers Responsible for Their Creations." Medium (blog), April 5, 2019. https://towardsdatascience.com/solving-the-ai-accountability-gap-dd35698249fe.

Barton, Greg. "ASIO's Language Shift on Terrorism Is a Welcome Acknowledgment of the Power of Words." The Conversation.

March 21, 2021. http://theconversation.com/asios-language-shift-on-terrorism-is-a-welcome-acknowledgment-of-the-power-of-words-157400.

Basit, Abdul. "Conspiracy Theories and Violent Extremism: Similarities, Differences and the Implications." Counter Terrorist Trends and Analyses 13, no. 3 (2021): 1–9.

Bathaee, Yavar. "The Artificial Intelligence Black Box and the Failure of Intent and Causation." Harvard Journal of Law & Technology 31 (2018): 889.

Bazzaz Abkenar, Sepideh, Mostafa Haghi Kashani, Ebrahim Mahdipour, and Seyed Mahdi Jameii. "Big Data Analytics Meets Social Media: A Systematic Review of Techniques, Open Issues, and Future Directions." Telematics and Informatics 57 (March 1, 2021): 101517. https://doi.org/10.1016/j.tele.2020.101517.

Beard, Charles A. "Time, Technology, and the Creative Spirit in Political Science." The American Political Science Review 21, no. 1 (1927): 1–11.

Benson, David C. "Why the Internet Is Not Increasing Terrorism." Security Studies 23, no. 2 (April 3, 2014): 293–328. https://doi.org/10.1080/09636412.2014.905353.

Bessi, Alessandro, Ferrara, Emilio. "Social bots distort the 2016 U.S. Presidential election online discussion." First Monday 21, no. 11 (November 7, 2016). http://dx.doi.org/10.5210/fm.v21i11.7090.

Bhaskar, Roy. The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences. New York: Routledge, 1998.

Bimber, Bruce. "Three Faces of Technological Determinism." In Does Technology Drive History, edited by Merrit Roe Smith and Leo Marx, 79–100. Cambridge, MA: MIT Press, 1994.

Birnhack, Michael, and Niva Elkin-Koren. "The Invisible Handshake: The Reemergence of the State in the Digital Environment." SSRN Scholarly Paper. April 10, 2003. https://doi.org/10.2139/ssrn.381020.

Bogain, Ariane. "Understanding Public Constructions of Counter-Terrorism: An Analysis of Online Comments during the State of Emergency in France (2015-2017)." Critical Studies on Terrorism 13, no. 4 (October 1, 2020): 591–615. https://doi.org/10.1080/17539153.2020.1810976.

Broderick, Ryan. "You Can't Always Blame Algorithms." Garbage Day, May 17, 2022. https://www.garbageday.email/p/you-cant-always-blame-algorithms.

Brown, Heather, Emily Guskin, and Amy Mitchell. "The Role of Social Media in the Arab Uprisings." Pew Research Center's Journalism Project (blog), November 28, 2012. https://www.pewresearch.org/journalism/2012/11/28/role-social-media-arab-uprisings/.

Brown, Michael E. The International Dimensions of Internal Conflict. Cambridge, MA: MIT Press, 1996.

Brown, Shea, Jovana Davidovic, and Ali Hasan. "The Algorithm Audit: Scoring the Algorithms That Score Us." Big Data & Society, January, 2021. https://doi.org/10.1177/2053951720983865.8.

Bucher, Taina. If...Then: Algorithmic Power and Politics. Oxford Studies in Digital Politics. New York: Oxford University Press, 2018. https://doi.org/10.1093/oso/9780190493028.001.0001.

Bundtzen, Sara. "What China's Sweeping Algorithm Regulation Means for Digital Governance Globally." Institute for Strategic Dialogue, May 31, 2022. https://www.isdglobal.org/digital_dispatches/chinas-sweeping-algorithm-regulation-and-global-digital-governance/.

Burrell, Jenna. "How the Machine Thinks: Understanding Opacity in Machine Learning Algorithms." Big Data & Society (January-

June 2016): 1-12. doi: 10.1177/2053951715622512

Byun, Chonghyun Christie, and Ethan J. Hollander. "Explaining the Intensity of the Arab Spring." Digest of Middle East Studies 24, no. 1 (2015): 26–46. https://doi.org/10.1111/dome.12057.

Caiani, Manuela, and Linda Parenti. "The Dark Side of the Web: Italian Right-Wing Extremist Groups and the Internet." South European Society and Politics 14, no. 3 (September 1, 2009): 273–94. https://doi.org/10.1080/13608740903342491.

Cairney, Paul, and Christopher M. Weible. "The New Policy Sciences: Combining the Cognitive Science of Choice, Multiple Theories of Context, and Basic and Applied Analysis." Policy Sciences 50, no. 4 (December 1, 2017): 619–27. https://doi.org/10.1007/s11077-017-9304-2.

Camus, Renaud. Le grand remplacement. Plieux: Renaud Camus, 2012.

Carlisle, Rodney P. Encyclopedia of Politics: The Left and the Right Vol. 2. Thousand Oaks, CA: SAGE, 2005.

Chakma, Anurug. "Does State Repression Stimulate Terrorism? A Panel Data Analysis on South Asia." Journal of Policing, Intelligence and Counter Terrorism 17, no. 2 (2021). https://www.tandfonline.com/doi/full/10.1080/18335330.2021.2022184.

Chandler, David. "Human Security: The Dog That Didn't Bark." Security Dialogue 39, no. 4 (2008): 427–38.

