

# **Lessons for Extremist Content Moderation: Capturing a Visual Style of Extreme Right Imagery**

Connor Rees

Submitted to Swansea University in fulfilment of the requirements for the Degree of Doctor of Philosophy in Computer Science

Swansea University

2025

## **Summary (Abstract)**

Online platforms have been subject to a recent surge in legislation from governments across the globe that mandate the swift detection and removal of harmful content (including extremist content) from their platforms, often under tight deadlines and with substantial fines for non-compliance. This trend, which opposes the historical self-regulation by online platforms, stems from several factors, not least being the apparent connection between online extremist rhetoric and real-world violence. Identifying and eliminating extremist content poses significant challenges, particularly concerning material from the extreme right. The extreme right frequently operates on the fringes of tolerable communication and utilises tactics like dog whistles (hiding secret messages within content such as imagery). This research focuses on extreme right imagery due to how challenging it can be to moderate. Research that analyses the extreme right has historically received less scholarly attention compared to other forms of extremism, namely Islamist extremism. The study involves an analysis of a dataset of designated extreme right group images taken from the encrypted messenger platform Telegram. This dataset is used in a mixed methods approach to iteratively design an extreme right image analysis framework. The purpose of this framework is to characterise the unique qualities and identifiers of extreme right content. The results and findings from creating the framework and using it to assess over 25000 images used to inform data-driven policy recommendations for online platforms to enhance their moderation practices. In doing so, this thesis aims to contribute new knowledge to the field and aid in moderation against the extreme right whilst protecting the freedom of expression of other online platform service users.

**Declarations**

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed.....  .....

Date..... 24/02/2025 .....

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed.....  .....

Date..... 24/02/2025 .....

I hereby give consent for my thesis, if accepted, to be available for electronic sharing

Signed..  .....

Date..... 24/02/2025 .....

The University's ethical procedures have been followed and, where appropriate, that ethical approval has been granted.

Signed.....  .....

Date..... 24/02/2025 .....

# Table of Contents

<b>Chapter 1: Introduction.....</b>	<b>9</b>
<b>Research Context Overview.....</b>	<b>11</b>
<b>Research Objectives.....</b>	<b>14</b>
<b>Structure and Methodology.....</b>	<b>16</b>
<b>Originality and Contribution.....</b>	<b>18</b>
<b>Chapter 2: Extreme Right Online Propaganda and Content Moderation .....</b>	<b>20</b>
<b>Contextualising and Defining Online Extreme Right Imagery .....</b>	<b>20</b>
Understanding Radicalisation .....	23
The Extreme Right.....	26
Online Extremism .....	33
Case Study: The Proud Boys.....	39
<b>Challenges of Extreme Right Content Moderation.....</b>	<b>43</b>
Modern Content Moderation Systems .....	45
The Core Problem: A Disjointed Policy Landscape.....	53
Platform Responsibilities, Limitations, and Support .....	57
Balancing Human Rights.....	63
<b>Conclusion .....</b>	<b>65</b>
<b>Chapter 3: Existing Studies Analysing Extremist Images .....</b>	<b>67</b>
<b>Introduction .....</b>	<b>67</b>
<b>Studies that Analyse Extremist Images Online.....</b>	<b>68</b>
<b>Methodologies and Findings.....</b>	<b>72</b>
Language and Imagery in Extremist Analysis .....	73
Theory-Driven Frameworks and Narrative Framing.....	76
Data-Driven Approaches: Grounded Theory and Inductive Analysis .....	77
<b>Innovative Approaches and the Unique Challenges of Extreme Right Content.....</b>	<b>79</b>
The Adaptation of Baele et al.'s Framework.....	81
<b>Conclusion .....</b>	<b>82</b>
<b>Chapter 4: Framework Development.....</b>	<b>83</b>
<b>Introduction .....</b>	<b>83</b>
<b>Building the Framework.....</b>	<b>83</b>
<b>Conclusion.....</b>	<b>92</b>
<b>Chapter 5: The Method.....</b>	<b>92</b>
<b>Introduction .....</b>	<b>92</b>
<b>Data and Sample.....</b>	<b>92</b>
<b>Extreme Right Groups .....</b>	<b>96</b>
<b>Measures.....</b>	<b>99</b>
<b>Analytic Strategy.....</b>	<b>100</b>
<b>Conclusion .....</b>	<b>103</b>
<b>Chapter 6: Results.....</b>	<b>103</b>

<b>Introduction .....</b>	<b>103</b>
<b>Dataset Characteristics.....</b>	<b>104</b>
<b>Duration .....</b>	<b>106</b>
<b>Frequency.....</b>	<b>107</b>
<b>Master Frames.....</b>	<b>109</b>
<b>Narratives.....</b>	<b>110</b>
<b>Shock Value .....</b>	<b>111</b>
<b>Image Format .....</b>	<b>112</b>
<b>Symbols.....</b>	<b>113</b>
<b>Conclusion .....</b>	<b>116</b>
<b><i>Chapter 7: Discussion.....</i></b>	<b><i>116</i></b>
<b>Introduction .....</b>	<b>116</b>
<b>RQ1: To what extent is imagery used as a medium of communication by extreme right groups?</b> <b>.....</b>	<b>117</b>
<b>RQ2: To what extent is graphic imagery used in the communicative strategies within extreme right imagery? .....</b>	<b>120</b>
<b>RQ3: Is there an extreme right visual style that is communicated and reinforced through the use of imagery?.....</b>	<b>124</b>
<b>RQ4: Do extreme right groups have differing visual styles that are communicated and reinforced through the use of imagery?.....</b>	<b>128</b>
<b>Policy Recommendations .....</b>	<b>133</b>
Policy Recommendation 1: Developing Knowledge Through Research .....	134
Policy Recommendation 2: Utilising Extreme Right Symbols for Content Moderation .....	140
Policy Recommendation 3: Increasing Understanding of Extreme Right Groups .....	143
<b>Conclusion .....</b>	<b>148</b>
<b><i>Chapter 8: Conclusion .....</i></b>	<b><i>149</i></b>
<b>A New Method for Understanding Extreme Right Imagery .....</b>	<b>150</b>
<b>Key Findings.....</b>	<b>151</b>
<b>Contributions of Study to Future Applications.....</b>	<b>154</b>
<b>Limitations and Future Work .....</b>	<b>155</b>
<b><i>Appendix I.....</i></b>	<b><i>158</i></b>
<b><i>Bibliography.....</i></b>	<b><i>161</i></b>

## **Acknowledgements**

Firstly, I am thankful to my family and friends for motivating, inspiring, and believing in me, you have been my bedrock. To Angharad, I am eternally grateful for your love, kindness, and belief in me. And to the Devereux family, who have continuously shown me love and support.

I want to express my gratitude to Swansea University, the EPSRC, and Meta for granting me the opportunity to pursue this thesis. Special thanks go to Stuart for encouraging me to apply and to Matt for taking a chance on me. Throughout my academic journey, the support and encouragement from the faculty and staff have been invaluable. I am especially thankful to the Centre for Doctoral Training (CDT) for the continuous guidance and professional development opportunities. I am grateful to Cohort 1, the friendship and support I received during a challenging master's program helped me discover new boundaries.

To the Criminology Department at Swansea University where I began my studies, I want to express my thanks. If it were not for the amazing people who formed the department and the positive experience I felt, I could not have made it this far. In particular, I want to thank Lella for her time, attention, and unwavering belief in me and my potential from the very beginning of my academic career.

I am profoundly thankful to my supervisors, Lella and Bertie, for their guidance, expertise, and patience throughout this endeavour. The support, insights, and constructive feedback have been pivotal in shaping the direction of my thesis and developing my skills as a researcher. I consider myself fortunate to have had the privilege of growing under their mentorship.

To the friends I made along the way, I offer my thanks and appreciation. To Simon for always making the time, to Ben for the dips, and to Dean for the days on the mountain, and to the many others. Namely, Jakub, whose friendship, support, and positive demeanour have made it impossible not to enjoy the journey from beginning to end.

Lastly, I want to acknowledge my late father, Anthony, whose memory always serves as my inspiration. I hope to make him proud and strive to emulate the example he set during his life.

## List of Tables and Figures

### List of Tables

Table 1: Extreme Right Dataset’s Groups, Telegram Groups, and Unique Telegram Groups Code.....	96
Table 2: Master Frame Results (%).....	110
Table 3: Narratives Results (%).....	111
Table 4: Shock Value Results (%).....	112
Table 5: Image Format Results (%).....	113
Table 6: Symbol Results (%).....	115

### List of Figures

<i>Figure 2: TAT's Designation of Extreme Right Entities Captured by the TCAP in 2022 .....</i>	<i>96</i>
<i>Figure 3: The Number of Unique Images per Extremist Right Group Telegram Group .....</i>	<i>105</i>
<i>Figure 4: The Number of Images Posted by the Combined Telegram groups of Each Extreme Right Group .....</i>	<i>106</i>
<i>Figure 5: The Number of Days Each Extremist Group Telegram Groups Was Live at Data Collection.....</i>	<i>107</i>
<i>Figure 6: The Average Frequency of Posts Per Day Per Telegram Group.....</i>	<i>108</i>
<i>Figure 7: The Number of Images, the Duration a Telegram Group was Live, and the Frequency of Posts Per Day.....</i>	<i>109</i>

## Abbreviations

Artificial Intelligence	AI
Atomwaffen Division	AWD
Blood & Honour	B&H
Content Incident Protocol	CIP
Feuerkrieg Division	FKD
Global Internet Forum to Counter Terrorism	GIFCT
International Covenant on Civil and Political Rights	ICCPR
International Human Rights Law	IHRL
Islamic State of Iraq and Syria	ISIS
Machine Learning	ML
National Socialist Order	NSO
Research Question	RQ
Systemic Functional Multimodal Discourse Analysis	SF-MDA
Tech Against Terrorism	TAT
Terrorist Content Analytics Platform	TCAP
Terrorists and Violent Extremist	TVE
The Base	TB
The Proud Boys	PB
User-Generated Content	UGC

## Chapter 1: Introduction

Over the last 20 years, the removal of harmful content online has been a process which has undergone significant change. Throughout this period terrorists and violent extremists (TVE) have widely adopted the internet throughout this period and taken varied approaches to their use of the internet (Gill et al., 2017; Bright, 2018). It is understood by governments and academia alike that extremists have widely adopted the internet. Content moderation has been used as a tool to mitigate and minimise the extremist presence on the internet. Online platforms are at the forefront of this process and are, by extension, held to a high standard to carry out this task. However, removing extremist content from online platforms entirely and continually is a problem yet to be completely solved. In practice, a more realistic aim is to develop content moderation practices so that service users do not encounter TVE content, and in cases where users do, said users can quickly have the content removed. This highlights the significance and impact of this thesis on shaping content moderation practices.

Achieving consensus on which extremist groups and what content should be removed has in large part evaded governments and online platforms alike to date. Whilst there have been clear-cut cases such as such as ISIS and Al Qaeda whose graphic content and terrorist designation by Western nations provide a distinct and recognisable visual style it is the fringe cases that create the most complication. Regulators face the complex task of protecting service users from harmful content whilst preserving users' rights to share ideas and opinions. In this vein, there is no shortage of examples of potentially harmful individuals, groups, and ideologies that exist on online platforms, from Andrew Tate, to Incel, to QAnon just to name a few. While there is a notable quantity of research into these examples and their harms (Helm et al., 2024; Hoffman et al., 2020, Sayogie et al., 2023), this research looks at the moderation of extreme right content.

The extreme right collective presents a complex case whereby, it is difficult for online platforms to determine what speech and communication is acceptable and what is not. Individuals with extreme right beliefs have been linked to real-world harm, from being linked to acts of graffiti (Nouri and Morgan, 2023) to riots (Kydd, 2021) to mass shootings (Capellan, 2015). The place of extreme right ideology and rhetoric has continually shifted around the

Overton Window (the range of societally acceptable discourse (Mackinac Centre, 2024) and will likely change in the years to come. This link to harm and taboo public conversation makes the extreme right an important and relevant case study on moderating content that divides public opinion whilst being potentially harmful. In such cases, it is vital to accurately remove only the content that is deemed harmful and not to overextend moderation practices to infringe on user's rights. However, in the case of fringe topics like the extreme right, making a distinction between the content that should and should not be removed is a complex task that risks encroaching on and eroding users' freedom of expression.

Content moderation refers to the organised practice of assessing user-generated content (UGC) posted to an online platform to determine whether that content should be subsequently removed, or down-ranked, should it infringe on a platform's policies (Roberts, 2017). It is the stance of this thesis that improving and refining content moderation practices is a more realistic and achievable outcome compared to eliminating TVE activity. However, this is not without its unique set of complications. TVE content is inherently complex in several ways. What is understood as extremist is subject to when in time, where, and from what perspective it is being considered (Berger, 2018). Therefore, if content moderation were to be carried out at a single point in time, in a specific region, on a small scale with definable parameters, the task would be significantly more achievable. However, as a consequence of both the expectations of online platform users in combination with international regulation, online platforms are compelled to moderate themselves regardless of their resources and capacity, their level of understanding of the topic, and/or whether or not any external support is offered.

By creating a framework to analyse extreme right imagery this thesis provides critical insight into how extreme right imagery can be understood. This understanding will be used to suggest improvements that can be made to content moderation practices across numerous platforms to accurately capture extreme right content. In doing so, this thesis will offer multiple actionable data-driven policy recommendations in addition to a practical framework which can be expanded upon to keep up to date with new trends in extreme right behaviour. To create a framework to analyse extreme right imagery, over 25,000 images were collected from the encrypted messaging platform Telegram. Images were collected from this platform

due to its categorisation as a beacon platform; beacons are platforms that store, or signpost extremist content found in another location (Fisher et al., 2019). Therefore, by collecting images from Telegram the sample contains images likely spread across numerous other platforms. By using data from a beacon platform (Telegram) the policy recommendations drawn from this data will be relevant to a range of online platforms dealing with extreme right imagery beyond Telegram.

### Research Context Overview

A global rise in use and ease of access to the internet has correlated with nefarious actors, including extremists, quickly adopting the technology (Weimann, 2004). This trend became particularly apparent to wider society in the 2010s when social media established itself as an industry and extremist groups used savvy video and image editing to create online campaigns (Conway and Courtney, 2018). The terrorist organisation Islamic State of Iraq and Syria (ISIS) is a pertinent example of this in 2013-2014, which was touted as the ISIS Twitter 'Golden Age' (Conway et al., 2020). Showcasing highly sophisticated extremist propaganda across major social media platforms, the group's activities demanded a response. In turn, this led to online platforms receiving significant pressure from governments and institutions to moderate against harmful and extremist content in its many forms; an ever-present pressure. As a result, content moderation practices and standards have been curated to ensure service users are subject to as little harmful content as possible.

For governmental institutions, the process of creating legislation is lengthy. From the point of recognising and responding to a threat by working out adequate legislative options to formally introducing such legislation to getting the legislation passed, little scope is left for a timely response to such threats through traditional Western legal systems. Threats such as extremism are likely to have evolved during the legislative creation process, necessitating new types of responses and creating a cycle of action and reaction. With the absence of a third-party body to regulate online platforms during this period, the platforms themselves have been forced to write the playbook on extremist content moderation in the short to medium term (Koetsier, 2020). This process has taken place with a lack of support to platforms and an increasing presence of potential sanctions if deemed not to be carried out effectively enough. Examples of legislation which places sanctions on content moderation can be found primarily

in Western nations, including the German NetzDG (German Federal Ministry of Justice and Consumer Protection, 2017), the French Avia Law (EPRA, 2021), and Australian Criminal Code Amendment (Australia Legislation, 2019).

The above legislation has made it clear that the onus is placed on online platforms to moderate themselves. This approach has generated significant debate on whether the sanctions are fair, whether support should be offered to platforms and where the content moderation line starts and ends. A significant volume of the pressures, demands, and proposed sanctions aimed at the process and quality of content moderation are seemingly based on the false assumption that these problems are quick and easy to solve (Feamster, 2018). This has encouraged platforms to move away from difficult-to-scale solutions like employing human experts and content reviewers and toward more scalable automated systems where feasible. However, Douek surmises that content moderation is not a process that can be perfected, and it should not be approached as if it can (Douek, 2020). This problem is exacerbated by the fact that there is limited transparency when it comes to online platforms sharing content moderation practices, leaving little opportunity for work to be done to develop these technologies beyond the work done by the platforms themselves.