Chang, Cleo. "The Unlikely Connection Between Wellness Influencers and the Pro-Trump Rioters." Cosmopolitan, January 12, 2021. https://www.cosmopolitan.com/health-fitness/a35056548/wellness-fitness-influencers-qanon-conspiracy-theories/.

Chen, Lucia Lushi, Walid Magdy, and Maria K. Wolters. "The Effect of User Psychology on the Content of Social Media Posts: Originality and Transitions Matter." Frontiers in Psychology 11 (2020). https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00526.

Cho, Gyeongcheol, Jinyeong Yim, Younyoung Choi, Jungmin Ko, and Seoung-Hwan Lee. "Review of Machine Learning Algorithms for Diagnosing Mental Illness." Psychiatry Investigation 16, no. 4 (April 2019): 262–69. https://doi.org/10.30773/pi.2018.12.21.2.

Chokoshvili, Davit. "The Role of the Internet in Democratic Transition: Case Study of the Arab Spring." Budapest, Central European University: Master of Arts in Public Policy 51, 2011.

Cohen, James N. "Exploring Echo-Systems: How Algorithms Shape Immersive Media Environments." Journal of Media Literacy Education 10, no. 2 (2018): 139–51.

Collins, Ben. "The Buffalo Shooting Suspect Apparently Posted a Manifesto Citing 'Great Replacement' Theory." NBC News, May 15, 2022. https://www.nbcnews.com/news/us-news/buffalo-supermarket-shooting-suspect-posted-apparent-manifesto-repeate-rcna28889.

Condon, Bernard, and Michael Hill. "Buffalo Mass Shooting Suspect: 'Lonely,' 'Nerdy' Teenager Showed Signs of Trouble." Global News, May 17, 2022. https://globalnews.ca/news/8842287/buffalo-mass-shooting-suspect-payton-gendron/.

"Content Personalisation and the Online Dissemination of Terrorist and Violent Extremist Content." Tech Against Terrorism, 2021. https://www.techagainstterrorism.org/wp-content/uploads/2021/02/TAT-Position-Paper-content-personalisation-and-online-dissemination-of-terrorist-content1.pdf.

"Continuing Our Work to Improve Recommendations on YouTube." YouTube (blog), January 25, 2019. https://blog.youtube/news-and-events/continuing-our-work-to-improve/.

Cooley, Asya, Skye Cooley, Robert Hinck, and Sara Kitsch. "Influencing Public Behavior: Takeaways From Public Communication Scholarship." The Media Ecology and Strategic Analysis Group, October 1, 2020. https://apps.dtic.mil/sti/citations/AD1118281.

Coyne, Sarah M., Adam A. Rogers, Jessica D. Zurcher, Laura Stockdale, and McCall Booth. "Does Time Spent Using Social Media Impact Mental Health?: An Eight Year Longitudinal Study." Computers in Human Behavior 104 (2020): 106160.

Crenshaw, Martha. Terrorism in Context. University Park, PA: Penn State Press, 1995.

Culliford, Elizabeth. "Facebook and Tech Giants to Target Attacker Manifestos, Far-Right Militias in Database." Reuters, July 26, 2021. https://www.reuters.com/technology/exclusive-facebook-tech-giants-target-manifestos-militias-database-2021-07-26/.

"Cyber-Physical Systems - a Concept Map." Berkeley CPS Publications, n.d. https://ptolemy.berkeley.edu/projects/cps/.

Dalacoura, Katerina. "Terrorism, Democracy and Islamist Terrorism." In Islamist Terrorism and Democracy in the Middle East, 21–39. Cambridge University Press, 2011.

Danner, Chas. "What We Know About the Racist Attack at a Buffalo Supermarket." Intelligencer, May 17, 2022. https://nymag.com/intelligencer/2022/05/ten-dead-after-gunman-attacks-buffalo-supermarket-updates.html.

"Defining Extremism: A Glossary of White Supremacist Terms, Movements and Philosophies." Anti Defamation League, n.d., https://www.adl.org/resources/glossary-term/defining-extremism-glossary-white-supremacist-terms-movements-and.

Deji, Olanike F. Gender and Rural Development: Introduction. Vol. 1. Berlin: LIT Verlag Münster, 2011.

Desai, Deven R., and Joshua A. Kroll. "Trust but Verify: A Guide to Algorithms and the Law." Harv. JL & Tech. 31 (2017): 1–64.

Dhariwal, Kunal. "Cryptocurrency Mining Algorithms and Popular Cryptocurrencies." Medium (blog), March 3, 2018. https://medium.com/@Mr.dhariwal/cryptocurrency-mining-algorithms-and-popular-cryptocurrencies-48176d3559d6.

Dignum, Virginia. "The ART of AI – Accountability, Responsibility, Transparency." Medium (blog), March 4, 2018. https://medium.com/@virginiadignum/the-art-of-ai-accountability-responsibility-transparency-48666ec92ea5.

Duffield, Mark, and Nicholas Waddell. "Human Security and Global Danger: Exploring a Governmental Assemblage." University of Lancaster, ESRC New Security Challenges programme, 2004, https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.116.141&rep=rep1&type=pdf.

Duffield, Mark, and Nicholas Waddell. "Securing Humans in a Dangerous World." International Politics 43 (2006): 1–23.

Dunn Cavelty, Myriam. "Breaking the Cyber-Security Dilemma: Aligning Security Needs and Removing Vulnerabilities." Science and Engineering Ethics 20, no. 3 (September 1, 2014): 701–15. https://doi.org/10.1007/s11948-014-9551-y.

eSafety Commissioner Australia. "Abhorrent Violent Conduct Powers: Regulatory Guidance." 2021. https://www.esafety.gov.au/sites/default/files/2022-03/Abhorrent%20Violent%20Conduct%20Powers%20Regulatory%20Guidance.pdf.