Content moderation varies in accuracy depending on depending on numerous factors including the context in which it is applied. In the cases of ISIS and Al-Qaeda-created content, Facebook has developed automated systems that can remove 99% of ISIS and Al-Qaeda content before it is uploaded or reported by service users, including suspect images, messages, posts, and accounts (BBC, 2017). However, these statistics are not publicised and are more difficult to locate in the case of the extreme right likely because the same accuracy figured cannot be replicated across different contexts. The extreme right has historically been found to widely use major platforms including Telegram, Instagram, 4chan and Reddit, among many others. They are adept at using a multitude of means of content creation including text, video, and audio to maximise their reach (Kimmel, 2017). Particularly it is the use of images that is the hallmark of the extreme right. These images and memes are used to communicate, express a collective identity, and reinforce political views through humour and satire (Decook, 2018). The extreme right uses these mediums to communicate an extremist discourse that is closely aligned with contemporary mainstream discourse and politics. It is this combination

of characteristics that makes the extreme right such a fringe and difficult-to-moderate demographic (Conway et al., 2019).

As a result of the complexity of content moderation, online platforms have evolved their approaches over the years. Zittrain separates digital governance into three eras: the rights era, the public health era, and the process era (Zittrain, 2019). With the large social media companies primarily being set up in the US, the rights era reflects the period where users' right to expression was the primary function. Hence ISIS and other extremist groups were able to capitalise and share their content without repercussion as a trade-off to protect a democratised speech culture. Then came the public health era where platforms began to consider the risks and benefits of allowing content such as extremism, leading to new policies which banned the use of hate speech and extremist associations. This came as a response to societal pressure to protect users online in the wake of the first era. According to Zittrain, we are now in an era where processes are implemented to individually assess content uploaded to online platforms, evaluate its contents, and decide whether the content should be removed, downranked, or kept online. These sentiments are shared by Douek who in the context of content moderation defines Zittrain's rights era as a "posts-as-trumps" era (otherwise meaning that users' posts and connectivity took precedence over all other factors), or the early content moderation era. Sharing Zittrain's findings of a public health era and process era, Douek argues for a second era defined by probability and proportionality in the governance of speech online (Douek, 2020).

This leaves the state of content moderation in a precarious position. There is increased pressure and expectation to carry out an 'imperfect process' with a high degree of accuracy concerning groups that have proved challenging to define, let alone identify. Simultaneously, the extreme right has been described as "difficult to moderate" due to its nuance and complexity. This, therefore, is where this thesis looks to contribute to solving the identified problem by producing a framework to understand and characterise extreme right imagery so that online platforms can use this information to improve extreme right content moderation practices. This framework is designed by analysing over 25,000 extreme right images collected from the online platform Telegram. Whilst the data analysis and framework development characterise the visual style of the extreme right, the findings from this analysis also inform

the creation of actionable policy recommendations for online platforms to consider to further improve and refine the moderation of the extreme right.

### Research Objectives

This thesis sets out to create a framework for online platforms to use to help discern what is or is not extreme right content i.e. what should or should not be allowed to remain on a platform based on whether it constitutes as being extreme right content. This is fundamentally the aim of the thesis because it is a highly complex issue to solve and, likely because of this, no such equivalent frameworks currently exist or are made publicly available.

Therefore, the place of the research questions that underpin this thesis is to extract key traits and characteristics of the extreme right, according to the dataset. The research questions are, therefore, an extension of the extreme right image analysis framework, answers to these questions are intended to help determine the traits and characteristics of the extreme right visual style so that this information can be used to accurately identify extreme right content. The research questions incorporate the themes of extreme right behaviour, communication, and use of imagery to gain a comprehensive understanding of the extreme right's visual style to achieve this aim.

The first research question asks, to what extent is imagery used as a medium of communication by extreme right groups? This question is pivotal to this thesis, as addressing it will help build a clear picture of the posting behaviours of a range of extreme right groups necessary to facilitate their moderation. This question fundamentally underpins the findings and the implications of the policy recommendations this thesis offers. There have been numerous pivotal works showcasing that the extreme right does utilise imagery as a means of communication and that the extreme right has a communication style which differs from other extremist groups. This finding is also present in the results, where there is significant variability in the usage of imagery and visual style within the extreme right, which makes its content difficult to identify and determine. Assessing the extent to which imagery is a medium used by the extreme right is necessary because it is a premise which the rest of the study relies upon. As such, without exploring this within the data, the framework along with the other contributions of this thesis would be undermined.

The second research question asks to what extent graphic imagery is used in the communicative strategies within extreme right imagery. This question is important to this thesis for two reasons. Firstly, for its ability to contribute to the field of knowledge and secondly, for its content moderation practice implications. The results and findings clarify that whilst graphic imagery is present within several extreme right groups' imagery, it is to a minimal extent. This finding does differ between extremist groups and within across a single groups array of Telegram groups, however, whilst being present within the dataset, it is not a definitive characteristic. Exploring the presence of graphic imagery within extreme right imagery is necessary as it tests existing understanding in the literature and provides the opportunity to suggest the prioritisation or de-prioritisation of specific content moderation practices.

The third research question asks, is there an extreme right visual style that is communicated and reinforced through the use of imagery? This is a pivotal research question within this thesis. It is simultaneously the most ambitious of the four research questions. The foundations which underpin this question are laid out in the literature review chapters. Within these, it is clear there are distinct extreme right visual characteristics communicated by the literature, but no definitive style consistent across the multiple forms of extremist groups encompassed within the extreme right umbrella. However, the data does provide a foundational visual style of the extreme right made up of five distinct characteristics. Exploring this research question is essential to achieving this thesis' aim of creating a collection of data-driven extreme right imagery characteristics for targeted and accurate content moderation.

The fourth and final research question asks, do extreme right groups have differing visual styles that are communicated and reinforced through the use of imagery? This research question builds on the knowledge established in the third research question, whereby the characteristics of extreme right visuals are found in the literature. However, the results of the study evidence distinct differences not only between extreme right groups but also within them. Whilst the previous question queries the presence of a collection of generic extreme right image characteristics, this question explores whether this is the case for specific groups.

Answering this question explores the feasibility of targeted content removal of specific groups.

The answers to these questions are threaded throughout this thesis. These research questions are covered in both the Literature Review and empirical analysis. As a result, this thesis establishes a base of understanding from the literature review. To assess this knowledge through empirical analysis, the methodology chapters illustrate the novel dataset and the design of a bespoke image analysis framework. Then the Results and Discussion chapters communicate the study's outputs, applying these to form new conclusions and contributions to the field, ultimately resulting in three policy recommendations for online platforms. The structure and research questions of this thesis are all targeted at aiding online platforms to help discern what is or is not extreme right content by exploring what potential extreme right imagery traits and characteristics can signpost moderation practices to accurately identify extreme right content.

### Structure and Methodology

This thesis follows a linear structure spread over a total of eight chapters, including the Introduction. These chapters can be grouped into the following categories: Introduction, Literature Review, Methodology, Results, Discussion, and Conclusion. Each chapter was selected to foster understanding and context around the four research questions that were identified and outlined in the research objective section of this introduction. The literature review chapters 'Extreme Right Online Propaganda' and 'Content Moderation and Existing Studies Analysing Extremist Images' establish key themes and findings that bring out new insights to form a bespoke image analysis framework. Two chapters form the methodology; these are 'Developing the Framework' and 'The Method'. These chapters use the gaps identified in the literature review to communicate how this study structurally approaches analysing and understanding extreme right imagery. The results are directly drawn from the bespoke methodology being applied to a novel dataset. This is where the novel dataset, bespoke framework, and understandings from the literature review are compiled to draw new conclusions and three policy recommendations.

Immediately following the Introduction is the literature review component of this thesis, which is split into two distinct chapters. The first chapter is the context and background chapter, the second is the challenges of extreme right content moderation chapter. The first section covers the extreme right, online propaganda, and why images are important. Having established the necessary context, the second section covers the challenges of extreme right content moderation which frames how this study looks to overcome some of these challenges.

Following the two sections that form the literature review are the two chapters that constitute this thesis' methodology. The literature review chapters present the complexity associated with categorising and identifying extreme right content, which makes moderation difficult. This, in turn, suggests the need for more work to be done to help categorise and identify such content. The methodology chapters present a novel approach to tackling this issue. This is achieved through the collection of a novel dataset and the creation of a bespoke image analysis framework. Looking first at the dataset, this component of the thesis is one of its central strengths through the collection of over 25,000 images from Telegram in the first quarter of 2021. This dataset has not previously been analysed and, as a result, offers insight into the contemporary landscape of many strands of the extreme right. These images were collected from 6 different extremist groups over 17 different Telegram groups. This dataset was collected for the single purpose of providing useful data to this study, meaning it contained the necessary characteristics. The three main dataset characteristics being referred to are the capturing of multiple extreme right groups, the presence of imagery, and the collection from a source not widely researched in the literature at the time.

Looking secondly at the framework, the data collected for this research were fed into the development of an extreme right image analysis framework. The purpose of this framework is first and foremost to explore the possibility of capturing a visual style of the extreme right, and secondly to answer the subsequent research questions of this thesis. This led to three iterative rounds of coding using a modified grounded theory approach. First, a review of the literature was undertaken to determine the key master frames i.e., the central themes of extreme right imagery. From this starting point, the data then determined the creation of

several more refined categories. This ultimately led to the creation of a framework comprised of five categories and 35 subcategories that categorise and capture extreme right imagery.

The Results chapter draws on the application of the novel dataset to the extreme right image analysis framework. This chapter takes the structure of being separated by each of the five categories and the 35 subcategories and provides a figure for each of the 17 extreme right Telegram groups. This therefore provides a broad overview of the results for each subcategory of each Telegram group. These results and their descriptions offer the raw data and subsequent analysis that underpins the contributions of this research.

The Discussion chapter draws on the data from the Results chapter in combination with the knowledge gathered in the literature review chapters. This chapter is structured by first identifying and answering each of the four research questions set out at the beginning of this thesis, extracting the relevant data from the literature and the results to evidence each answer given. After the research questions have been answered the findings drawn from these questions are used to develop and present three novel policy recommendations for online platforms to consider in response to the extreme right use of imagery. These key findings and conclusions are then positioned to consider future applications of the study and its format whilst summarising and reflecting on the thesis in the Conclusion chapter.

### Originality and Contribution

This thesis makes four key contributions to existing knowledge on the extreme right and content moderation. First, by synthesising existing research, a new understanding of the extreme right and the internet is presented. Secondly, by extracting data from the encrypted messaging platform Telegram, a novel dataset is created for empirical research. Thirdly, developing a new framework for extreme right imagery analysis, based on these findings offers new insights and understandings. Introducing this framework further creates scope for future work involving the application of other datasets. Finally, the creation of the policy recommendations offers actionable evidence-based policies for online platforms to consider incorporating into their systems and policies. These four contributions are characterised as empirical (new knowledge), data (new dataset), methodological (new framework), and policy (policy recommendations). As a result, this thesis offers contributions not only to offer new

insights into trends and behaviours of extreme right image use for research purposes but also to bridge the gap between academia and practice.

The methodological innovation in this thesis is the first of the four contributions. To meet this criterion, a bespoke theoretical framework was created for extreme right image analysis. The process of creating this framework was carried out in a robust and data-driven manner, utilising a novel dataset in addition to consulting contemporary research on the extreme right. The development of this framework offers a platform for future work, including potential adoption into the content moderation practices of online platforms.

The second contribution offered by this thesis is its data contribution. As mentioned, the theoretical framework for this thesis was created and refined through the analysis of a large dataset that captures many elements of the extreme right. The establishment of a new dataset is a significant contribution to the field. The characteristics of this novel dataset include the collection of over 25,000 images collected from six extreme right groups and 17 distinct groups on Telegram. The collection of this dataset contributes to the field's increased understanding of trends and characteristics of extreme right groups and their use of images. Such insights can be found in the Results chapter, whereby, from looking at the broader themes showcased in the dataset, there are novel insights into how the extreme right behaves in terms of the duration a group has had a presence on a platform and the image posting frequency. This unique dataset is a significant resource for future research on the extreme right.

The empirical contribution is the third contribution offered by this thesis. Through the novelty of a bespoke framework and dataset, the findings of this research offer new insight and knowledge to the field. Through data-driven and empirical reasoning, existing understandings of how the extreme right communicates online are developed and furthermore, an understanding of the extreme right visual style is offered. This contribution in turn gives increased certainty around the posting behaviours of a potentially dangerous population. These findings themselves can, directly and indirectly, support future content moderation solutions and responses through the creation of a better understanding of trends and behaviours.

The policy contribution is the last contribution offered by this research. Leveraging the data-driven results and findings, this thesis transforms these data into policy recommendations. These recommendations have been developed with online platforms in mind, with the intended purpose of being considered for incorporation into content moderation practices. Their credibility is linked to the novelty of the framework, dataset, and the strength of the existing body of literature which were used to inform them.

The identification of these four key contributions illustrates the strength and utility of this thesis. These contributions not only strengthen the field but also help bridge the gap between academia and practice.

## Chapter 2: Extreme Right Online Propaganda and Content Moderation

### Contextualising and Defining Online Extreme Right Imagery

The notion of extremism is relatively new, first emerging in Germany in the mid-20th century (Bötticher, 2017). Since then, numerous iterations and interpretations of the term have evolved. However, extremism remains an inherently complex concept, a fact underscored by the extensive academic literature critiquing, comparing, and proposing new definitions of the term (Coleman and Bartoli, 2003; Hassan et al., 2023; Lowe, 2017). This lack of a universal definition underscores the core issue of this thesis, the challenge of detecting extreme right imagery. The difficulties in defining extremism more broadly contribute to the complexities of identifying and categorising such content. Therefore, in the pursuit of clarity, it is essential to first define what is meant by 'extremism' within the context of this research. Moreover, as extremism, violent extremism, and terrorism are terms often used interchangeably in the field of preventing and countering violent extremism and related research (Stephens et al., 2019), a brief explanation of violent extremism and terrorism is also essential to establishing a clear analytical framework for the thesis. Clarity in this context is crucial because the purpose of this thesis is to develop a framework for detecting extreme right imagery that can be adapted by various platforms. Given that platforms often use differing definitions of terms such as extremism, violent extremism, and terrorism, applying the framework according to

inconsistent definitions could produce misleading results and conclusions. This would undermine one of the primary objectives of the framework: to promote greater clarity in understanding and moderating the diverse and overwhelming volume of such content.

The definitions used in this thesis will be grounded in existing literature on the topic, without delving into an exhaustive exploration of the ultimate meaning of extremism. While there may be value in exploring this topic in more detail than what is to come, it would constitute a separate thesis and fall outside the scope of this study's objective, which is to define extremism for the purposes of analysing extreme right imagery. The lack of consensus on the definition of extremism can be attributed to variations in inclusion and exclusion criteria influenced by factors such as the defining body (e.g., academia, industry, or government), disciplinary perspective (e.g., political science, criminology, or law), and geographical, temporal, and political contexts. These factors often lead to tensions when definitions are adapted or evolved, as illustrated by the recent expansion of the UK Government's definition of extremism (Quinn, 2024; UK Government, 2024). Despite these challenges, adhering to a specific definition is essential to ensure clarity and transparency, particularly in a study like this one, which focuses on analysing extreme right content in the form of imagery.

As alluded to, there are numerous definitions of extremism proposed by academics and industry alike. One of the most significant academic contributions towards achieving a consensus on this concept is J. M. Berger's definition, which describes extremism as "the belief that an in-group's success or survival can never be separated from the need for hostile action against an out-group. The hostile action must be part of the in-group's definition of success. Hostile acts can range from verbal attacks and diminishment to discriminatory behaviour, violence, and even genocide" (Berger, 2015, p. 44). This definition integrates the concept of physical violence into its understanding of extremism. However, this thesis adopts the view that distinguishing between extremism and violent extremism offers greater clarity, as it separates two often-interchangeable terms and presents and a spectrum of extremist actions. While Berger's definition highlights "hostile action," it leaves the concepts of the "in-group" and "out-group" open to interpretation. These constructs can be traced back to Social Identity Theory (Abrams & Hogg, 2006; Tajfel, 1981; Tajfel & Turner, 2004), which frames the "us versus them" perspective as being shaped by factors such as race, religion, or ideology

(Rasoulkolamaki et al., 2023). Berger applies these long-standing constructs specifically to extremism, defining the in-group as the group to which individuals belong and the out-group as those excluded from it (Berger, 2021). A practical example of this is ISIS, which defines its in-group as “Muslims” or the “camp of faith” and its out-group as “disbelievers” (Kufr), whom they hold responsible for a perceived crisis (Al-Tamimi, 2023; Ingram, 2017).

Using this framework, extremism can be understood as a belief system where the in-group’s survival is inseparable from hostile action against the out-group. For instance, ISIS’s success is ideologically tied to hostile actions against its perceived enemies, illustrating the inseparability of these components. Adopting this understanding, this thesis defines extremism as ‘the belief that an in-group’s success or survival cannot be separated from hostile action against an out-group, where hostile actions may range from verbal attacks to diminishment but exclude the necessity of physical violence’. This distinction facilitates the separation of extremism from violent extremism. Building on this, Berger (2017) defines violent extremism as the belief that an in-group’s success is inseparable from violence against an out-group. He emphasises that violent extremist ideologies often frame violence as defensive, offensive, or pre-emptive and describes violent extremism as requiring violence to achieve the in-group’s goals. This distinction ensures that violence is not incidental but intended and intrinsic to the ideology of violent extremism. For example, while extremism may include verbal or psychological tactics, violent extremism necessitates acts of physical violence.

A challenge arises when considering the overlap between extremism, violent extremism, and terrorism. To maintain clear distinctions, this thesis adopts Berger’s definition of violent extremism without modification while emphasising that extremism can exist without violence. In contrast, terrorism is better understood as a tactic rather than an ideology. This view aligns with the United States Department of State’s definition, which describes terrorism as “premeditated, politically motivated violence perpetrated against non-combatant targets by sub-national groups or clandestine agents” (United States Department of State, 2006). Schmid’s (2013) definition further supports this interpretation, framing terrorism as “a doctrine about the presumed effectiveness of a special form or tactic of fear-generating, coercive political violence.” This conceptualisation underscores the flexibility of terrorism as

a methodology that can be employed by violent extremist groups. For example, Antifa, is a far-left collective that opposes fascism and right-wing ideologies and engages in acts of violence but has not carried out direct or symbolic acts of terrorism (LaFree, 2018). Conversely, violent extremist groups like ISIS frequently employ terrorism as a key tactic. Thus, while violent extremism and terrorism may intersect, they are not synonymous. Violent extremism represents an ideology, while terrorism is a specific method used to advance ideological goals.

Understanding extremism, violent extremism, and terrorism through clear definitions and inclusion/exclusion criteria is essential for research that aims to capture the online imagery of a diverse and global network such as the extreme right. By adapting and refining Berger's definitions, this thesis establishes a coherent framework that distinguishes between these concepts. This foundation supports the exploration of the symbiotic relationships between extremism, violent extremism, and terrorism while providing a clear basis for building a framework to analyse extreme right imagery. These definitions will underpin the subsequent chapters and contribute to addressing the gaps in the existing literature.

### Understanding Radicalisation

Building upon the definitions of extremism and related terms, it is important to explore why individuals may ingest and produce extremist online content, through the process of radicalisation. Understanding this process is central to this thesis, as it highlights why extreme right imagery online is a potential societal issue. Such imagery plays a role in radicalising internet users, acting as both a medium and catalyst for the dissemination of extremist ideologies. This section clarifies the concept of radicalisation, outlines its complexities, and provides a foundation for later discussions on how content moderation systems can address the challenges posed by online radicalisation.

Radicalisation refers to the process through which an individual develops extremist views or behaviours (Neumann, 2013). Like extremism, the term lacks a universally agreed definition, adding to its complexity. The term radicalisation has evolved over time. Originating in the 18th century within the field of medicine, it was first applied to political contexts in the late

1790s (Bötticher, 2017). In the modern era, it gained prominence following the September 11th attacks, as governments sought to address homegrown terrorism and its underlying causes (Hafez & Mullins, 2015). Although earlier research explored pathways to extremism, the term radicalisation became more closely associated with these processes in the early 2000s. For instance, Martha Crenshaw's seminal 1981 paper, *The Causes of Terrorism*, identified three primary factors influencing an individual's path to extremism: individual motivation and belief systems, strategic decision-making within a terrorist movement, and the broader political and social context (Crenshaw, 1981). These findings remain relevant today, underscoring both the continuity and complexity of understanding radicalisation.

Radicalisation is often described as a process, though its pathways are rarely linear or uniform. A prominent example of this framing can be found in the UK government's counterterrorism strategy, CONTEST, launched in 2003 (HM Government, 2011, 2023). One of its four strands, Prevent, focuses specifically on tackling radicalisation and employs the concept of a 'radicalisation process.' Initially, Prevent was underpinned by the 'Conveyor Belt' theory, which depicts radicalisation as a linear progression from grievance to extremist violence (Sabir, 2016; Powell, 2016). While this model provides a simplified framework for understanding radicalisation, it has been criticised for its limitations, particularly its presumption that radicalisation inevitably culminates in violence (Francis, 2012). The Conveyor Belt theory's reductive nature fails to account for alternative outcomes, such as individuals adopting non-violent extremist ideologies or disengaging from the process entirely. Additionally, it overemphasises the role of ideology while neglecting the many social, psychological, and contextual factors that shape an individual's trajectory (Malthaner, 2017). These critiques highlight the need for more nuanced approaches to understanding radicalisation.

In response to the limitations of process-based models, more recent understandings frame radicalisation as a pathway. This metaphor allows for multiple routes, outcomes, and contributing factors, recognising the variability of individual experiences. For instance, Gill's pathway model documents the steps leading to specific extremist behaviours, such as becoming a suicide bomber (Gill, 2007; Christmann, 2012). Other models focus on broader influences, including predisposing risk factors (Horgan, 2009), network dynamics (Sageman,

2008), and socio-political mechanisms (McCauley and Moskalenko, 2011). These approaches move beyond rigid, linear frameworks, acknowledging the complex interplay of factors that contribute to radicalisation.

This thesis argues that extreme right imagery online can serve as a tool for radicalisation, influencing individuals through visual and symbolic representations of extremist ideologies. Such imagery often exploits socio-political grievances, reinforces in-group/out-group dynamics, and normalises extremist narratives. By disseminating these visuals on accessible platforms, extremist groups create pathways for radicalisation that may escalate to violent or non-violent extremism. Radicalisation, when viewed through the lens of this thesis, highlights the dangers of online content that goes un-moderated. Unlike traditional recruitment methods, digital platforms allow for the mass dissemination of extremist material, often targeted at vulnerable or impressionable individuals. Understanding radicalisation is therefore relevant for addressing the thesis's core question: how can content moderation systems effectively distinguish between what is and is not extreme right imagery?

The academic literature on radicalisation often emphasises the endpoint (committal of extremist behaviour) while overlooking the broader phenomenon of radicalism itself. Radicalism, defined as the pursuit of significant societal change through unconventional ideas, is not inherently negative. Historical figures like Martin Luther King Jr. and Rosa Parks were considered radicals in their time, yet their contributions are now celebrated (Gartenstein-Ross and Blackman, 2019). This distinction underscores the importance of not conflating radicalism with radicalisation or extremism. For the purposes of this thesis, radicalisation is defined as the process of progressing toward extremist ideologies or behaviours driven by a combination of social, psychological, and contextual factors. This thesis views that extreme right imagery can serve as one of these contributing factors by shaping how individuals perceive and interpret other influences in their lives. This nuanced understanding enables a more comprehensive analysis of the role extreme right imagery plays in fostering radicalisation online.

## The Extreme Right

Like other key terms in terrorism and extremism studies, defining the “extreme right” is inherently complex. This complexity arises from the diversity and in-group variability of extreme right collectives, which span ideologies, grievances, and socio-political contexts (Piazza, 2017; Van Prooijen and Kuijper, 2020; Scrivens et al., 2023). While the extreme right may appear, at first glance, to be a unified global movement, this perception oversimplifies a fragmented phenomenon. According to current understanding, the extreme right is a “melting pot” of groups, each shaped by geographical, cultural, and temporal factors, as well as their tolerance for ideological flexibility and cross-pollination (Bliuc et al., 2020; Caiani et al., 2012; Schain, 2006). Recent efforts to define and understand the extreme right are limited to a point in time. The extreme right can be characterised differently depending on when in time a definition is created. Caiani and Della Porta (2011) recently evidence this when identifying a shift in attributes by the extreme right in recent years from fascism to populism. Historical understandings of the extreme right were linked with associations with fascism, including ultra-nationalism, beliefs in societal decline, and the propagation of conspiracy theories (Eatwell, 2003; Caiani and Della Porta, 2011). More contemporary understandings around the extreme right harness associations with populism, these characteristics including nationalism, xenophobia, and authoritarianism (Mudde 2007). Populism can be understood in several ways; however, this term is understood as an ideology dividing society into two antagonistic groups, the “us” versus “them” or the in-groups versus the outgroup, advocating for politics to reflect the will of the people (Mudde 2004). In this light, a simplified understanding of the extreme right can be understood as collectives of extremist groups that share populist and right-wing ideologies.

From the perspective of this thesis, the term extreme right captures the overarching commonalities across extremist groups with conservative political ideologies. This is an approach of using one term to capture the remit of others that has been previously adopted by other authors. Leading scholar in the domain Cas Mudde uses the term far-right for the same purpose (Mudde, 2019, Ravndal and Bjorgo, 2018). Traditionally, these groups are captured and characterised by several terms that look to distinguish key differences between groups united by these traits. In this context, many terms look to achieve this, for example, the “Radical Right”, “Far-Right”, and “Right Wing Extremist” to name a select few. The radical right refers to an umbrella term for extreme and far-right groups that share a core ideology

that incorporates nativism, authoritarianism, and populism (Mudde, 2015; Nouri et al., 2020; Scrivens et al., 2020). Similarly, the far-right refers to an umbrella concept which captures the “radical” and “extreme” collectives of right-wing politics (Pirro, 2023). The term right-wing extremist has been posed as an ideology that contains the following three characteristics: authoritarianism, anti-democracy, and exclusionary holistic nationalism (Carter, 2018). However, there are numerous ways to define and differentiate between these terms and others like them. The Royal United Services Institute, for example, differentiate between the radical right and the far right via what these actors publicly say about violence (RUSI, 2024).

Despite such variability, each of these terms has its place in academic discussion and practical implementation, however, whilst it is necessary to identify that there are differences between extreme right groups and ways to categorise them, it is equally necessary to determine that it is not the role nor the intent of this research to explore these concepts further. Consequently, drawing from pivotal academic works by Cas Mudde and others, this thesis adopts the understanding that the term extreme right can be understood as “a collective term of radicals and extremists united by acceptance of social inequality, authoritarianism, and nativism”. Thus, following the development of a critical understanding of the extreme right, this research will develop an understanding of the key contributors that account for the variability of members of the extreme right.

Having determined and clarified how this thesis interprets the term extreme right, it is necessary to apply context to this term. This is a necessary step not only to understand the extreme right further but also to understand what differentiates the extreme right from extremist collectives and what impact this may have on wider society. Academia and industry have long identified the inherently political nature of extremist groups and collectives. Hostility toward these political actors (including the extreme right) relates to how these collectives operate on or beyond fringes of tolerable political speech, otherwise referred to as the Overton Window (Dan, 2022). Understanding the extreme right as a fringe collective on the outskirts of societally tolerated political views is to dismiss the concept that these collectives may otherwise have the potential to have a meaningful impact on existing political structures. Framed differently, in a broad sense, the academic literature identifies that whilst extremist groups come and go, the practice of extremism persists and is likely to continue doing so (Rapoport, 2004; Suttmoeller et al., 2018). Furthermore, this understanding argues

that collectives on the fringes or beyond the scope of the Overton Window have a role to play within a broadly accepted political structure, regardless of their popularity.

In understanding extremist collectives as part of broader political ecosystems, Cas Mudde's *Normal Pathology Thesis* (2010) offers a valuable perspective. Mudde positions the extreme right as a by-product of political systems, triggered by crises—whether real or perceived. Yet, this theoretical framework remains limited in identifying the role of perception and its exploitation, a gap addressed by recent research on the extreme right's manipulation of contemporary crises, such as the COVID-19 pandemic (Pantucci, 2022; Backhaus, 2023). While Mudde acknowledges the adaptability of extremist collectives, the author does not address the extent to which digital spaces amplify this adaptability. This detail is particularly significant as the extreme right increasingly leverages imagery, memes, and satirical content to propagate ideologies, evade content moderation, and attract new members.

The literature on extremism studies also reflects a historic underrepresentation of the extreme right (Conway, 2017; Schuurman, 2019). This imbalance can be traced back to the academic and political focus on jihadist extremism following the War on Terror and high-profile attacks in the early 2000s (Amnesty International, 2012; Koeler, 2016). While this focus is justified, it has left critical gaps in understanding how extreme right collectives operate in digital spaces and use visual culture as a tool for recruitment, radicalisation, and mainstreaming extremist ideologies (Philips, 2023; Nouri et al., 2020). Recent studies have started addressing this gap, particularly regarding the alt-right's use of internet culture (Hawley, 2017; Nagle, 2017). However, there remains a need for systematic frameworks capable of detecting and analysing extreme right imagery across different platforms. This thesis aims to fill this gap by developing a framework that examines the role of visual culture in promoting extreme right ideologies.

The extreme right cannot be understood in isolation; it must be contextualised within broader contexts, namely, the influence of socio-political systems. While extremist groups are often dismissed as transient or peripheral, Rapoport (2004) and Suttmoeller et al. (2018) argue that extremism is a persistent and adaptable phenomenon. Mudde's *Normal Pathology Thesis* aligns with this perspective but falls short of addressing how modern digital environments have amplified the extreme right's reach and adaptability. The COVID-19 pandemic is a

relevant example where extreme right groups exploited public uncertainty to spread disinformation and conspiracy theories, inserting themselves into mainstream discourse (Backahus, 2023; Pantucci, 2022). Recognising the importance of socio-political circumstances, this thesis will integrate these contextual factors into the development of its framework to ensure a more nuanced and adaptable approach to detecting extreme right imagery.

A critical gap in the literature lies in the study of digital visual culture as a tool for extreme right communication and mobilisation. While memes, images, and internet culture have been increasingly recognised as integral to extreme right strategies (Doerr, 2021; Wimberly, 2020), existing research often treats these visual elements as anecdotal or secondary. There is a lack of systematic analysis exploring how extreme right imagery is created, disseminated, and engaged with across multiple platforms. This is significant because much of the current research focuses on isolated platforms—such as Twitter, Gab, or Reddit—without fully accounting for the extreme right’s cross-platform behaviour (Zannettou et al., 2018). The ability of extreme right actors to adapt their visual content and tone to suit platform-specific norms and policies remains underexplored, creating a critical need for more holistic, platform-agnostic studies. This thesis addresses this gap by focusing on Telegram, a platform that functions as a beacon platform within the extremist digital ecosystem. Beacon platforms, as defined by Fisher et al. (2019), store and signpost extremist content that originates or circulates across multiple online platforms. Telegram’s role as a repository and connector of extreme right content makes it an ideal data source for this research. Unlike studies limited to a single platform, analysing imagery from Telegram provides insights into the broader context of extreme right visual content. Users engaging with Telegram are likely consuming content that is hyperlinked, replicated, or referenced across various other platforms, reflecting the interconnected nature of extreme right digital strategies. By examining extreme right imagery collected from Telegram, this research captures visual content that is not platform-specific but representative of a wider online ecosystem. The conclusions drawn from this data are, therefore, relevant beyond Telegram and speak to the cross-platform challenges faced by content moderation systems. This study aims to fill a critical gap in the literature by developing a framework capable of systematically detecting and analysing extreme right imagery across platforms, offering insights into its creation, adaptation, and

role in mainstreaming extremist ideology. In doing so, it contributes to a deeper understanding of how visual communication is weaponised in digital spaces and how moderation practices can be improved to counter these evolving tactics effectively.

Furthermore, due to the ambiguity of imagery, particularly through the use of irony, satire, and nihilism, the detection and analysis of extreme right visual is further complicated. Scholars such as McKelvey et al. (2015) and Milner (2014) have identified how humour masks extremist messaging, but there remains insufficient understanding of the implications, especially its role in radicalisation. This thesis aims to bridge this gap by examining how humour-based and seemingly innocuous visual content can disguise extremist intent, making it more palatable to mainstream audiences while bypassing moderation systems.

The ideological and operational variability of the extreme right across contexts creates significant challenges for its detection and effective response. Comparative studies of Germany and the United States demonstrate how national political systems, speech protections, and cultural norms shape the structures and strategies of extreme right movements (Michael and Minkenberg, 2007; Pitcavage, 2019). For example, Germany's militant democracy model enforces strict controls on anti-democratic ideologies, including the power to ban extremist groups and symbols. In contrast, the United States' First Amendment protections limit government intervention in speech and association but marginalise extreme right political parties through the dominance of its two-party electoral system. These divergent approaches illustrate how geopolitical differences complicate a comprehensive understanding of extreme right movements. In the U.S., the extreme right is characterised by ideological fragmentation, comprising diverse subgroups such as racialsists, militias, and neo-Confederates, making it politically diffused but resilient (Pitcavage, 2019). Conversely, German extreme right movements are more ideologically cohesive, often united by ethnonationalist and fascist narratives (Michael and Minkenberg, 2007). Despite these structural differences, both contexts highlight the unifying role of digital visual culture. Online imagery, memes, and other visual media enable extreme right collectives to disseminate their ideologies, attract recruits, and evade detection, transcending national and ideological boundaries.