———. "An overview of eSafety's role and functions." 2021. https://www.esafety.gov.au/sites/default/files/2021-07/Overview%20of%20role%20and%20functions_0.pdf.

Esposito, Elena. "Artificial Communication? The Production of Contingency by Algorithms." Zeitschrift Für Soziologie 46, no. 4 (2017): 249–65.

"Facebook's Algorithm: A Major Threat to Public Health." Avaaz, August 19, 2020. https://avaazimages.avaaz.org/facebook_threat_health.pdf.

Fedoruk, Benjamin, Harrison Nelson, Russell Frost, and Kai Fucile Ladouceur. "The Plebeian Algorithm: A Democratic Approach to Censorship and Moderation." JMIR Formative Research 5, no. 12 (2021): e32427.

Ferryman, Kadija, and Mikaela Pitcan. "Fairness in Precision Medicine," 2018.

Fisher, Ali. "Swarmcast: How Jihadist Networks Maintain a Persistent Online Presence." Perspectives on Terrorism 9, no. 3 (2015): 3–20.

Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI." SSRN Scholarly Paper. January 15, 2020. https://doi.org/10.2139/ssrn.3518482.

Floridi, Luciano. The Fourth Revolution: How the Infosphere Is Reshaping Human Reality. Oxford, UK: Oxford University Press, 2014.

Fournier, Philippe. "The Neoliberal/Neurotic Citizen and Security as Discourse." Critical Studies on Security 2, no. 3 (2014): 309–22.

Fukuda-Parr, Sakiko, and Elizabeth Gibbons. "Emerging Consensus on 'Ethical AI': Human Rights Critique of Stakeholder Guidelines." Global Policy 12 (2021): 32–44.

Gall, Richard. "Machine Learning Explainability vs Interpretability: Two Concepts That Could Help Restore Trust in AI." KDnuggets (blog), 2018. https://www.kdnuggets.com/machine-learning-explainability-vs-interpretability-two-concepts-that-could-help-restore-trust-in-ai.html/.

Galtung, Johan. "The Specific Contribution of Peace Research to the Study of the Causes of Violence: Typologies." UNESCO Interdisciplinary Expert Meeting on the Study of the Causes of Violence, 1975. https://www.amazon.com/specific-contribution-research-violence-typologies/dp/B0006CWC26.

Garcia, Blake Evan. "International Migration and Extreme-Right Terrorism." PhD Thesis, A&M University, 2015. https://oaktrust.library.tamu.edu/handle/1969.1/155240.

Gaudette, Tiana, Ryan Scrivens, Garth Davies, and Richard Frank. "Upvoting Extremism: Collective Identity Formation and the Extreme Right on Reddit." New Media & Society 23, no. 12 (2020): 3491–3508.

General Assembly Resolution 66/290 (2012, 10 September) Follow-up to Paragraph 143 on Human Security of the 2005 World Summit Outcome, A/RES/66/290. United Nations General Assembly, 2012. https://digitallibrary.un.org/record/737105.

Giddens, Anthony. The Constitution of Society: Outline of the Theory of Structuration. Univ of California Press, 1984.

GIFCT CAPPI Working Group. "Content-Sharing Algorithms, Processes, and Positive Interventions Working Group. Part 1: Content-Sharing Algorithms & Processes." Global Internet Forum to Counter Terrorism, 2021. https://gifct.org/wp-content/uploads/2021/07/GIFCT-CAPI1-2021.pdf.

Gill, Paul, Emily Corner, Maura Conway, Amy Thornton, Mia Bloom, and John Horgan. "Terrorist Use of the Internet by the Numbers: Quantifying Behaviors, Patterns, and Processes." Criminology & Public Policy 16, no. 1 (2017): 99–117.

Gillespie, Tarleton. Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. New Haven, CT: Yale University Press, 2018.

———. "The Relevance of Algorithms." Media Technologies: Essays on Communication, Materiality, and Society 167, no. 2014 (2012): 167.

Gladwell, Malcolm. "Small Change." The New Yorker, October 4, 2010. https://www.newyorker.com/magazine/2010/10/04/small-change-malcolm-gladwell.

Gluyas, Lee, and Stefanie Day. "Artificial Intelligence – Who Is Liable When AI Fails to Perform?" CMS, 2018. https://cms.law/en/gbr/publication/artificial-intelligence-who-is-liable-when-ai-fails-to-perform.

Goldman, Eric. "The Constitutionality of Mandating Editorial Transparency." Hastings Law Journal 73 (2022).

"The Government's Commitment to Address Online Safety." Government of Canada, July 8, 2022. https://www.canada.ca/en/canadian-heritage/campaigns/harmful-online-content.html.

Hadi, Hiba Jasim, Ammar Hameed Shnain, Sarah Hadishaheed, and Aziz Ahmad. "Big Data and Five V's Characteristics." International Journal of Advances in Electronics and Computer Science 2, no. 1 (2015). https://www.researchgate.net/profile/Ammar-Hameed-Shnain/publication/332230305_BIG_DATA_AND_FIVE_V%27S_CHARACTERISTICS/links/5ca76bbca6fdcca26d011d6a/BIG-DATA-AND-FIVE-VS-CHARACTERISTICS.pdf?origin=publication_detail.

Haidt, Jonathan. "Why the Past 10 Years of American Life Have Been Uniquely Stupid." The Atlantic, April 11, 2022. https://www.theatlantic.com/magazine/archive/2022/05/social-media-democracy-trust-babel/629369/.