Speech restrictions and platform-specific policies further complicate the detection of extreme right content, revealing notable gaps in the literature. Digital platforms often align their moderation strategies with national legal frameworks, allowing extremist groups to adapt. They frequently exploit ambiguities in satire, irony, and humour to bypass moderation and shift to alternative platforms when restrictions are enforced. This variability in how content is created and disseminated across platforms remains underexplored, underscoring the need for platform-agnostic frameworks to detect and analyse extremist visual content effectively. The role of social media platforms in facilitating extreme right communication is particularly evident in Zannettou et al.'s (2018) large-scale study, which analysed 160 million images from Twitter, Reddit, 4chan's Politically Incorrect board (/pol/), and Gab over 13 months. Their findings reveal significant differences in the nature of content across platforms. For instance, memes on Reddit tend to be more light-hearted and playful, while content on platforms like Gab and 4chan often contains overtly racist or politically charged messages (Nouri et al., 2020). Prisk (2017) further highlights that memes on 4chan frequently exhibit 'hyperreality', a concept introduced by Baudrillard (1994), where content becomes detached from or reconstructs an alternative version of reality. This hyperreal content, along with the heavy use of irony, can make content indecipherable to individuals outside of the ideology. Similarly, McKelvey et al. (2015) describe 4chan content as "ironic, sarcastic, nonsensical, and subversive," highlighting its deliberate ambiguity. This highlights a gap in the literature, as the fragmented understanding of extreme right visual strategies impedes the development of effective detection frameworks. Addressing this gap is essential to improve content moderation systems, particularly in understanding how extremist groups manipulate visual media to exploit digital spaces.

The alt-right exemplifies a modern evolution of the extreme right, offering critical insights into the direction of the extreme right more broadly. Unlike earlier iterations of the extreme right, the alt-right thrives in digital spaces, strategically integrating internet culture (e.g. memes, humour, and trolling) into its communication and radicalisation processes (Hawley, 2017; Greene, 2019). This reliance on visual content enables the alt-right to engage a younger, tech-savvy audience, circumvent traditional content moderation, and mainstream extremist ideologies through irony and satire (Doerr, 2021). These tactics reduce the visibility of overt extremism, creating an adaptable and evasive strategy that reflects broader trends within the

extreme right. The centrality of imagery to the alt-right's strategy signals a wider shift in how the extreme right communicates and disseminates its ideology. Visual symbols, such as Pepe the Frog, have become emblematic of the alt-right's ability to rebrand and repurpose popular culture for extremist purposes (Yiannopoulos and Bokhari, 2016). While these images may appear innocuous or humorous to outsiders, they often embed deeper extremist messages that appeal to the in-group while masking their intent from external observers (Milner, 2014; McKelvey et al., 2015). This deliberate ambiguity (where irony and nihilism blur the boundaries between satire and genuine extremism) makes such content particularly difficult to identify, moderate, and address.

Further complicating matters is the alt-right's ability to disseminate content across multiple platforms. However, research into this cross-platform dissemination remains limited, failing to capture the full scope of how alt-right content circulates and influences broader extreme right networks. The alt-right's use of visual culture highlights a critical gap in the existing literature: while studies have begun to examine memes and internet culture in extremist communication, they often treat these phenomena as secondary or anecdotal (Doerr, 2021; Wimberly, 2020). There remains insufficient analysis of how alt-right imagery evolves, circulates, and radicalises audiences. As the alt-right's methods—especially its use of imagery—become increasingly influential, they serve as a bellwether for broader shifts in extreme right behaviour and tactics. This thesis develops a systematic framework to detect and analyse extreme right imagery, addressing gaps in understanding its cross-platform dissemination, contextual ambiguities, and radicalising potential. It focuses on the extreme right's use of digital visual content, including irony, humour, and memes, which complicate detection and moderation. By defining the extreme right as a diverse but unified movement centred on nativism, authoritarianism, and social inequality, the thesis highlights how these groups weaponised internet culture to mainstream their ideology and evade moderation. This research contributes to both academic knowledge and practical improvements in content moderation systems by offering tools to better identify and contextualise extreme right imagery in the digital age.

## Online Extremism

Building on the definitions of extremism and the extreme right established earlier, this section delves deeper into the critical role of the digital domain in shaping contemporary extremist activities. With the advent of Web 2.0 which is characterised by interactive, user-generated content, the internet has fundamentally transformed the ways extremist groups operate and expand their influence. The extreme right has historically been an early adopter of digital communications technology, leveraging it to promote ideologies, recruit members, and create global networks. This section will explore online extremism, focusing on its definition, historical evolution, and contemporary manifestations, with particular attention to the extreme right's strategic use of imagery and digital tools. These insights directly contextualise the central goal of this thesis: to develop a framework for accurately moderating extreme right imagery on online platforms. Online extremism has evolved into a sophisticated phenomenon, with extremists employing innovative digital strategies, including bots, multi-platform dissemination, and beacon systems, to amplify their reach. This section also examines how the community-building potential of the internet has made it an indispensable tool for extremist movements. By offering a seemingly borderless and anonymous environment, the internet has enabled extremists to recruit, radicalise, and build communities that transcend physical and ideological boundaries. Understanding these dynamics is essential to equipping researchers, policymakers, and digital platforms with the knowledge required to counter the growing influence of online extremism.

The rise of Web 2.0 marked a paradigm shift in how individuals and groups interact online. Unlike the static and one-directional communication of Web 1.0, Web 2.0 introduced interactivity and user-generated content, revolutionising the digital landscape (Hiremath and Kenchakkanavar, 2016; Cormode and Krishnamurthy, 2008). While mainstream society benefited immensely from this shift, extremist groups similarly capitalised on these advancements to propagate their ideologies. Consequently, online extremism emerged as a distinct form of extremism, characterised by the exploitation of digital platforms to achieve extremist goals. Paralleling the earlier discussed issues with definitions, defining online extremism has proven challenging, with no universally accepted definition in academic or policy discourse (Berger, 2018). However, Winter et al. (2020) provide a useful framework, defining online extremism as: "Internet activism that is related to, engaged in, or perpetrated

by groups or individuals that hold views considered to be doctrinally extremist." This definition effectively contextualises extremism within the digital realm, highlighting the interplay between traditional extremist ideologies and the unique affordances of online platforms. While some scholars question the necessity of distinguishing between online and offline extremism (Floridi, 2015; Valentini et al., 2020), this distinction remains critical for this thesis, which focuses on platform-specific behaviours, such as the dissemination and moderation of extreme right imagery.

The integration of extremist activities with the internet began during the Web 1.0 era, as groups sought to leverage its reach for communication and recruitment. Early examples, such as the "White Aryan Resistance" bulletin board predating the World Wide Web (Conway et al., 2019), illustrate how the extreme right recognised the potential of digital platforms. However, it was the establishment of platforms like *Stormfront* in 1995 that exemplified the strategic adoption of online technologies by extremists. Founded by former Ku Klux Klan leader Dan Black, Stormfront became a hub for white supremacist ideologies, demonstrating the long-term viability of an extremist online community (Bowman-Grieve, 2009). Stormfront's success highlights several key strategies that remain relevant today. Initially a static, one-way communication platform, it evolved into an interactive community, incorporating features such as regional sections and discussion boards. The site also facilitated indirect signalling of extremist beliefs through usernames, avatars, and imagery which is a practice still prevalent in extreme right circles (Caren et al., 2012). By fostering a virtual "community of practice" (Hundeide, 2003), Stormfront created conditions for ideological reinforcement, recruitment, and, in some cases, violence. Notably, users of the platform have been linked to high-profile violent acts, including Anders Behring Breivik's 2011 attacks (Pantucci, 2011). Stormfront serves as an early blueprint for how the extreme right leverages online platforms, laying the groundwork for the contemporary strategies observed in the Web 2.0 era. Understanding this historical trajectory is crucial for developing moderation frameworks that address both the overt and covert tactics employed by modern extremist groups.

Having given a brief overview of the history of online extremism, it is necessary to now contextualise the evolution of modern online extremism, with a specific focus on the extreme

right and its adaptation to digital spaces. In the current climate, it is commonplace for an extremist group or movement to have an online presence (Tech Against Terrorism, 2023). This online presence can vary significantly not only across different extremist ideologies (extreme right, Islamist extremism, left-wing extremism) but also within them. There are numerous online platform types, and they each have their affordances that would appeal to extremist groups (MacDonald et al., 2022). However, despite the variety of roles and purposes of online platforms, the use of social media is commonly found at the forefront of online extremism discussion (Alava et al., 2017). Having established an example of how extreme right collectives used the internet in Web 1.0, it is necessary to consider the risks associated with online extremism and its role in impacting the lifespan of an extremist group.

A broad spectrum of extremist groups such as the alt-right, the far left, and animal rights, to non-extremist groups such as the Hippies have benefitted from the internet. Through attracting attention, birthing new loyalties, and ultimately expanding their existing network (Garrett, 2006; Norris, 2004). However, the opposite is equally true, in that through existing in online public spaces, individuals with contrasting opinions are equally eligible to engage. A broad and diverse spectrum of individuals who hold a diverse set of views, opinions, and beliefs use the internet. The accessibility to extremist groups and content through social media can lead users toward engaging with and learning more about extremist groups and ideologies. This in turn can lead to consequences that reach beyond the internet. This plays to the narrative of radicalisation, whereby, extremist recruiters aim to convert potential recruits from an online environment to an offline environment (Gaudette et al., 2022; Galloway and Scrivens, 2018). The presence of online extremism has the potential to indoctrinate an individual into a new ideology or to push those for and against an ideology further to the fringes of their views. Although most users are unlikely to come into contact with or even identify extremist content in a contemporary setting, there are risks associated with one instance of radicalisation that can result in acts of violence in an offline setting.

Online congregations of a wide spectrum of extremists may not be in itself an act of violence. However, when platforms such as Stormfront continue to exist, there is always a present risk of a progression of one of its members leading to violence. Particularly in the case of the extreme right, due to the numerous sub-factions which fall within the label's remit, there is a

significant spectrum of individuals, ideologies, and intentions. These interactions, for example, contributed to the first “alt-right killer”, Elliot Rodger, killing six students in Isla Vista in 2014 (Witt, 2020). This reinforces the notion that seemingly harmless online images and humour-sharing culture can and have been linked to acts of violence.

The extreme right has amassed at least 6,000 extreme right groups in the United States since 1990 (Chermak et al., 2013). Whilst the internet has numerous affordances from the perspective of extreme right groups, it is not yet understood how this impacts their ability to exist over time. Extreme right groups akin to other extremist and terrorist organisations cannot endlessly persist. Extremist groups are born and mature until they cease to exist (Long, 1990; Rapoport, 2014; Vittori, 2009). Research in this area unveils numerous findings on the factors which determine a group’s ability to exist over time. Factors such as a group’s participation in violence, for example, have been shown some correlation with a group’s longevity (Cronin, 2009). However, this factor alone does not determine the longevity of the group. Furthermore, membership count has also been identified as a telling factor. For example, a large non-violent group is likely to cease to exist sooner than an equally large violent group. On this matter Cronin (2009) suggests that internal group disagreements over the use of violence often lead to group instability. There is an additional line to be drawn within the frequency of attacks, Blomberg et al., (2010) and Chermak et al., (2013) identify that some groups may conduct a single violent attack, while others may commit multiple. Thus, the complex nature of what makes a group end is exacerbated in this case due to the in-group variability of the extreme right. And therefore, questioning to what extent violence informs how long the movement will persist. Furthermore, other differences exist between violent and nonviolent groups beyond ideologically motivated violence (Chermak et al., 2013). In recent years, the adoption of technology, and the internet more specifically, has had a positive correlation with group longevity. However, the more complex the technology is relative to a population, the less likely it is that all its members will migrate. In a UK context, this can be seen in the case of Britain First. Whereby the UK-based extreme right group Britain First in March 2018 had their Facebook group disrupted and removed from the platform. This led the group to start a new following on the social media platform Gab. Despite having a digitally active following, Britain First’s removal from Facebook also caused the group to amass a much smaller and less diverse recruitment following (Nouri et al., 2019). Therefore,

whilst the extreme right readily uses the internet, they are still susceptible to a drop in membership when moving to increasingly complex technologies.

Despite the many benefits of the internet to support an extremist group's ability to persist, the uptake of complex technology has also been linked to their demise (Hodge and Hallgrimsdottir, 2021). However, the extreme right offers a unique perspective due to their history of adopting the Internet. Furthermore, when considered a subset of the extreme right like the alt-right, an increasingly unique perspective can be offered. Being a byproduct of internet culture appears to be a strength instead of a limitation due to its tactics being synonymous with internet culture. The absence of a consistent and comprehensive ideology creates a broad spectrum of nuanced propaganda. With a large array of political opinions coming across in this material, the right has seemingly stumbled upon an effective recruitment model. Although seemingly humorous, playful, and controversial, this activity can already be associated with inspiring several individuals to commit to a path of violence under the extreme right's ideological umbrella.

The unique emergence and evolution of the alt-right offer critical insights into contemporary and future cases of online extremism. This overview of the alt-right's online extremism builds on the understanding found in the previous section to show how the collective leverages internet culture and digital tools to propagate its ideology. While the alt-right may appear to represent a break from earlier extremist movements, its use of memes, trolling, and online forums can be viewed as a continuation of older extremist strategies adapted to the affordances of the digital age (Greene, 2019; Topinka, 2018; Finlayson, 2021). The alt-right's adoption of internet culture as a core tactic, particularly through imagery and humour, represents a shift in how extremist communication occurs, reflecting both innovation and continuity within the extreme right movement. The rapid growth of the alt-right is intrinsically linked to its decentralised, online nature and its ability to adapt to evolving technological platforms. Initially emerging in 2008, the movement gained traction around anonymous forums like 4chan, 8chan, Reddit, and extreme right websites such as VDARE and Breitbart (Hawley, 2018; Taylor, 2020). While its foundational figures, including Paul Gottfried and Richard Spencer, played pivotal roles in its conceptualisation, the alt-right's resurgence in 2015—particularly during the U.S. presidential election—was propelled by its appropriation

of digital tools to connect disparate factions under a unified banner (Ohlheiser and Dewey, 2016; Powers, 2022). This leaderless and decentralised nature has enabled the alt-right to thrive, blurring the lines between traditional social media and niche extremist platforms (Lorenzo-Dus and Nouri, 2021).

A key factor in the alt-right's online success lies in its strategic use of mainstream and niche platforms. Larger social media sites initially provided visibility and audience exposure, facilitating recruitment and ideological dissemination (Nouri et al., 2020; Gaudette et al., 2021). However, as moderation policies improved, the alt-right adapted by migrating to less-regulated platforms such as Gab, 4chan, and 8chan—spaces where extremist content is less punished under the guise of free speech (Colley and Moore, 2019; Zannettou et al., 2018). These platforms allow the alt-right to share hate speech, conspiracy theories, and radical content while evading traditional moderation systems (Price, 2017; Benson, 2016). The digital dispersion of extremists following moderation on mainstream platforms underscores a key limitation of current moderation approaches: while content can be removed, extremist behaviour persists, often in more concentrated and radicalised online spaces (Kelly, 2019).

The alt-right's reliance on visual content, particularly memes, is a hallmark of its strategy. Humour, irony, and trolling are employed to mask extremist messaging, making it appear harmless or satirical to outsiders while fostering in-group cohesion and radicalisation (Milner, 2014; McKelvey et al., 2015). Memes and imagery are particularly effective because of their ability to spread rapidly across platforms, embedding extremist narratives into popular internet culture. Yet, there remains a significant gap in understanding how these images evolve, circulate, and impact audiences across platforms. The limited focus on visual propaganda, particularly within niche platforms, highlights the need for systematic research on extreme right imagery.

This section has underscored the critical role of digital platforms in enabling extremist group operations while highlighting significant gaps in the literature regarding the strategies employed by the extreme right to recruit, radicalise, and sustain their movements. Although mainstream social media platforms have received substantial scholarly attention, smaller, less-regulated platforms like Telegram remain notably underexplored, despite their importance for extremist content dissemination. A key area of neglect is the role of visual

propaganda, such as memes, satirical imagery, and cultural symbols, in fostering radicalisation, maintaining group longevity, and building cohesive in-group identities within decentralised online ecosystems. This thesis directly addresses these gaps through its central research question: *To what extent is imagery used as a medium of communication by extreme right groups?* (Research Question 1). By focusing on Telegram as a “*beacon platform*” (Fisher et al., 2019), this research systematically examines the use of extreme right imagery in recruitment, ideological dissemination, and community-building. The study develops a framework for detecting, analysing, and contextualising such visual content, providing actionable insights to improve content moderation across platforms. Crucially, this approach captures the layered and evolving nature of visual communication, including its cultural nuances, symbolic subtexts, and role in radicalisation pathways. By bridging these critical gaps, the thesis advances understanding of how extreme right groups strategically exploit visual culture and digital tools to amplify their influence. It also contributes to broader efforts to disrupt the amplification and normalisation of extremist ideologies online, offering a robust foundation for enhancing moderation strategies in increasingly decentralised digital spaces.

#### Case Study: The Proud Boys

The Proud Boys, a prominent extreme right group, exemplify the integration of online and offline strategies to gain influence, legitimacy, and a sense of community. As a case study, their methods, evolution, and impact provide critical insights into the contemporary extreme right. The group’s use of online imagery, memes, and trolling not only facilitates recruitment but also fosters a cohesive in-group identity. This analysis contextualises the group within the framework of the alt-right and illustrates their prominence as the most represented groups in the data examined in this thesis.

The Proud Boys emerged in 2016 amidst the political landscape shaped by Donald Trump’s presidency. Trump’s rhetoric and reluctance to condemn extremist groups (Kenes, 2021) emboldened organisations like the Proud Boys, elevating their visibility. Trump’s infamous “Proud Boys, stand back and stand by” comment during the 2020 presidential debate (McBain, 2020) cemented the group’s association with his administration and thrust them into the global spotlight. Canada’s designation of the group as a terrorist entity (Public Safety

Canada, 2021) further underscores their notoriety and influence. The group's identity as a male-only, anti-immigrant, Islamophobic organisation with ties to misogynistic online spaces such as the "manosphere" (Nagle, 2017) highlights its alignment with broader alt-right ideologies. Their blend of anti-political correctness, disdain for "white guilt," and appeals to "Western chauvinism" (McBain, 2020; SPLC, 2021) situates them within the extreme right spectrum while leveraging shared grievances to recruit a diverse membership. This flexibility allows the group to mask its more violent and exclusionary undertones with a veneer of playful satire and fraternity, widening its appeal.

The Proud Boys' success in building influence stems significantly from their adept use of digital platforms. Social media enabled the group to disseminate ideology, recruit members, and build an in-group identity. Memes played a pivotal role in their strategy. Visual content such as the co-opted cartoon character Pepe the Frog, dressed in Proud Boys attire and making white supremacist gestures, served as both an ideological signal and a recruiting tool (Anti-Defamation League, 2016; Kenes, 2020). The group's mantra, "West is the Best," often paired with such imagery, underscores their underlying chauvinistic ethos. The use of memes and satire not only attracted members but also provided plausible deniability for their extreme rhetoric, complicating content moderation efforts on mainstream platforms. By embedding ideological messages within seemingly harmless humour, the group leveraged the internet's participatory culture to amplify its reach (Shifman, 2014; Mina, 2018).

The Proud Boys' offline activities demonstrate their dual role as both an online alt-lite group and an offline alt-right organisation. Early on, the group projected an image of a non-threatening fraternity, but their actions at events like the 2017 Charlottesville "Unite the Right" rally and the 2021 Capitol riots revealed their propensity for violence (Anti-Defamation League, 2018; Hall, 2021). These events showcased the group's capacity to mobilise offline, aligning their online presence with physical acts of aggression. The group's membership rituals, including violent initiation practices such as the "fourth degree" requiring members to commit acts of violence on behalf of the organisation (Kutner, 2020), highlight the overlap between their public image and private conduct. The Proud Boys' offline behaviour further solidifies their role as a gateway to more extreme ideologies and actions, despite their efforts

to maintain plausible deniability through disavowals by founder Gavin McInnes (Hope Not Hate, 2021).

De-platforming significantly disrupted the Proud Boys' mainstream online presence, forcing them onto smaller platforms like Gab, Telegram, and Parler (Crawford, 2020). While this migration reduced their audience size, it concentrated their content among more ideologically aligned users, fostering a more radicalised community. This shift underscores the challenges of content moderation, as the removal of extremist content from mainstream platforms often pushes users to less regulated spaces, where harmful ideologies may intensify. Despite the challenges posed by de-platforming, the Proud Boys' persistence on niche platforms demonstrates the resilience of alt-right groups in adapting to changing digital landscapes. Their continued use of online imagery and memes on alternative platforms highlights the enduring role of visual content in building in-group cohesion and maintaining ideological dissemination.

Founder Gavin McInnes played a critical role in shaping the group's identity and appeal. McInnes' provocative humour, misogynistic commentary, and alt-right affiliations provided a gateway for potential recruits to engage with the group (Miller, 2018). His role as both a public figure and a symbolic leader underscores the interplay between charismatic leadership and online community-building in the extreme right. The group's fraternity-like structure and shared grievances create a strong sense of community among members. The Proud Boys' use of shared symbols, rituals, and online content fosters solidarity and a sense of belonging, key factors in their ability to attract and retain members. This dynamic mirrors broader trends in the alt-right, where humour, trolling, and satire mask deeper ideological commitments.

The case study of the Proud Boys provides a compelling example of the modern manifestations of the extreme right and their utilisation of online and offline domains to build influence and reinforce in-group cohesion. The Proud Boys demonstrate how humour, satire, and imagery serve as powerful tools for recruitment, narrative dissemination, and ideological reinforcement. These tools create an accessible and appealing veneer that masks the group's more violent and extremist undertones, making it easier to draw in potential recruits who may later radicalise further. The Proud Boys leveraged their affiliation with high-profile political figures, such as former US President Donald Trump, and significant events like the

Capitol riots and Charlottesville rally to cement their presence within the alt-right ecosystem. Their adoption of internet culture, particularly the use of memes and satirical content, enhances their ability to disseminate ideology subtly while maintaining plausible deniability. This strategy of embedding ideological messages within seemingly playful or innocuous imagery complicates content moderation efforts, as it requires contextual and cultural understanding to identify the subtext.

This case study highlights the complex challenges associated with de-platforming extreme right groups like the Proud Boys. While removing them from mainstream platforms disrupts their visibility and reach, it often drives members toward smaller, less-regulated spaces where harmful ideologies can thrive and become more concentrated. This dispersion underscores the urgent need for cross-platform collaboration and the development of sophisticated frameworks capable of detecting and moderating coded extremist content across diverse digital ecosystems. The Proud Boys' evolution reveals two critical insights into the role of visual imagery within extreme right communication strategies. First, internet culture serves as a cornerstone for fostering in-group identity, with symbols like Pepe the Frog and phrases such as "West is the Best" acting as unifying markers that reinforce shared belonging and purpose. Second, the strategic use of imagery functions as a gateway to radicalisation, using layered communication tactics—grounded in humour, satire, and cultural grievances—to obscure extremist ideologies. This deliberate ambiguity makes extremist messaging appear innocuous to outsiders, complicating detection and moderation efforts, particularly on mainstream platforms where such content often blurs the line between satire and harm.

However, significant gaps remain in understanding the visual strategies employed by extreme right groups, especially on smaller, underexamined platforms like Telegram. While mainstream platforms have been the focus of considerable research, there is a critical need to explore alternative digital spaces where extremist visual propaganda continues to evolve (Crawford, 2020; Fisher et al., 2019). Additionally, the role of imagery in sustaining group longevity, reinforcing ideological cohesion, and enabling decentralised communication across fragmented ecosystems remains underexplored. This thesis addresses these gaps through another of its central research questions: *Do extreme right groups have differing visual styles that are communicated and reinforced through the use of imagery?* (Research Question 4). By

focusing on Telegram as a “*beacon platform*” (Fisher et al., 2019), this study develops a systematic framework for detecting, analysing, and contextualising extreme right visual content. It investigates how visual styles are adapted and reinforced, contributing to group sustainability, recruitment, and ideological dissemination. Ultimately, this research advances the understanding of how extreme right groups exploit visual imagery as a strategic tool, bridging critical gaps in the literature and providing actionable insights for improving content moderation systems. By equipping platforms with the tools to identify and interpret extremist imagery, this study strengthens efforts to disrupt pathways of online radicalisation and counter the spread of extremist ideologies in increasingly decentralised digital environments.

### Challenges of Extreme Right Content Moderation

Despite significant efforts by online platforms, governments, and supranational organisations, extremist content continues to persist across the digital domain. While proactive measures have mitigated some risks, the complete removal of online extremism remains an unattainable goal (Weimann, 2004). Social media platforms particularly offer extremists unparalleled opportunities to disseminate their ideologies to large audiences with speed, minimal financial cost, and relative ease (Weimann, 2006; Conway and Courtney, 2018). As a result, content moderation has shifted from being a desirable feature to an essential tool in digital governance, particularly as user-generated content (UGC) increasingly dominates these platforms. Defined as “an organised practice of assessing UGC [user-generated content] posted to an online platform to determine whether or not that content will be subsequently removed or downranked” (Roberts, 2017), content moderation represents a critical form of damage control. It seeks to address illegal and harmful content, including extremism, through ongoing and adaptive strategies aimed at countering the evolving threat landscape.

Understanding the intricacies of content moderation is essential for this thesis, which seeks to develop a framework for detecting and contextualising extreme right imagery. As highlighted in this literature review, extreme right groups leverage visual propaganda that is often layered, ambiguous, and heavily reliant on cultural and contextual cues, making it particularly resistant to traditional moderation methods. Existing moderation systems, which

rely heavily on automated algorithms and human moderators, face substantial limitations in identifying and addressing such content effectively (Gillespie, 2018). Additionally, the migration of extremist groups to smaller, less-regulated platforms like Telegram exacerbates these challenges, necessitating cross-platform solutions that can adapt to fragmented and decentralised digital ecosystems. Complicating these efforts further is the shifting Overton Window, which constantly redefines the boundaries of harmful content (Conway and MacDonald, 2023). Platforms that succeed in moderating one form of extremist material are often immediately faced with emergent threats, such as the shift from Islamist extremism to the rise of extreme right content and the proliferation of disinformation (Piazza, 2022). While some platforms have achieved notable success in content moderation, public sentiment frequently perceives these efforts as insufficient (Watkin, 2023). This underscores the need for a nuanced and transparent approach to moderation that balances the protection of users with safeguarding freedom of expression.

In response to these complexities, this section of the literature review aims to achieve two objectives. First, it provides an overview of existing content moderation practices, establishing the foundational strategies employed by platforms to address extremist content. Second, it outlines the key challenges inherent in these approaches. It is argued that the challenges associated with the diverse and fragmented nature of content moderation policy have placed disproportionate responsibility on individual platforms to create their in-house systems without consistent external guidance or oversight. This, in turn, has created problems in consistency of benchmarks and enforcement, transparency and difficulty in balancing human rights with this disjointed approach. The problems identified in this literature review, such as the inherent difficulties in defining extreme right imagery, are intensified by this fractured governance landscape. The lack of a unified approach to content moderation not only hampers the ability of platforms to respond effectively but also creates inconsistencies in how extremist content is addressed across different digital spaces. This section argues that the development in this thesis of a more transparent, systematic, and collaborative framework is a step towards encouraging this approach across content moderation policy and practice. By addressing these gaps, this research aims to support platforms in their efforts to detect and moderate extreme right imagery by proposing a universally adaptable framework. Such a framework seeks to enhance content moderation efforts across platforms, fostering a

joined-up approach to combatting online extremism in an increasingly fragmented digital environment.

### Modern Content Moderation Systems

This section explores the hybrid human-automated approach to content moderation, with a particular focus on its application in detecting and removing extremist content. Hybrid systems combine machine learning (ML) techniques with human oversight to address the complexities of moderating harmful material online. While these systems represent a critical evolution in content moderation, they also expose significant gaps in implementation, particularly regarding scalability, context sensitivity, and operational limitations. This PhD research aims to develop a data-driven content removal framework and policy recommendations for online platforms, making it essential to critically examine the strengths and weaknesses of hybrid systems. Such an analysis is crucial for designing more effective, scalable, and ethically robust strategies to address the challenges posed by extreme right content online.

Most online platforms rely on hybrid content moderation processes to manage extremist content. This approach, frequently cited in academic literature and transparency reports, consists of two interdependent components: automated systems and human oversight. The automated component usually leverages ML techniques to classify and label content. In essence, ML systems function as sophisticated labelling tools (Kozyrkov, 2018), designed to determine whether content meets predefined criteria for extremism. These tools are employed at various stages of content moderation, either pre-emptively (before content is uploaded) or reactively (after it has been flagged). For instance, ML systems might be trained to identify extremist symbols, slogans, or patterns within an image. The second component of hybrid moderation is the human element, which plays a crucial role in managing the nuanced and context-dependent aspects of extremist content. Human moderators are essential for addressing complex cases where automated systems lack certainty, such as ambiguous content or appeals against removed material. Due to the scalability limitations of human resources, their involvement is often reserved for scenarios requiring judgment,

cultural understanding, or deeper contextual analysis that ML systems cannot yet replicate. To provide a comprehensive understanding of hybrid content moderation, this section will examine the strengths and limitations of both components. By dissecting these elements, this analysis will inform the development of a more effective and ethically grounded framework capable of addressing the unique challenges posed by extreme right content online.

### *The Human Component*

The human element of hybrid human-automated content moderation is critical to addressing the complexities and nuances of extremist content, particularly where context, language, and cultural subtleties come into play. Automated systems, while capable of identifying straightforward cases, struggle with ambiguities such as coded language, satire, or fair use scenarios, making human oversight indispensable (Feamster, 2018). This is especially true for the moderation of extreme right content, which often relies on coded and anonymised communication (IRTL, 2022). While no credible arguments exist for entirely automated processes, the division of responsibilities between human moderators and machine algorithms remains a contentious and evolving challenge (Link et al., 2016). Due to the harm to human content moderators, the aim of the content moderation practice should be to prioritise the safety of users whilst proactively minimising the harm to moderators until a new solution prevails.

Human moderators play an essential role in reviewing complex or uncertain cases flagged by automated systems, often serving as the final arbiter in the decision-making process. Large social media platforms like Facebook employ tens of thousands of moderators, with the most recent figure standing at 15,000 (Thomas, 2020). However, many of these moderators are employed through third-party subcontractors, highlighting a cost-saving approach that has been criticised for its lack of transparency and ethical considerations (Financial Times, 2023a). Moderators are often required to review a wide spectrum of flagged content, including graphic violence, child exploitation, and terrorism, under tight time constraints, typically analysing up to 400 pieces of content per day with an average of 30 seconds allocated per item (Newton, 2019). The task of human moderators is fundamentally to determine whether flagged content violates a platform's terms of service. Depending on the platform, moderators may have multiple options when reviewing content, such as ignoring it, escalating

it for further review, or deleting it outright (Bickert and Fishman, 2017). These decisions are guided by varying levels of training, often consisting of short courses and manuals that may inadequately prepare moderators for the diverse and complex scenarios they encounter (Hopkins, 2017). The limited training and support provided to moderators underscore broader concerns about the lack of investment in human content moderation infrastructure.

Human moderators face significant challenges in identifying and contextualising graphic content, particularly when it is used by extreme-right groups to convey ideologies or incite hostility. While some forms of graphic content, such as depictions of violence or explicit hate symbols, are more easily recognised, subtler forms can evade detection. For example, imagery containing historical references, symbolic violence, or grotesque caricatures can communicate extremist narratives without overtly violating platform policies. These visuals often provoke emotional reactions, reinforcing in-group cohesion and vilifying out-groups. Research Question 2, *'To what extent is graphic imagery used in the communicative strategies within extreme-right imagery'* is essential for understanding how such content operates within the broader framework of visual propaganda. Graphic imagery often plays a dual role: it shocks or intimidates adversaries while reinforcing the ideological bonds of the in-group. Importantly, this thesis focuses on graphic content because of its proven effectiveness in other extremist contexts. For instance, ISIS's online propaganda became infamous for its graphic nature, which not only served to terrorise but also attracted attention and amplified its message. This recognisability facilitated the development of moderation tools specifically designed to detect ISIS content. Following this logic, if extreme-right content were similarly graphic, elements of these existing tools could potentially be repurposed to moderate such content more effectively. Therefore, this thesis will investigate the prevalence of graphic imagery within extreme right spaces. By mapping the usage patterns of graphic imagery across different extreme right groups on Telegram, this study aims to provide actionable insights that enhance moderators' ability to anticipate and contextualise such content.

The role of human moderators becomes particularly challenging when dealing with subtle or coded extremist content. For example, extreme right groups often employ symbols, euphemisms, and numeric codes such as "18" (representing Adolf Hitler's initials) or "88" ("Heil Hitler") to communicate their ideologies covertly (IRTL, 2022). These codes require

specialised knowledge to identify, yet only 150 of Facebook's 15,000 moderators are reportedly dedicated to counterterrorism, representing just 1% of the moderation workforce (Bickert and Fishman, 2018). This disparity highlights the limitations of relying on generalist moderators to address highly specialised and context-dependent extremist content. Furthermore, the absence of transparency regarding ongoing training and expertise development raises concerns about the effectiveness of human moderators in adapting to the evolving tactics of extremist groups.

Human moderators are exposed to significant psychological risks due to the nature of their work, which often involves prolonged exposure to graphic and harmful content. Studies have documented cases of vicarious trauma, post-traumatic stress disorder, and desensitisation among moderators (Newton, 2019; Steiger et al., 2021). These mental health challenges not only affect the well-being of moderators but also have implications for the accuracy and consistency of their decision-making processes. Moderators experiencing desensitisation may develop a skewed perception of what constitutes harmful content, potentially allowing borderline material to evade removal (Dwoskin, 2019, and Spence et al., 2023). The reported normalisation of extreme content by moderators further underscores the ethical and operational challenges associated with this role. Despite these risks, human moderators are often poorly compensated and feel undervalued, with many expressing frustrations at being treated as a lower-tier workforce (Newton, 2019). The lack of adequate mental health support and financial compensation exacerbates the ethical dilemmas surrounding the use of human moderators, raising questions about the sustainability and fairness of current practices.