Hansen, Lene, and Helen Nissenbaum. "Digital Disaster, Cyber Security, and the Copenhagen School." International Studies Quarterly 53, no. 4 (2009): 1155–75.

Hauer, Thomas. "Technological Determinism and New Media." International Journal of English Literature and Social Sciences 2, no. 2 (2017): 1–4.

Herzog, Christian. "On the Risk of Confusing Interpretability with Explicability." AI and Ethics 2, no. 1 (2022): 219–25.

Hodgson, Geoffrey M. "On Defining Institutions: Rules versus Equilibria." Journal of Institutional Economics 11, no. 3 (2015): 497–505.

Howard, Philip N., and Muzammil M. Hussain. Democracy's Fourth Wave?: Digital Media and the Arab Spring. Oxford, UK: Oxford University Press, 2013. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199936953.001.0001/acprof-9780199936953.

"Human Development Report 1994." United Nations Development Programme (UNDP). New York: Oxford University Press, 1994.

"Human Rights in the Age of Artificial Intelligence." Access Now, 2018. https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf.

"Human Security: An Approach and Methodology for Business Contributions to Peace and Sustainable Development." London School of Economics, n.d. https://www.lse.ac.uk/ideas/Assets/Documents/project-docs/un-at-lse/LSE-IDEAS-Human-Security-Background.pdf.

Human Security: Safety for People in a Changing World. Ottawa: Department of Home Affairs and Trade, 1999.

"Human Security Handbook: An Integrated Approach for the Realization of the Sustainable Development Goals and the Priority Areas of the International Community and the United Nations System." New York: United Nations, 2016. https://www.un.org/humansecurity/wp-content/uploads/2017/10/h2.pdf.

Huntington, Samuel P. Political Order in Changing Societies. New Haven, CT: Yale University Press, 2006.

Huszár, Ferenc, Sofia Ira Ktena, Conor O'Brien, Luca Belli, Andrew Schlaikjer, and Moritz Hardt. "Algorithmic Amplification of Politics on Twitter." Proceedings of the National Academy of Sciences 119, no. 1 (2022): e2025334119.

Islam, Mir Riyanul, Mobyen Uddin Ahmed, Shaibal Barua, and Shahina Begum. "A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks." Applied Sciences 12, no. 3 (2022): 1353.

Jenkins, Bert, D.B. Subedi, and Kathy Jenkins. Reconciliation in Conflict-Affected Communities. Singapore: Springer Nature Publication, 2018.

Jobin, Anna, Marcello Ienca, and Effy Vayena. "The Global Landscape of AI Ethics Guidelines." Nature Machine Intelligence 1, no. 9 (2019): 389–99.

Johansson Dahre, Ulf. "Searching for a Middle Ground: Anthropologists and the Debate on the Universalism and the Cultural Relativism of Human Rights." The International Journal of Human Rights 21, no. 5 (2017): 611–28.

Johnson, Bridget. "10 Killed in Buffalo Supermarket Attack Allegedly Inspired by Christchurch Terrorist," HS Today.US, May 15, 2022. https://www.hstoday.us/featured/10-killed-in-buffalo-supermarket-attack-allegedly-inspired-by-christchurch-terrorist/.

Jolly, Richard, and Deepayan Basu Ray. "The Human Security Framework and National Human Development Reports: A Review of Experiences and Current Debates." NHDR Occasional Paper 5 (2006).

Jones, Seth G., Catrina Doxsee, and Nicholas Harrington. "The Escalating Terrorism Problem in the United States." Centre for Strategic and International Studies, 2020. https://www.csis.org/analysis/escalating-terrorismproblem-united-states.

Jost, John T., Christopher M. Federico, and Jaime L. Napier. "Political Ideology: Its Structure, Functions, and Elective Affinities." Annual Review of Psychology 60, no. 1 (2009): 307–37.

Kassem, Julia. "Ten Years After 'Arab Spring.'" Al Mayadeen, October 10, 2021. https://english.almayadeen.net/articles/blog/ten-years-after-arab-spring.

Kassir, Sara. "Algorithmic Auditing: The Key to Making Machine Learning in the Public Interest." The Business of Government, 2020. https://www.businessofgovernment.org/sites/default/files/Algorithmic%20Auditing.pdf.

Kearns, Michael, and Aaron Roth. "Ethical Algorithm Design Should Guide Technology Regulation." Brookings (blog), January 13, 2020. https://www.brookings.edu/research/ethical-algorithm-design-should-guide-technology-regulation/.

Keller, Daphne. "Five Big Problems with Canada's Proposed Regulatory Framework for 'Harmful Online Content.'" Tech Policy Press, August 31, 2021. https://techpolicy.press/five-big-problems-with-canadas-proposed-regulatory-framework-for-harmful-online-content/.

Kim, Soomin, Changhoon Oh, Won Ik Cho, Donghoon Shin, Bongwon Suh, and Joonhwan Lee. "Trkic G00gle: Why and How Users Game Translation Algorithms." Proceedings of the ACM on Human-Computer Interaction 5, no. CSCW2 (2021): 1–24.

Kimmich, Christian, and Ferdinand Wenzlaff. "The Structure–Agency Relation of Growth Imperative Hypotheses in a Credit Economy." New Political Economy 27, no. 2 (2022): 277–95.

Kirchgaessner, Stephanie, and Jason Burke. "Rwanda Dissidents Suspect Paul Rusesabagina Was Under Surveillance." The Guardian, September 3, 2020. https://www.theguardian.com/world/2020/sep/03/rwanda-dissidents-suspect-paul-rusesabagina-was-under-surveillance.