The scalability of human moderation is constrained by financial and logistical factors, particularly for smaller platforms with limited revenue streams. Unlike automated tools, which can be shared across platforms through initiatives such as hash databases, human moderators are a resource-intensive solution that smaller platforms cannot afford (Tech Against Terrorism, 2023a). This disparity leaves smaller platforms more vulnerable to extremist exploitation, as they lack the resources to implement effective moderation systems. Even for larger platforms, the cost of maintaining a significant human moderation workforce can be prohibitive, leading to inconsistent investment in this critical area (Financial Times, 2023).

A notable gap in current practices is the lack of publicly available data on the accuracy of human moderation. While platforms frequently report accuracy statistics for automated systems—typically citing an error rate of around 10% (Koetsier, 2020)—similar metrics are absent for human moderators. The combination of ethical challenges, psychological strain, and limited training raises questions about the reliability of human decision-making in content moderation. Greater transparency is needed to assess the effectiveness of human moderators and to develop strategies for improving their accuracy and well-being.

### *The Automated Component*

The automated component of hybrid content moderation relies on ML models to identify, assess, and label content as extremist or non-extremist. These systems are underpinned by decades of technological development, with early examples of ML dating back to 1952 (Shenkman et al., 2021). Modern ML models analyse datasets to extract features that help classify content, whether by detecting specific words, symbols, or patterns. For example, an automated content moderation system for a platform like Twitter might be trained on a dataset containing both extremist and non-extremist tweets. The model would then identify linguistic or visual traits that correlate with extremism, such as the use of specific terms or images, to predict and flag future instances for review. However, the effectiveness of such systems depends heavily on the quality and diversity of their training data. Transparency regarding the implementation of automated moderation tools varies significantly among platforms. While Meta provides some insights into the methods it employs through transparency blogs, other platforms, such as Telegram, offer minimal disclosure about their moderation practices (Facebook, 2024). This opacity complicates efforts to understand how these tools are used and their effectiveness.

ML-driven moderation can be broadly divided into two approaches, matching algorithms and predictive models. Matching algorithms adopt a reactive stance, comparing new content against a database of known extremist material. For instance, an image-matching tool might flag an image already identified in its training data as extremist. This approach has the advantage of simplicity, making it easier to explain and implement, and lends itself to resource-sharing initiatives across platforms. For example, perceptual matching identifies images with minor variations, while cryptographic matching requires an exact match

(Shenkman et al., 2021). However, matching algorithms are inherently limited by their reliance on pre-existing datasets, leaving them unable to detect novel extremist content. Therefore, while they can efficiently flag harmful content, they struggle to differentiate between extremist material and legitimate expressions, such as journalism or human rights documentation. For instance, during the Black Lives Matter protests, graphic content intended to raise awareness was erroneously flagged as extremist by algorithms (Alexander, 2020). Similarly, in Myanmar, attempts to curb hate speech led to the removal of critical human rights documentation (Morada, 2023). Although these tools can reduce false positives by fine-tuning their thresholds, a degree of trade-off persists. Cryptographic matching excels in precision but lacks flexibility, whereas perceptual matching accommodates minor variations, which can help capture content that evolves incrementally (Engstrom and Feamster, 2017). Meta's advancements in identifying slight image variations reflect the necessity of striking this balance (Facebook, 2024).

Predictive models, in contrast, adopt a proactive stance, using training datasets to identify patterns and characteristics that may indicate extremism in previously unseen content. These models employ various techniques, such as classifiers, object detection, semantic and instance segmentation, and scene understanding, to analyse and predict the nature of content (Batra, 2019). For example, classifiers might predict whether an image contains extremist symbols, while object detection can identify multiple elements within a single image. While predictive models hold significant promise, they are computationally demanding and require vast, high-quality datasets. Even with robust training data, these models are not immune to error, as they may fail to generalise insights to more complex or nuanced cases (Hendrycks and Dietterich, 2019). The challenge is particularly present in the case of extreme right content, where symbols, codes, and implicit messaging complicate detection efforts. For example, predictive models might struggle to identify the context or intent behind ambiguous images, such as a coded extremist message embedded in an otherwise innocuous graphic.

Automated content moderation systems face several limitations, including the risk of false positives and false negatives. While false positives involve mistakenly identifying non-extremist content as harmful, false negatives occur when extremist content escapes

detection. Predictive models are particularly prone to these errors, as demonstrated by notable cases in other domains, such as a US military ML model that mistakenly identified friendly tanks as enemy vehicles (Yudkowsky, 2008). These issues underscore the limitations of current technologies, particularly those relying on perceptual hashing rather than cryptographic methods, which may inadvertently capture legitimate content. The lack of explainability, especially in predictive models, further compounds these challenges. Unlike matching algorithms, whose decisions are more transparent and interpretable, predictive models operate as "black boxes," making it difficult to diagnose why errors occur (Duarte et al., 2017). This opacity necessitates careful management of user and stakeholder expectations regarding the capabilities and limitations of automated systems. Another significant limitation is the reliance on high-quality training data. Biases in datasets can skew the performance of ML models, potentially leading to discriminatory outcomes or reinforcing existing biases in content moderation (Shenkman et al., 2021). Furthermore, the evolving nature of extremist tactics, particularly among the extreme right, requires constant updates to datasets and models to maintain their relevance and effectiveness.

Automated content moderation systems are effective at identifying specific symbols or patterns from pre-existing datasets but often falter when tasked with detecting abstract or stylistic elements in imagery. Many groups, including online extremist communities, use cohesive visual styles characterised by specific colour palettes, recurring motifs, and consistent aesthetic choices. These stylistic features subtly communicate group identity and values without overt reliance on explicit symbols or text. For instance, elements such as shock value, humour, or calls to activism often reinforce a group's ideological messaging and brand identity. If extreme-right groups systematically employ such concepts through distinct stylistic choices, there is significant potential for improving automated moderation systems, which frequently overlook these subtler forms of visual communication. Research Question 3, *'Is there an extreme-right visual style that is communicated and reinforced through the use of imagery'* seeks to address this gap by exploring whether extreme-right groups intentionally leverage specific visual styles to bolster their brand and messaging. By identifying and categorising these patterns, this research aims to refine predictive models, allowing platforms to more effectively detect nuanced extremist content. Additionally, it will test and provide a

foundational framework for further research into these less overt methods of ideological dissemination.

While automated content moderation has been praised as a scalable solution to the challenges posed by online extremism, it remains an imperfect tool. Matching algorithms provide an essential baseline for identifying known extremist content but are inherently reactive. Predictive models offer a more proactive approach but require significant resources and expertise to implement effectively. Both approaches demand transparency and accountability to ensure their ethical and effective use. As platforms continue to refine their automated moderation systems, the need for hybrid approaches that combine automation with human oversight becomes increasingly evident. By acknowledging the limitations of automated systems and integrating them with the strengths of human moderation, platforms can create more resilient and adaptive frameworks for addressing online extremism. Ultimately, automated tools must be seen as complementary to human expertise rather than as standalone solutions to effectively counter the evolving landscape of extreme right content online.

### *A Hybrid Approach*

The hybrid approach to content moderation represents an essential yet imperfect response to the challenges posed by online extremism. Automated systems provide scalability and efficiency, while human oversight offers nuanced contextual understanding and judgment. However, the limitations of both components are evident: automated tools struggle to interpret complex cultural and contextual cues, adapt to novel extremist tactics, and ensure transparency, while human moderation is constrained by scalability, resource demands, and ethical concerns. Despite these challenges, the hybrid approach remains the most viable framework for moderating extreme right content, where ambiguity, coded language, and cultural specificity present unique obstacles. The framework developed in this thesis will adopt a hybrid approach, recognising its ability to combine the strengths of human and automated systems. Automated tools will be leveraged for their capacity to handle large-scale data and flag potential extremist content rapidly, while human moderators will play a pivotal role in addressing the nuanced and context-dependent aspects of extreme right imagery. This dual-system approach ensures a balance between scalability and precision, making it

particularly suited to the complexities of moderating content that relies heavily on implicit symbols, layered messaging, and cultural subtleties.

To maximise the effectiveness of this hybrid framework, this research will address critical weaknesses identified in current systems. These include improving the cultural and contextual sensitivity of machine learning models, enhancing the training and support for human moderators, and integrating transparency measures to build trust and accountability. Furthermore, the framework will explore opportunities for cross-platform collaboration and shared resources to ensure scalability and equitable application across platforms of varying sizes and capacities. Given the dynamic and decentralised nature of online extremism, a hybrid approach is the most appropriate choice for moderating extreme right imagery. It offers the flexibility to adapt to emerging threats while maintaining the precision required to address the specific challenges posed by this form of content. By refining and advancing the hybrid model, the framework proposed in this thesis aims to contribute to the development of scalable, ethical, and effective strategies for combating extreme right content online.

#### The Core Problem: A Disjointed Policy Landscape

While significant advancements have been made in understanding and improving content moderation systems, including the hybrid human-automated approach, substantial challenges persist. Chief among these is the inability of current systems to accurately and consistently detect extreme right content with all its cultural and contextual nuances. Even with growing insight into the strengths and weaknesses of hybrid models, a pressing question remains: why has a universally applicable system not been developed and adopted across platforms to foster shared learning and uniformity in approach? This persistent gap can be attributed to the broader framework where content moderation operates i.e., digital governance.

Digital governance is a regulatory structure designed to manage technology use and mitigate harm, evolved alongside the internet's rise, driven in part by the rapid adoption of online platforms by bad actors, including extremists (Zittrain, 2019). These actors exploited digital spaces for propaganda, recruitment, and other harmful purposes, creating an urgent need for countermeasures to protect end-users from exposure to extremist material (Lewis, 2005).

However, the emergence of these threats highlighted a fundamental weakness: the absence of comprehensive, unified frameworks for content moderation. This regulatory vacuum left online platforms to independently develop their own moderation policies, often under intense governmental and societal pressure to act swiftly. Governments tasked with creating legislative responses, face the challenge of long and complex policy development processes that are ill-suited to the immediacy and dynamism of emerging threats. In the absence of a third-party regulatory body to enforce consistent standards, platforms have been forced to 'write the playbook' on extremist content moderation, frequently relying on ad hoc measures, limited guidance, and facing potential sanctions for perceived failures. This lack of cohesion in policy and approach has not only fragmented efforts across platforms but also hindered the development of cross-platform solutions. Consequently, the ability to share learning, resources, and best practices remains limited, perpetuating inefficiencies and inconsistencies in addressing extreme right content. Solving this core problem requires not only a deeper understanding of hybrid systems but also an awareness of the need to develop collaborative frameworks that align digital governance. This section offers a brief overview of the policy landscape of content moderation.

By default, extremist content moderation has historically been spearheaded by mainstream platforms, with social media giants like Facebook and Twitter shouldering much of the responsibility. During the "golden age" of the Islamic State (2013–2014), the group's sophisticated use of social media to propagate its ideology exposed significant gaps in existing digital governance structures (Conway et al., 2020). With no pre-existing frameworks tailored to countering online extremism, platforms were compelled to innovate their content moderation strategies in real time. These initial efforts laid the groundwork for the hybrid human-automated moderation frameworks that dominate today.

Zittrain (2019) outlines the evolution of digital governance through three distinct eras: the *rights era*, the *public health era*, and the *process era*. The rights era prioritised free expression, with platforms reluctant to impose restrictions that could be seen as stifling democratic values. This approach inadvertently allowed extremist groups like ISIS to capitalise on the permissiveness of online platforms. The subsequent public health era marked a shift toward balancing free expression with the need to mitigate harm. In this phase, platforms began

imposing restrictions to counter risks such as extremism. The current process era reflects the recognition that not all views can be accommodated, prompting platforms to adopt procedural mechanisms to address disputes and enforce standards effectively. Douek (2021) similarly categorises this evolution, framing it as a shift from a "posts-as-trumps" model to one centred on proportionality and contextual decision-making. In response to the challenges posed by extremist content, platforms have developed hybrid moderation systems combining algorithmic tools and human oversight (Van der Vegt et al., 2019).

As explained in the previous section, this two-pronged approach leverages automation for scalability while retaining human moderators for nuanced decision-making. Automation excels at identifying patterns and processing large volumes of data but struggles with contextual understanding, which human moderators provide. However, the unassisted nature of these efforts has led to unique challenges, including inconsistent application of standards and a lack of transparency (Koetsier, 2020).

The increasing demand for improved content moderation has emerged from diverse sources, including civil society organisations such as Glitch (2024) and the Centre for Countering Digital Hate (CCDH, 2024), as well as non-governmental initiatives like Stop Hate for Profit (Color of Change, 2024) and Sleeping Giants (2024). These groups reflect a broader civil society concern regarding the harm caused by extremist content online. While their advocacy has successfully highlighted the urgent need for action, the solutions proposed often fall short of addressing the complexities involved in moderating extreme right content. This gap underscores the need for a systematic approach to support online platforms in developing, implementing, and refining content moderation policies.

Since 2018, many Western nations have introduced legislation aimed at addressing harmful online content, each offering varying degrees of success and significant limitations. For example, Germany's Network Enforcement Act (NetzDG), which took effect on 1 January 2019, set a precedent by imposing fines of up to €50 million on platforms that fail to remove illegal content promptly (Network Enforcement Act, 2017). Similarly, Australia's Criminal Code Amendment (2024), passed after the Christchurch attacks, criminalised the sharing of harmful material and introduced provisions for fines and even criminal liability for non-compliant platform employees. In France, the Avia Law (European Digital Rights, 2020)

mandated the removal of hateful content within 24 hours of being flagged, though its broad scope was ultimately deemed unconstitutional (Conseil Constitutionnel, 2020). Despite their differences, these legislative measures share common challenges. They frequently prioritise takedown rates without addressing broader issues such as cross-sector collaboration, transparency, and the ethical implications of content removal. Additionally, their emphasis on financial sanctions often fails to consider the disparities between resource-rich platforms and smaller operators, potentially disadvantaging less-resourced platforms that struggle to comply with rigorous legal standards.

Contrasting with punitive legislation, the EU Code of Conduct to Tackle Online Hate offers an alternative model focused on collaboration rather than sanctions (European Commission, 2022). Developed in 2016 following the Brussels terrorist attacks, this code encouraged platforms to review flagged content within 24 hours while preserving freedom of speech. Although non-binding and without penalties for non-compliance, the code exemplifies how platforms can take responsibility for moderation while maintaining a balance between regulation and expression. This collaborative approach underscores the potential for frameworks that support, rather than penalise, platforms in their moderation efforts.

As discussed in the first part of this literature review, the global variability in definitions of unlawful and harmful content further complicates the policy landscape. For instance, Germany's strict prohibition of Nazi propaganda contrasts sharply with the broader free speech protections under the US First Amendment. Similarly, unique national regulations, such as Spain's prohibition on insulting the King (Amnesty International, 2021), highlight the difficulties in achieving a universal standard for online harms. These disparities also affect the operational capabilities of content moderation systems. Terms like "hate speech" often overlap with extremism, creating ambiguities that automated moderation tools struggle to resolve. For instance, machine learning systems designed to flag extremist symbols in images may not be equipped to interpret nuanced language in audio files or complex visual contexts. This lack of standardisation hinders the development of scalable, effective content moderation strategies, leaving platforms to navigate an inconsistent regulatory environment with limited external support.

While academia and industry have made strides in building an understanding of what is defined as extreme right content and in understanding content moderation systems, a critical gap persists in providing practical support for platforms, particularly in moderating extreme right content. Notably, Tech Against Terrorism's (TAT) Online Regulation Series (Tech Against Terrorism, 2023b) has offered valuable resources to clarify the global regulatory landscape. Similarly, academic contributions, such as Reed and Henschke's (2021) co-regulation approach, provide theoretical insights into balancing accountability and governance. However, these efforts often stop short of delivering actionable frameworks tailored to the operational realities of online platforms. This thesis seeks to address this gap by developing a hybrid framework that combines the scalability of automated systems with the nuanced judgment of human moderators. The proposed framework will be designed to account for the unique challenges of moderating extreme right content, such as its reliance on coded language, cultural symbols, and evolving tactics. By integrating best practices from existing legislative and collaborative models, this research will provide online platforms with practical tools to navigate the fragmented regulatory environment while aligning with ethical and legal standards. To best undertake this, the resulting challenge that has been unearthed in this section, that of the role of social media platforms in regulating extreme right content, will be investigated further.