Kitchens, Brent, Steven L. Johnson, and Peter Gray. "Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption." MIS Quarterly 44, no. 4 (2020): 1619–49.

Knott, Alistair, Kate Hannah, Dino Pedreschi, Tapabrata Chakraborti, Sanjana Hattotuwa, Andrew Trotman, and Ricardo Baeza-Yates. "Responsible AI for Social Media Governance: A Proposed Collaborative Method for Studying the Effects of Social Media Recommender Systems on Users." The Global Partnership on Artificial Intelligence, 2021. https://gpai.ai/projects/responsible-ai/social-media-governance/responsible-ai-for-social-media-governance.pdf.

Koene, Ansgar, Chris Clifton, Yohko Hatada, Helena Webb, and Rashida Richardson. "A Governance Framework for Algorithmic Accountability and Transparency." European Parliamentary Research Service, 2019. https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf.

Krause, Keith. "The Key to a Powerful Agenda, If Properly Delimited." Security Dialogue 35, no. 3 (September 1, 2004): 367–68. https://doi.org/10.1177/096701060403500324.

Kroll, Joshua A., Joanna Huey, Solon Barocas, Edward Felten, Joel Reidenberg, David Robinson, and Harlan Yu. "Accountable Algorithms." University of Pennsylvania Law Review 165 (2017): 633–705.

Kumar, Puneet, Dharminder Kumar, and Narendra Kumar. "E-Governance in India: Definitions, Challenges and Solutions." International Journal of Computer Applications 101, no. 16 (2014). https://ssrn.com/abstract=2501127.

LaFree, Gary, Laura Dugan, and Erin Miller. Putting Terrorism in Context: Lessons from the Global Terrorism Database. New York: Routledge, 2015.

Leaning, Jennifer, and Sam Arie. Human Security: A Framework for Assessment in Conflict and Transition. Cambridge, MA: Harvard Center for Population and Development Studies, 2000.

Lederach, John. The Little Book of Conflict Transformation. New York: Good Books, 2003.

Ledwich, Mark, and Anna Zaitsev. "Algorithmic Extremism: Examining YouTube's Rabbit Hole of Radicalization." ArXiv, 2019. http://arxiv.org/abs/1912.11211.

Lehr, Amy. "A Human Rights-Based Approach to AI." Center for Strategic & International Studies (podcast), 2019. https://www.csis.org/podcasts/humanity-wired/human-rights-based-approach-ai.

Lehr, David, and Paul Ohm. "Playing with the Data: What Legal Scholars Should Learn about Machine Learning." UCDL Rev. 51 (2017): 653–717.

LeVan, A. Carl. "Sectarian Rebellions in Post-Transition Nigeria Compared." Journal of Intervention and Statebuilding 7, no. 3 (2013): 335–52.

Lévy, Pierre. Becoming Virtual, Reality in the Digital Age. New York: Plenum Trade, 1998.

Lewis-Kraus, Gideon. "How Harmful Is Social Media?" The New Yorker, June 3, 2022. https://www.newyorker.com/culture/annals-of-inquiry/we-know-less-about-social-media-than-we-think.

Lieberman, Henry. "Interaction Is the Key to Machine Learning Applications." In Workshop on Learning from Examples and Programming by Demonstration, International Conference on Machine Learning. Lake Tahoe, California, 1995. https://web.media.mit.edu/~lieber/Lieberary/AI/Interaction-Is/Interaction-Is.html.

Linzer, Isabel. "Digital Technology Helps Governments Target Critics Across Borders." Slate, February 24, 2021. https://slate.com/technology/2021/02/paul-rusesabagina-rwanda-trial-digital-technology-critics-abroad.html.

Macdonald, Geoffrey, and Luke Waggoner. "Dashed Hopes and Extremism in Tunisia." Journal of Democracy 29, no. 1 (2018): 126–40.

Magalhães, João Carlos. "Do Algorithms Shape Character? Considering Algorithmic Ethical Subjectivation." Social Media+Society 4, no. 2 (2018): 1–10.

Magen, Amichai. "Fighting Terrorism: The Democracy Advantage." Journal of Democracy 29, no. 1 (2018): 111–25.

Malone, John Jack. "Examining the Rise of Right Wing Populist Parties in Western Europe." College of Saint Benedict/Saint John's University, 2014. https://digitalcommons.csbsju.edu/cgi/viewcontent.cgi?article=1044&context=honors_theses.

Martin, Mary, and Taylor Owen. "The Second Generation of Human Security: Lessons from the UN and EU Experience." International Affairs 86, no. 1 (2010): 211–24.

Mayne, Ruth, Duncan Green, Irene Guijt, Martin Walsh, Richard English, and Paul Cairney. "Using Evidence to Influence Policy: Oxfam's Experience." Palgrave Communications 4, no. 1 (2018): 1–10.

McClain, Colleen. "56% of Americans Oppose the Right to Sue Social Media Companies for What Users Post." Pew Research Center (blog), July 1, 2021. https://www.pewresearch.org/fact-tank/2021/07/01/56-of-americans-oppose-the-right-to-sue-social-media-companies-for-what-users-post/.

McCormack, Tara. "Power and Agency in the Human Security Framework." Cambridge Review of International Affairs 21, no. 1 (2008): 113–28.

McKay, Alistair. "The Arab Spring of Discontent." E-International Relations, 2011. http://www.e-ir.info/wp-content/uploads/arab-spring-collection-e-IR.pdf.

Menczer, Filippo. "How 'Engagement' Makes You Vulnerable to Manipulation and Misinformation on Social Media." The Conversation, September 20, 2021. http://theconversation.com/how-engagement-makes-you-vulnerable-to-manipulation-and-misinformation-on-social-media-145375.

Miller, Seumas. "Social Institutions." The Stanford Encyclopedia of Philosophy, 2019. https://plato.stanford.edu/archives/sum2019/entries/social-institutions/.

Miller-Idriss, Cynthia. The Extreme Gone Mainstream: Commercialization and Far Right Youth Culture in Germany. Princeton, NJ: Princeton University Press, 2018.

Mills, Albert, Durepos Gabrielle, and Wiebe Elden. Encyclopedia of Case Study Research. Los Angeles, CA: SAGE Publications, 2011.

Minh, Dang, H. Xiang Wang, Y. Fen Li, and Tan N. Nguyen. "Explainable Artificial Intelligence: A Comprehensive Review." Artificial Intelligence Review 55 (2022): 3503–68.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. "Model Cards for Model Reporting." In Proceedings of the Conference on Fairness, Accountability, and Transparency, 220–29, 2019.

Mittelstadt, Brent. "Principles Alone Cannot Guarantee Ethical AI." Nature Machine Intelligence 1, no. 11 (2019): 501–7.

Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. "The Ethics of Algorithms: Mapping the Debate." Big Data & Society 3, no. 2 (2016): 2053951716679679.

Moat, Kaelan A., John N. Lavis, and Julia Abelson. "How Contexts and Issues Influence the Use of Policy-Relevant Research Syntheses: A Critical Interpretive Synthesis." The Milbank Quarterly 91, no. 3 (2013): 604–48.

Müller, Marion G., and Celina Hübner. "How Facebook Facilitated the Jasmine Revolution. Conceptualizing the Functions of Online Social Network Communication." Journal of Social Media Studies 1, no. 1 (2014): 17–33.

Mumford, Lewis. Technics and Civilization. New York: Harcourt, Brace & World, 1963.

———. The Myth of the Machine: Technics and Human Development. Vol. 1. London: Secker & Warburg, 1967.

Munger, Kevin, and Joseph Phillips. "A Supply and Demand Framework for YouTube Politics." Unpublished paper, 2019. https://Osf. Io/73jys.

Nutley, Sandra M., Isabel Walter, and Huw T. O. Davies. Using Evidence: How Research Can Inform Public Services. Bristol, UK: Policy Press, 2007.

"Obama Tells Letterman How Algorithms Undermined Political Promise of Social Media." MarketWatch, January 18, 2018. https://www.marketwatch.com/story/obama-tells-letterman-how-algorithms-undermined-political-promise-of-social-media-2018-01-17.

O'Callaghan, Derek, Derek Greene, Maura Conway, Joe Carthy, and Pádraig Cunningham. "Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems." Social Science Computer Review 33, no. 4 (2015): 459–78.

Onbaşi Gençoğlu, Funda. "Social Media and the Kurdish Issue in Turkey: Hate Speech, Free Speech and Human Security." Turkish Studies 16, no. 1 (2015): 115–30.

Onwuegbuzie, Anthony J., Nancy L. Leech, and Kathleen M. T. Collins. "Innovative Data Collection Strategies in Qualitative Research." Qualitative Report 15, no. 3 (2010): 696–726.

Panagia, Davide. "The Algorithm Dispositif (Notes towards an Investigation)." UCLA School of Law: Program on Understanding Law, Science, & Evidence, 2019. https://escholarship.org/uc/item/154618gr.

Papadamou, Kostantinos, Savvas Zannettou, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Michael Sirivianos. "'It Is Just a Flu': Assessing the Effect of Watch History on YouTube's Pseudoscientific Video Recommendations." In Proceedings of the International AAAI Conference on Web and Social Media, 16: 723–34, 2020.

Paris, Roland. "Human Security: Paradigm Shift or Hot Air?" International Security 26, no. 2 (2001): 87–102.

———. "Still an Inscrutable Concept." Security Dialogue 35, no. 3 (2004): 370–72.

The Parliament of the Commonwealth of Australia. Treasury Laws Amendment (News Media and Digital Platforms Mandatory Bargaining Code) Bill 2020, 2020 § (2020). https://parlinfo.aph.gov.au/parlInfo/download/legislation/bills/r6652_first-reps/toc_pdf/20177b01.pdf;fileType=application%2Fpdf.

Pasquale, Frank. The Black Box Society: The Secret Algorithms That Control Money and Information. Cambridge, MA: Harvard University Press, 2015.

Pavliscak, Pamela. "How We Game the Algorithm to Tame the Algorithm." Medium (blog), May 19, 2016. https://medium.com/@paminthelab/how-we-game-the-algorithm-to-tame-the-algorithm-99f287d81a3b.

Perlis, Roy H., Jon Green, Matthew Simonson, Katherine Ognyanova, Mauricio Santillana, Jennifer Lin, and Alexi Quintana. "Association between Social Media Use and Self-Reported Symptoms of Depression in US Adults." JAMA Network Open 4, no. 11 (2021): e2136113–e2136113.

Prokupecz, Shimon, Christina Maxouris, Dakin Andone, Samantha Beech, and Amir Vera. "Payton Gendron: What We Know about the Buffalo Supermarket Shooting Suspect." CNN, May 15, 2022. https://www.cnn.com/2022/05/15/us/payton-gendron-buffalo-shooting-suspect-what-we-know/index.html.

Raji, Inioluwa Deborah, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. "Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing." In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 33–44, 2020.