#### Platform Responsibilities, Limitations, and Support

The regulation of online content remains inconsistent, with significant variability in policies and practices across platforms. This lack of uniformity is compounded by the absence of legal requirements mandating platforms to disclose their content moderation strategies. While some larger platforms, like Facebook, proactively report on their moderation practices, these reports are rarely subject to external verification, raising questions about their reliability. On the other hand, smaller platforms, particularly those prioritising freedom of speech, often provide little to no transparency about their processes. These discrepancies highlight the fragmented nature of content moderation and its reliance on platform-specific approaches.

At one end of the spectrum is Facebook, whose parent company Meta has developed a comprehensive and resource-intensive model for content moderation. Meta employs a multi-

tiered governance structure involving its Content Policy team, Integrity team, Global Operations team, and Research team. These teams collectively create, enforce, and refine the company's "Community Standards" and "Community Guidelines," which address complex issues such as hate speech, terrorism, and violence (Meta, 2022). Enforcement mechanisms are supported by policies targeting Violence and Incitement, Dangerous Organisations and Individuals, Hate Speech, and Graphic Content. This integrated approach enables Facebook to define and act against extremist content systematically. However, this resource-intensive model is feasible only for large platforms with significant financial and technical capabilities. Conversely, Telegram exemplifies a minimalist approach to moderation. Known for its emphasis on privacy and security, Telegram has historically resisted regulatory practices but has gradually introduced limited self-regulation (Counter Extremism Project, 2017; Europol, 2019). Telegram prohibits users from promoting violence on publicly viewable channels but provides little insight into its enforcement mechanisms. Its minimalistic approach is representative of many smaller platforms like Gab and Parler, which prioritise user privacy over proactive moderation. These differences highlight the challenges smaller platforms face in balancing regulatory demands with their operational and philosophical priorities.

To address content moderation gaps, some platforms have implemented oversight mechanisms such as human rights reviews and independent boards. For instance, Facebook's Oversight Board evaluates and reviews moderation practices, aiming to balance freedom of expression with effective content regulation (FOB, 2023). However, such initiatives often lack tangible evidence of their impact and are prohibitively expensive for smaller platforms (Reed and Henschke, 2021). The trend of downsizing trust and safety teams, as observed in 2023, further undermines these efforts, reflecting a broader tension between corporate cost-saving measures and ethical commitments (Motyl and Ellington, 2024). While oversight boards showcase a commitment to ethical regulation, their limited adoption highlights the disparity between well-resourced platforms and smaller entities. This gap emphasises the need for scalable solutions that can support platforms of varying sizes and capacities.

Recognising the limitations of platform-specific approaches, multi-organisational initiatives have emerged to provide shared resources and guidance. Prominent examples include the Global Internet Forum to Counter Terrorism (GIFCT) and TAT. These initiatives focus on

improving policies, providing tools, and fostering collaboration among platforms. TAT launched its mentorship programme in 2018 to support platforms in combating terrorist content while respecting human rights. This programme includes in-depth reviews of terms and guidelines, content moderation practices, and transparency reporting. By identifying strengths and areas for improvement, TAT provides smaller platforms with critical resources to develop effective policies and protect their users from extremist misuse. One of the benefits of the TAT membership program is that it qualifies online platforms to become GIFCT members. Therefore, the key difference between what is offered by TAT and GIFCT in the context of policy support is that to access GIFCT's support, online platforms must first qualify. However, upon accessing membership status, online platforms have access to community standard support (GIFCT, 2024b) and access to the Definitional Framework and Principles Project (GIFCT, 2024c). Most notably, the Definitional Frameworks and Principles project helps members understand and apply definitions of terrorism and violent extremism. In doing so, this tool allows platforms to understand the elements that form a definition, including understanding the context these definitions are used and the impact that has. This tool likely offers online platforms the means to develop their own definitions with the necessary information to do so.

The GIFCT was established in 2017 by Twitter, Facebook, Microsoft, and YouTube to combat the exploitation of online platforms by terrorists and violent extremists (TVE). Initially a collaborative effort among major tech companies, GIFCT evolved into an independent, not-for-profit organisation in 2020. This transformation enabled GIFCT to operate with a multidisciplinary approach, incorporating contributions from industry, government, civil society, and academia to address the complexities of online extremism. GIFCT aims to uphold and protect human rights, such as freedom of expression, while neutralising extremist activities online. GIFCT provides tools and resources that support cross-platform collaboration and content moderation. Key among these are the Content Incident Protocol (CIP) and the hash-sharing database, which have become significant assets in the fight against extremist content. The CIP is designed to act as a rapid-response mechanism during extremist events, alerting member platforms to the circulation of related content in near real time. This allows platforms to quickly identify and remove harmful material, preventing its proliferation. The

CIP is especially effective for high-traffic incidents where large volumes of content are generated in short periods.

The hash-sharing database is GIFCT's flagship initiative, enabling member platforms to securely share verified extremist content in an encrypted format. This tool supports the efficient identification and removal of harmful material by assigning each piece of content a unique "digital fingerprint" or hash. These fingerprints allow platforms to corroborate and detect extremist content without retaining personally identifiable information. The database, which has grown from 200,000 entries in 2019 to over 320,000 by 2021 (GIFCT, 2021), facilitates automated content removal and promotes cross-platform collaboration. This initiative also aids in establishing consensus definitions of extremist content, providing smaller platforms with resources they might otherwise lack. Despite its benefits, the hash-sharing database raises ethical concerns. The use of perceptual hashing is designed to account for minor variations in content, which can, in turn, open the door for increased cases of false positives (Du, et al., 2020). For instance, media documenting human rights abuses or protest footage intended to raise awareness may inadvertently be flagged as extremist. Such incidents have already been observed, including during the Black Lives Matter protests, where videos of police altercations were removed by algorithms mistaking them for harmful content (Alexander, 2020). Similarly, content related to human rights abuses in Myanmar was taken down as part of efforts to curb hate speech, despite its importance in documenting atrocities (Morada, 2023). These examples underscore the need for balancing efficiency with nuanced, context-sensitive moderation practices. GIFCT has also established GNET, its research arm, to improve understanding of extremist exploitation of the internet (GNET, 2024). GNET's role is to produce and promote research that informs content moderation strategies. By bridging the gap between academia and practice, GNET helps platforms address the evolving tactics of extremists.

GIFCT's approach highlights the challenges of balancing automated and human moderation. While tools like the hash-sharing database facilitate scalable content moderation, concerns about overreach and potential censorship persist. The risk of creating a monopolised standard for online speech is particularly acute, as shared tools and definitions may unintentionally suppress legitimate discourse (Llansó, 2020). Moreover, while GIFCT's initiatives promote

collaboration among large platforms, smaller platforms often face barriers to entry, such as meeting the qualifications required for membership. In conclusion, GIFCT exemplifies how multi-organisational efforts can enhance content moderation practices by fostering collaboration, sharing resources, and establishing frameworks for addressing extremist content. However, the ethical implications and potential limitations of these tools, particularly regarding false positives and access disparities, highlight the ongoing need for critical evaluation and adaptation to ensure both effectiveness and fairness in moderating online extremism.

Smaller-scale initiatives also contribute to content moderation efforts. These initiatives often focus on flagging extremist material for removal, leveraging expertise from both governmental and private organisations. The European Union's Internet Referral Unit (EU IRU) identifies and investigates harmful content online, referring flagged material to relevant stakeholders (European Union, 2023). Since its creation, the EU IRU has analysed over 42,066 pieces of content, resulting in 40,714 actions across more than 80 platforms in multiple languages. Similarly, TAT has developed the Terrorist Content Analytics Platform (TCAP), which combines open-source intelligence with AI-driven processes to detect and record verified terrorist content. Since its launch in 2020, TCAP has sent over 26,000 alerts containing 44,000 unique instances of terrorist content to 105 platforms.

Both the EU IRU and TCAP provide valuable support to online platforms, particularly smaller ones, by addressing gaps in their moderation systems. These organisations highlight content that may have bypassed moderation processes, ensuring a more comprehensive approach to identifying and removing harmful material. Although these initiatives may not operate at a scale sufficient to replace platform-specific systems, they provide critical supplementary support and insight into content moderation practices. Whilst these examples consider the improvement of pre-existing structures (improving pre-existing definitions, policies, and modes of regulation), online platforms must make a start somewhere to create policies and regulations to counteract online extremism.

As previously discussed, one of the foundational complexities of this task is that there is no global consensus on what extremism is and who extremists are. Resources such as the

European Union's and United Nation's terrorist designation lists would offer clarity on what specific persons, groups and entities are recognised as terrorist and extremist were they accessible to the public eye (Council of the European Union, 2024). However, many countries do not subscribe to these lists, such as in the cases of the UK and US, which have their own designation lists. From the perspective of an online platform, being provided clarity on who to protect against enables the development of specific and targeted policies and regulations. Having these qualities present in policy and regulation is essential in ensuring the fast removal of proscribed content and protects users from moderation practices overspilling into unintended cases. Taken together, these multi-organisational initiatives underscore the value of collaboration in addressing the complexities of online extremism. By providing resources, tools, and policy guidance, organisations like GIFCT and TAT help platforms navigate the challenges of content moderation, supporting efforts to foster safer and more accountable online environments.

This thesis seeks to address the gaps in the existing landscape of content moderation by developing a nuanced framework that supports platforms of all sizes. By integrating insights from initiatives like TAT and GIFCT and proposing scalable, ethically grounded solutions, this research aims to bridge the disparities between resource-rich platforms and smaller entities. The framework will prioritise transparency, collaboration, and adaptability, providing platforms with the tools they need to navigate the complexities of moderating extreme right content effectively. The disjointed policy landscape for moderating extreme right content highlights the challenges platforms face in implementing consistent and effective practices. While larger platforms like Facebook demonstrate the potential of resource-intensive models, smaller platforms like Telegram underscore the limitations of minimalist approaches. Collaborative initiatives like GIFCT and TAT offer valuable support, but their reach and impact are both hindered. In addressing these organisational gaps, this thesis aims to contribute to a more unified and equitable approach to content moderation. Through the development of an adaptable framework and comprehensive policy recommendations, this research seeks to advise platforms on how to better respond to the evolving challenges of extreme right imagery, fostering a collaborative and effective digital governance ecosystem.

## Balancing Human Rights

The fractured and complex nature of online regulation presents profound challenges, particularly in balancing the regulation of extremist content with the preservation of fundamental human rights, such as freedom of speech and expression. Poorly designed or implemented regulations for online speech can be invasive and carry significant consequences for civil liberties. Counter-extremism measures, in particular, have historically been exploited to infringe upon rights, necessitating a critical examination of how content moderation intersects with human rights frameworks. To address this effectively, it is vital to explore the relationship between online speech regulation and human rights laws, weighing the trade-offs and implications inherent in moderating extremist content.

The need for clarity in online speech regulation is recognised not only by social media users but also by the platforms themselves. Since 2017, a surge in legislation targeting online extremism has reflected broader regulatory efforts grappling with freedom of expression. These contemporary measures build upon the principles of international human rights law (IHRL), which provide a foundational global framework for safeguarding individual rights. However, IHRL has not been systematically employed as a guiding framework for content moderation, often leaving platforms uncertain about the broader human rights impacts of their policies and how to achieve a balance between competing imperatives. IHRL encompasses treaties such as the Universal Declaration of Human Rights (UDHR, 1948), the International Covenant on Civil and Political Rights (ICCPR, 1966), and the International Covenant on Economic, Social and Cultural Rights. These documents define individual rights and provide critical guidance on how content moderation could and should be conducted. Their global scope aligns well with the transnational nature of online platforms, making them more suitable than narrower national laws. David Kaye, former UN Special Rapporteur on Freedom of Expression, advocates for prioritising IHRL over national laws in internet governance, emphasising its global applicability and suitability for the borderless digital environment (Dvoskin, 2023).

Aligning platforms with IHRL offers several advantages. First, IHRL provides a globally recognised ethical framework that resists the encroachment of overly restrictive content moderation policies, as highlighted by Benesch (2020) and Sander (2020). Second, IHRL's

inherent flexibility allows for contextual interpretation, enabling platforms to tailor their policies to unique challenges while adhering to universal standards. For example, Article 19 of the ICCPR guarantees the right to freedom of expression but recognises it as a limited right. Under Article 20, this right can be restricted to prevent propaganda for war, advocacy of hatred, or incitement to violence, discrimination, or hostility. These provisions grant platforms legal grounds to act against harmful content while leaving room for nuanced case-by-case decisions. Such flexibility is essential in addressing the complexities of online speech. However, IHRL also has limitations. Its broad scope was not designed to account for the speed and scale of communication enabled by the internet, particularly in the context of extremist speech. For instance, IHRL does not address risks stemming from human or algorithmic errors in content moderation, nor does it provide mechanisms to safeguard individuals from the consequences of such errors. Additionally, platforms may lack the capacity to assess speech against IHRL standards with precision, further complicating implementation.

One significant challenge related to the lack of a global approach to content moderation policy is the reliance on U.S.-based platforms on the First Amendment of the US Constitution as the default framework for moderating speech. Shifting to IHRL-based approaches introduces conflicts for platforms like Facebook, YouTube, and Twitter/X, which operate within US legal and cultural contexts. Furthermore, commercial imperatives pose additional hurdles. As profit-driven entities, platforms are motivated to maximise user engagement, which can conflict with restrictive content moderation measures. Balancing freedom of expression with moderation policies raises questions about whether it is realistic, or fair, to expect platforms to adopt frameworks that might undermine financial goals. A related issue involves the scalability of regulatory expectations across platforms of varying sizes and resources. Larger platforms like Facebook, with their robust legal and financial infrastructure, can invest in comprehensive content moderation frameworks, whereas smaller platforms, such as JustPasteIt, may lack the resources to do so (Weimann and Vellante, 2021). Striking a balance between proportionality and effectiveness in regulation is, therefore, essential. Regulatory frameworks must account for disparities in platform capacities to ensure fairness and efficacy.

This thesis argues that addressing the challenge of balancing human rights with content moderation requires a nuanced, multi-layered approach. Policymakers must recognise the reality of the policy landscape we find ourselves in today with its limitations of both IHRL and national laws. Expecting global human rights frameworks to be perfectly adhered to is practically difficult, given the regulatory landscape that exists today. Therefore, there is a need for tailoring and supporting solutions that reflect the varying capacities and contexts of platforms. Moreover, regulatory frameworks must balance ethical imperatives with practical feasibility, ensuring that measures to combat extremism do not come at the expense of fundamental freedoms. By interrogating these tensions, this research seeks to advance equitable and effective content moderation policies that respect human rights while addressing the critical challenge of online extremism and the extreme right more specifically.

## Conclusion

This chapter has explored the multifaceted challenges of moderating extreme right online propaganda, contextualising the complexities of defining, analysing, and addressing this form of content. It has established that the extreme right's use of digital platforms, particularly its reliance on imagery, is a sophisticated and evolving phenomenon. By leveraging visual culture, memes, satirical images, and coded symbols, extreme right groups have developed strategies that mask their intentions, evade content moderation, and radicalise audiences. These tactics underscore the importance of systematically understanding and analysing extremist imagery, particularly given its ability to transcend platforms and influence cross-platform behaviours.

This chapter has highlighted significant gaps in the existing literature that this thesis seeks to address. There is a limited exploration of how extreme right imagery evolves and circulates across platforms, particularly within decentralised ecosystems like Telegram, where much of this content thrives. The use of layered communication strategies, including humour, irony, and coded symbols, further complicates detection and remains underexplored in academic and industry discourse. Existing definitions of extremism, violent extremism, and terrorism often fail to account for the operational nuances of online extremist imagery. This limitation hinders the development of effective content moderation strategies, as platforms struggle to

apply consistent and actionable frameworks. Moreover, the variability in national and international definitions of extremism exacerbates this challenge, leaving platforms to navigate an inconsistent and fragmented regulatory landscape with minimal external support.

This chapter has introduced several key research questions that will guide the thesis. These research questions have been chosen over many potential questions that have emerged from this chapter due to their particular use in developing an effective framework for capturing extreme right imagery:

**1. To what extent is imagery used as a medium of communication by extreme right groups?**

This question explores the centrality of visual content in extreme right propaganda, focusing on its role in recruitment, ideological dissemination, and community-building.

**2. To what extent is graphic imagery used in the communicative strategies within extreme right imagery?**

This question investigates the specific use of graphic content and its significance in conveying extreme right ideologies and narratives.

**3. Is there an extreme right visual style that is communicated and reinforced through the use of imagery?**

This inquiry examines whether extreme right groups develop and propagate a cohesive visual style that contributes to their broader ideological and operational goals.

**4. Do extreme right groups have differing visual styles that are communicated and reinforced through the use of imagery?**

This question delves into the visual diversity among extreme right subgroups, exploring how their imagery reflects and reinforces distinct ideological identities.