Reece, Andrew G., and Christopher M. Danforth. "Instagram Photos Reveal Predictive Markers of Depression." EPJ Data Science 6, no. 1 (2016): 15.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1 § (2016).

Reynolds, Sean C., and Mohammed M. Hafez. "Social Network Analysis of German Foreign Fighters in Syria and Iraq." Terrorism and Political Violence 31, no. 4 (2017): 661–86.

Ribeiro, Manoel Horta, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, and Wagner Meira, Jr. "Auditing Radicalization Pathways on YouTube." In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 131–41, 2019.

Rich, Michael L. "Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment." University of Pennsylvania Law Review 164 (2016): 871–929.

Rodrigues, Rowena. "Legal and Human Rights Issues of AI: Gaps, Challenges and Vulnerabilities." Journal of Responsible Technology 4 (2020): 100005.

Roff, Heather M. Advancing Human Security through Artificial Intelligence. London: Chatham House, 2017.

Rowa, Yvonne. "Liminal Boundaries and Vulnerabilities to Radicalisation in the Context of Securitisation of Migration." PhD Thesis, University of Adelaide, 2019.

Rowa, Yvonne Jazz. "Part 1: Algorithmic Deconstruction in the Context of Online Extremism." GNET (blog), September 15, 2020. https://gnet-research.org/2020/09/15/part-1-algorithmic-deconstruction-in-the-context-of-online-extremism/.

———. "Part 2: Algorithmic Agency in Online Extremism: The Bigger Picture." GNET (blog). September 21, 2021. https://gnet-research.org/2020/09/21/part-2-algorithmic-agency-in-online-extremism-the-bigger-picture/.

Rudin, Cynthia, and Joanna Radin. "Why Are We Using Black Box Models in AI When We Don't Need to? A Lesson from an Explainable AI Competition." Harvard Data Science Review 1, no. 2 (2019). https://doi.org/10.1162/99608f92.5a8a3a3d.

"The Russians and Ukrainians Translating the Christchurch Shooter's Manifesto." Bellingcat, August 14, 2019. https://www.bellingcat.com/news/uk-and-europe/2019/08/14/the-russians-and-ukrainians-translating-the-christchurch-shooters-manifesto/.

Salmela, Mikko, and Christian Von Scheve. "Emotional Roots of Right-Wing Political Populism." Social Science Information 56, no. 4 (2017): 567–95.

Samin, Nadav. "Saudi Arabia, Egypt, and the Social Media Moment." Arab Media & Society 15, no. 1 (2012): 46–65.

Schindler, Hans-Jakob. "Emerging Challenges for Combating the Financing of Terrorism in the European Union: Financing of Violent Right-Wing Extremism and Misuse of New Technologies." Global Affairs 7, no. 5 (2021): 795–812.

Schlosser, Markus. "Agency." The Stanford Encyclopedia of Philosophy, n.d. https://plato.stanford.edu/archives/win2019/entries/agency/.

Schmitt, Josephine B., Diana Rieger, Olivia Rutkowski, and Julian Ernst. "Counter-Messages as Prevention or Promotion of Extremism?!: The Potential Role of Youtube Recommendation Algorithms." Journal of Communication 68, no. 4 (2018): 780–808.

Schuurman, Bart. "Topics in Terrorism Research: Reviewing Trends and Gaps, 2007-2016." Critical Studies on Terrorism 12, no. 3 (2019): 463–80.

Scrivens, Ryan. "Examining Online Indicators of Extremism among Violent and Non-Violent Right-Wing Extremists." Terrorism and Political Violence, 2022, 1–21.

Selbst, Andrew D., Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. "Fairness and Abstraction in Sociotechnical Systems." In Proceedings of the Conference on Fairness, Accountability, and Transparency, 59–68, 2019.

Sewell Jr., William H. "A Theory of Structure: Duality, Agency, and Transformation." American Journal of Sociology 98, no. 1 (1992): 1–29.

Soucy, Robert. "Barrès and Fascism." French Historical Studies 5, no. 1 (1967): 67–97.

Sovacool, Benjamin K., Xiaojing Xu, Gerardo Zarazua De Rubens, and Chien-Fei Chen. "Social Media and Disasters: Human Security, Environmental Racism, and Crisis Communication in Hurricane Irma Response." Environmental Sociology 6, no. 3 (2020): 291–306.

Stone, Peter. "US Far-Right Extremists Making Millions via Social Media and Cryptocurrency." The Guardian, March 10, 2021. https://www.theguardian.com/world/2021/mar/10/us-far-right-extremists-millions-social-cryptocurrency.

Striegher, Jason-Leigh. "Violent-Extremism: An Examination of a Definitional Dilemma." In 8th Australian Security and Intelligence Conference, Held from the 30 November – 2 December, 2015, 75–86. Edith Cowan University Joondalup Campus, Perth, Australia: SRI Security Research Institute, 2015. https://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1046&context=asi.

"Tech Against Terrorism Annual Report 2020–2021." Tech Against Terrorism, August 2021. https://www.techagainstterrorism.org/wp-content/uploads/2021/09/TAT-ANNUAL-REPORT_2020-21%E2%80%93FINAL.pdf.

"Tech Against Terrorism Calls for Removal of Taliban-Produced Content on Online Platforms, Noting Challenges for Smaller Tech Companies." Tech Against Terrorism, 2021. https://www.techagainstterrorism.org/2021/08/24/statement-tech-against-terrorism-calls-for-removal-of-taliban-produced-content-on-online-platforms-noting-challenges-for-smaller-tech-companies/.

"Tech Against Terrorism's Assessment of New Zealand's Digital Violent Extremism Report." Tech Against Terrorism, May 18, 2022. https://www.techagainstterrorism.org/2022/05/18/tech-against-terrorisms-assessment-of-new-zealands-digital-violent-extremism-report/.

Thompson, Robin. "Radicalization and the Use of Social Media." Journal of Strategic Security 4, no. 4 (2011): 167–90.

Thorburn, Luke, Jonathan Stray, and Priyanjana Bengani. "What Will 'Amplification' Mean in Court?" Tech Policy Press, May 19, 2022. https://techpolicy.press/what-will-amplification-mean-in-court/.

Tutt, Andrew. "An FDA for Algorithms." Admin. L. Rev. 69 (2016): 83.

"2021 Digital Violent Extremism Transparency Report." Department of Internal Affairs, New Zealand Government, March, 2022. https://www.dia.govt.nz/diawebsite.nsf/Files/Countering-violent-extremism-online/$file/DVE-Transparency-Report-2021-a.pdf.

Udupa, Sahana. "Enterprise Hindutva and Social Media in Urban India." Contemporary South Asia 26, no. 4 (2018): 453–67.

"Ugandan Intelligence Confirm Kakwenza Rukirabashaija Is Rwandan Agent." Kampala Post, January 4, 2022. https://kampalapost.com/content/ugandan-intelligence-confirm-kakwenza-rukirabashaija-rwandan-agent.

Valentini, Daniele, Anna Maria Lorusso, and Achim Stephan. "Onlife Extremism: Dynamic Integration of Digital and Physical Spaces in Radicalization." Frontiers in Psychology 11 (2020): 524.

Valkenburg, Patti M. "Social Media Use and Well-Being: What We Know and What We Need to Know." Current Opinion in Psychology 45; 101294, 2021.

Veblen, Thorstein. The Engineers and the Price System. New York: Kelley, 1965.

———. Imperial Germany and the Industrial Revolution. Livingston, NJ: Transaction Publishers, 1990.

———. The Instinct of Workmanship and the State of the Industrial Arts. New York: Augustus Kelley, 1914/1990.

Vigderman, Aliza, and Gabe Turner. "The Data Big Tech Companies Have On You." Security.Org (blog), 23 2022. https://www.security.org/resources/data-tech-companies-have/.

Vilone, Giulia, and Luca Longo. "Explainable Artificial Intelligence: A Systematic Review." ArXiv, 2020, https://arxiv.org/abs/2006.00093.

Vincent, Nicholas, Hanlin Li, Nicole Tilly, Stevie Chancellor, and Brent Hecht. "Data Leverage: A Framework for Empowering the Public in Its Relationship with Technology Companies." In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 215–27, 2021.

Visentin, Lisa. "NSW Election 2019: Labor's Michael Daley Claims Foreigners Taking Young People's Jobs." The Sydney Morning Herald, March 18, 2019. https://www.smh.com.au/nsw-election-2019/michael-daley-claims-foreigners-taking-young-people-s-jobs-20190318-p51591.html.

Wallace, Nick. "EU's Right to Explanation: A Harmful Restriction on Artificial Intelligence." TechZone360, 2017. https://www.techzone360.com/topics/techzone/articles/2017/01/25/429101-eus-right-explanation-harmful-restriction-artificial-intelligence.htm.

Welsh, Mike. "The Future is Phygital: Physical and Digital," Mobiquity, April 19, 2021, https://www.mobiquity.com/insights/the-future-is-phygital.

"What Is the Difference Between CPS and IoT?" Vanderbilt School of Engineering, February 28, 2022. https://blog.engineering.vanderbilt.edu/what-is-the-difference-between-cps-and-iot.

Whittaker, Joe. "Recommendation Algorithms and Extremist Content: A Review of Empirical Evidence." Global Internet Forum to Counter Terrorism, 2022.

———. "The Online Behaviors of Islamic State Terrorists in the United States." Criminology & Public Policy 20, no. 1 (2021): 177–203.

Whittaker, Joe, Seán Looney, Alastair Reed, and Fabio Votta. "Recommender Systems and the Amplification of Extremist Content." Internet Policy Review 10, no. 2 (2021): 1–29.

Wieviorka, Michel. "ETA and Basque Political Violence." In The Legitimization of Violence, edited by David E. Apter, 292–349. Springer, 1997.

Wiktorowicz, Quintan. Radical Islam Rising: Muslim Extremism in the West. Oxford, UK: Rowman & Littlefield Publishers, 2005.

Winner, Langdon. Autonomous Technology: Technics-Out-of-Control as a Theme in Political Thought. Cambridge, MA: MIT Press, 1977.

Woolley, Samuel C., and Philip N. Howard. Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media. Oxford, UK: Oxford University Press, 2018.

"Working Group on Infodemics: Policy Framework. Forum on Information & Democracy." Forum on Information & Democracy, 2020. https://informationdemocracy.org/wp-content/uploads/2020/11/ForumID_Report-on-infodemics_101120.pdf.

Youmans, William Lafi, and Jillian C. York. "Social Media and the Activist Toolkit: User Agreements, Corporate Interests, and the Information Infrastructure of Modern Social Movements." Journal of Communication 62, no. 2 (2012): 315–29.

Zarsky, Tal. "The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making." Science, Technology, & Human Values 41, no. 1 (2016): 118–32.

Zorthian, Julia. "Washington Wants to Regulate Facebook's Algorithm. That Might Be Unconstitutional," Time, October 13, 2021 https://time.com/6106643/facebook-algorithm-regulation-legal-challenge/.