Building on these gaps and research questions, the thesis will carry forward several critical objectives. First, it will develop a systematic framework for detecting and analysing extreme right imagery that accounts for its cultural, symbolic, and contextual nuances. This framework will be designed to adapt to the cross-platform and decentralised nature of extreme right digital strategies, ensuring its applicability beyond single-platform studies. Lastly, the research

will offer policy recommendations to address the limitations of current content moderation systems discovered in this chapter and propose enhancements to address the unique challenges posed by extreme right imagery. This includes developing solutions that are scalable for platforms of varying sizes and resources and that foster a collaborative digital governance ecosystem.

Chapter 2 has laid the groundwork for the thesis by identifying the challenges of defining and moderating extreme right online propaganda and highlighting the critical gaps in the literature. These insights establish the necessity of a systematic and adaptable framework for detecting and moderating extreme right imagery. Chapter 3 will build on this foundation by synthesising existing research on extremist visual content, offering a comprehensive overview of studies that inform the development of the proposed framework. By addressing the gaps identified in this chapter, the thesis aims to contribute to the academic discourse and provide actionable solutions for improving content moderation in an increasingly complex digital landscape.

## Chapter 3: Existing Studies Analysing Extremist Images

### Introduction

The study of extremist imagery has evolved in recent years, with increasing academic attention directed toward the unique visual strategies employed by extreme right groups. Early research into extremist imagery largely focused on Islamist propaganda, emphasising its print and press, text-heavy and graphic ideological narratives (O'Halloran 2015; O'Halloran et al., 2019). However, as online platforms have become closely linked with extremist content dissemination, the unique characteristics of extreme right imagery have become an area of study. These characteristics commonly utilise implicit cultural, historical, and nationalistic references to resonate with audiences, making their detection and analysis more complex than traditional explicit and text-reliant propaganda (Mattheis, 2022; Windisch and Simi, 2022). The aim of this literature review is twofold. Firstly, to condense insights from existing studies on extremist imagery to better understand the strategies and narratives used by the extreme right, focusing on extreme right imagery. Secondly, to evaluate the methodologies

utilised in prior research to inform the development of a robust framework for detecting extreme right imagery later in this thesis. By examining both the findings and the methods of existing studies, this review provides a foundation for constructing an adaptable and systematic analytical approach tailored to the unique challenges posed by extreme right propaganda.

Extreme right groups commonly utilise imagery to promote in-group cohesion and demonise out-groups. Imagery in this context commonly uses historical and cultural symbols to communicate ideology and utilise emotional responses (Forchtner and Kølvråa, 2017). Frameworks which analyse extreme right imagery have categorised propaganda into themes such as symbolic representation, group identity, and the use of shocking or gruesome imagery (Baele et al., 2020a). Additionally, advancements in technology have given extreme right groups the ability to adapt and innovate, allowing their content to avoid being moderated and remain online. Despite these advancements, significant gaps remain. Existing multimodal analysis frameworks are less effective for standalone, text-free visuals, which dominate extreme right imagery (Forchtner and Kølvråa, 2017). While a grounded theory approach has been used to identify patterns in extreme right imagery, the limitations of this approach reference the lack of structure necessary for large-scale analysis. This thesis addresses these limitations by adapting Baele et al.'s (2020a) framework, which incorporates inductive and deductive methods to showcase a robust analytical approach. The adapted framework aims to uncover the implicit messaging strategies of extreme right imagery and contribute actionable insights for mitigating its proliferation. In doing so, this research aspires to enhance both academic understanding and real-world counter-extremism efforts. Thus, this literature review contextualises the underpinning of the methodological choices for the thesis. It ensures the framework is theoretically grounded and practically effective in addressing the challenges posed by extreme right visual propaganda.

### [Studies that Analyse Extremist Images Online](#)

When looking at the literature in the field, images play a significant role (Caiani, 2024). Studies on activism and political participation have increasingly focused on the relationship between images and political mobilisation (e.g., Halfmann and Young, 2010; Corrigan-Brown and Wilkes, 2014). The creation of visual materials by political actors, such as photos, posters,

leaflets, and memes, can uniquely communicate messages through symbolic association, which differs from the sole use of text, which depends on rational argument (Müller and Özcan, 2007). Analysing visuals allows researchers to analyse how content can be used to uniquely surface emotions (Joffe, 2008), which plays a role in shaping political views and behaviours (Gross, 2009). Furthermore, visual analysis is not only descriptive but can highlight other trends that can lead viewers to mobilisation (Adami, 2020). Visual theorists in Media Studies and Art History note that images are linked to a bed of cultural knowledge and experiences, frames, and identifications (Milman and Doerr, 2021). This knowledge is essential for understanding the extreme right, as their imagery capitalises on these evolving factors to reinforce group identity and mobilise supporters.

Several studies have highlighted how extreme right imagery differs from the explicit and text-heavy strategies typical of Islamist propaganda. Extreme right imagery often relies on implicit visuals, including nationalistic symbols and historical references. For instance, studies using grounded theory approaches have described how the extreme right uses these strategies in their fragmented and decentralised communication. Windisch and Simi (2022) found that these groups frequently draw on historical and cultural symbols, such as national flags, monuments, or mythologised historical figures, to resonate with audiences who share a common cultural or ideological background. This reliance on implicit messaging allows extreme right groups to communicate complex ideological messages while avoiding breaking online platforms terms of service. Similarly, Mattheis (2022) demonstrated how such visuals often depend on audience familiarity with specific cultural or historical contexts to fully communicate their intended meanings. This reliance on shared cultural knowledge underscores the role of context-specific semiotics in interpreting these images.

In addition to symbolic imagery, extreme right propaganda heavily emphasises fostering a sense of in-group identity while vilifying perceived out-groups. Baele et al. (2020a) categorised these strategies within a broader framework of visual propaganda analysis, identifying key recurring themes such as identity construction, dehumanisation of opponents, and the strategic use of graphic or shocking imagery to provoke emotional reactions. Their analysis highlights how visuals are designed to present pride, solidarity, or nostalgia among in-group members while portraying out-groups as existential threats or morally inferior.

Although initially developed to analyse Islamist propaganda, Baele et al.'s framework evidence an approach to understanding extremist contexts. The structural and thematic consistencies identified make it particularly valuable for adaptation to the study of extreme right imagery.

This is further supported by the concept of 'imaginaries', which represent the idealised, desired, and utopian presentations found in visual and social contexts. Imaginaries, as described across disciplines such as psychoanalysis (Lacan, 1944), political theory (Castoriadis 1987), and sociology (Taylor 2004), refer to how societies and groups envision and positively imagine structures and relationships. Imaginaries relate to visual representations because they idealise specific social relationships and roles depicted within the imagery. Building on Kress and van Leeuwen's framework (2020), images construct different types of imaginary social relationships through three key layers. First, they create an imagined connection between the content producer and the audience, which is inherent in all forms of visual communication. Second, they establish a relationship between the viewer and the subject represented in the image, particularly when the subject of the image addresses the viewer directly, e.g. through eye contact. Third, images involving multiple subjects construct relationships among those subjects, which the viewer observes as an 'invisible onlooker.' These layers of interaction encourage viewers to see themselves in specific roles, this enriches emotional and ideological engagement with the imagery. By using strategies such as this, extreme right imagery reinforces in-group identity while leveraging imaginaries to shape perceptions of both the in-group and the out-group.

Developments in technology have further enabled extreme right groups to innovate their visual communication. As noted by TAT (2022), the extreme right leverages the capabilities afforded by digital platforms to disseminate their content whilst working to avoid content moderation. The continued presence of extreme right content online is partly due to the decentralised nature of the extreme right as well as the savvy use of platform-specific characteristics, such as hashtags, memes, and video edits, which allow messages to blend seamlessly with mainstream media content. These strategies make it increasingly difficult for automated moderation tools to distinguish extremist material from benign content.

The work of Ingram (2014, 2015) and Milton (2016) sheds light on the strategic deployment of high-quality visual propaganda by extremist groups. Their research highlights how groups like ISIS have employed cinematic techniques, dramatic compositions, and professional editing to amplify the impact and reach of their messages. This focus on production quality serves to legitimise and reinforce extremist ideology to their target audiences. While these studies primarily focus on Islamist groups, they provide valuable insights into how extremist groups in general, including the extreme right, use visuals to project strength and authority. Therefore, it helps fill the gap of understanding that lies in a lack of extreme right imagery analysis. The deliberate blurring of lines between extremist propaganda and mainstream media content exemplified in their research further complicates detection and counter-extremism efforts.

Similarly, Forchtner and Kølvråa (2017) carry out a comprehensive analysis of the extreme right's use of imagery, particularly through the lens of radical authenticity. This analysis found that extreme right groups often use multimodal aesthetics to convey their messages, focusing on themes such as history, nature, and traditional gender roles. These visuals are carefully used to appeal to audiences' emotions and ideologies. By leveraging narratives of authenticity, the extreme right amplifies its ideological appeal, particularly among younger audiences. Forchtner and Kølvråa also highlight how images in this context contain symbolic meanings to resonate with audiences. For example, historical references in visuals can portray a romanticised view of the past, while presentations of perfect-looking landscapes may symbolise purity. Traditional gender roles, often portrayed through idealised images of masculinity and femininity, can be used to reinforce a vision of societal order that aligns with their ideology. Beyond the embedded ideology, images can also present the polished aesthetic seen in mainstream media, making content appear both relatable and credible. By mirroring mainstream media styles, these images can also enhance their reach and blur the boundaries between extremist content and popular culture. In turn, this makes it harder for online platforms to identify and remove the content. This study by Forchtner and Kølvråa's highlights the need for a nuanced analysis of extreme right imagery, as the array of strategies is key to extremists looking to normalise their ideologies through the appearance of authenticity.

Despite these approaches to understanding extreme right imagery, significant gaps remain. Despite the significance of images in political communication, empirical research has often overlooked visual data when analysing extreme right movements (Caiani 2024). Existing multimodal analysis frameworks, such as Systemic Functional Multimodal Discourse Analysis (SF-MDA), while effective for studying text-heavy content, are less suited to minimal, text-free visuals that are present in extreme right communication. Frameworks primarily concerned with text are at risk of overlooking the symbolic and implicit messaging strategies central to extreme right imagery. While grounded theory approaches have provided valuable exploratory insights into patterns and themes within extreme right imagery, they often lack the deductive structure necessary for consistent and large-scale analysis. This limitation becomes particularly evident when attempting to analyse extensive datasets, where an iterative and structured framework is essential for identifying recurring motifs and trends.

By integrating both inductive and deductive methods, this study aims to develop a robust analytical approach that captures the nuanced strategies employed in extreme right visual propaganda. This dual-method approach ensures that the framework is adaptable to new and trending patterns while grounded in established theoretical principles. This will provide a deeper understanding of how extreme right groups utilise imagery to advance their ideologies, particularly through implicit messaging strategies and decentralised dissemination networks. Furthermore, this research will contribute actionable insights for mitigating the proliferation of harmful content online, addressing both academic and practical challenges in countering extremist propaganda. To achieve this, the next section of this chapter will shift focus from an overview of research on extreme online imagery to the specific methods employed in the studies discussed above. By examining these methodological approaches in greater depth, this thesis seeks to identify strengths, limitations, and areas of innovation that will inform the development of a tailored framework for detecting and analysing extreme right imagery. This transition ensures that the chosen methodology is not only grounded in theoretical insights but is also practically effective in addressing the unique challenges posed by contemporary extreme right imagery.

## Methodologies and Findings

### Language and Imagery in Extremist Analysis

A significant thread in the literature on extremist propaganda is the analysis of the interplay between textual and visual elements (O'Halloran 2015; O'Halloran et al., 2019; Tan et al., 2018). This dual focus stems from the recognition that extremist groups strategically utilise multimodal communication to amplify their messages, enhance emotional resonance, and increase ideological impact. Research by O'Halloran et al. (2019) and Tan et al. (2018) have contributed to highlighting how semiotics (the study of signs and symbols) intersects with textual narratives in Islamist extremist propaganda. Their studies demonstrate that visuals rarely function in isolation; instead, they are often designed to frame, reinforce, or elaborate on accompanying text. For example, Islamist extremist propaganda frequently uses visual cues such as symbolic imagery, colour palettes, and structured layouts to portray themes of heroism, victimhood, and legitimacy, among others. These visuals link to textual narratives, creating a cohesive message that appeals to the emotions and identities of the audience. In this analysis, O'Halloran et al. highlighted the pairing of text and imagery being used to craft compelling narratives, such as portraying fighters as heroic martyrs or depicting adversaries as oppressive and inhumane.

Wignell et al. (2017) extended these insights by applying SF-MDA, as conceptualised by Jewitt et al. (2016). SF-MDA builds upon Halliday's Systemic Functional Theory (Halliday, 1978; Halliday and Matthiessen, 2014), which focuses on how language operates as a social semiotic resource within specific contexts. SF-MDA broadens this focus to include visual and other semiotics, analysing how they interact with language to present meaning. Using this framework, Wignell et al. explored how Islamist extremist groups use images and text to evoke emotional and ideological engagement. Key semiotic elements such as colour, composition, iconography, and spatial organisation were identified as central to the creation of powerful propaganda. For example, the use of dark or vibrant colours in imagery can portray fear, hope, or anger, depending on the context. Similarly, the inclusion of religious symbols or cultural motifs reinforces textual claims of divine authority or historical continuity. SF-MDA allows researchers to unpack the complex layering of meaning within propaganda, showing how visuals and text work together to appeal to the viewer's emotions, intellect, and identity. This multimodal approach has proven particularly effective in deconstructing the

elaborate messaging strategies of groups like ISIS, whose propaganda often combines cinematic-quality visuals with ideologically loaded text.

Despite the strengths of these methodologies, their reliance on the presence of textual elements limits their applicability in certain contexts, particularly when analysing extreme right propaganda. Unlike Islamist extremist groups, which often pair visuals with explicit textual narratives, extreme right propaganda frequently relies on standalone images to convey its ideological messages. These images often contain implicit meanings that draw on cultural, historical, or nationalistic references recognisable to the audience. For example, an extreme right image may feature a lone figure standing against a stormy backdrop, accompanied by nationalistic symbols such as flags or historical emblems. While the imagery evokes themes of resistance or defiance, the absence of textual explanation means that the message relies on the audience's ability to interpret these cues based on their cultural knowledge or ideological alignment. Traditional multimodal methodologies like SF-MDA struggle to address such text-free content, as they are designed to analyse the interaction between visual and linguistic elements. This limitation underscores the need for an image-focused analytical framework that can effectively capture the implicit messaging strategies often employed by extreme right groups. Such a framework must be capable of dissecting the visual semiotics of standalone images, identifying patterns, and interpreting their ideological functions without relying on accompanying text. By focusing solely on the visual elements, researchers can gain deeper insights into how extreme right propaganda operates within text-free datasets, addressing a critical gap in the existing literature.

Using a multimodal approach, Forchtner and Kølvråa (2017) captured meanings embedded in standalone images, showcasing how themes such as nostalgia, nature, and gender roles are moulded to appeal to specific audiences. This was achieved using a methodology combining network mapping and thematic categorisation with multimodal semiotic analysis. Starting with a seed node, "Balaclava Küche," a radical vegan nationalist group, they used snowball sampling to identify a network of extreme right Facebook pages. Using this dataset, the visuals were grouped into three key themes: (1) Historical References to the Third Reich, emphasising ideological continuity; (2) Nature and Naturalness, linking classical fascist ideals of purity to contemporary movements like veganism; and (3) Gender Roles and Family Life, portraying traditional values while engaging with modern cultural dynamics.

Their analysis relied on Multimodal Critical Discourse Analysis (MCDA), drawing on frameworks by Kress and van Leeuwen (2020), to decode semiotic factors such as composition, salience, and colour schemes. Importantly, Forchtner and Kølvråa introduced the notion of an *authoritative imaginary*, which represents hierarchical relationships and ideological discipline, and the *intimate imaginary*, fostering emotional resonance through personal and relational themes. This framework allowed the authors to analyse how images establish relationships with viewers through explicit demands for conformity or implicit appeals to intimacy and authenticity. However, the reliance on visual semiotics and the absence of accompanying text in much of the analysed material highlights a challenge that remains even with a more nuanced methodology, as interpretation often depends on the analyst's experiences, biases, and expertise, introducing a degree of subjectivity. Despite this, the study contributes to understanding extreme right propaganda by capturing ideological and emotional messaging present in visual communication. Simultaneously, this work also paves the way for further methodological innovation to address the unique characteristics of text-free datasets.

Caiani's 2024 study develops a framework to analyse the functionality of visuals in extreme right communication, with a focus on their discursive meaning and communicative function. This approach addresses how extreme right visuals strategically craft narratives to engage their audience. Discursive meaning examines the ideological and emotional content of visuals, such as their use of cultural, historical, and nationalistic symbols used to promote in-group cohesion and demonise the out-groups. The communicative function investigates how visuals increase visibility and mobilise support, exploring how they connect with audiences to create and promote a collective identity and belonging. The methodology uses a qualitative multi-method approach, drawing on visual data from extreme right groups' online platforms, websites, and offline materials. Through systematic content analysis, recurring themes and symbols were identified. The dataset was linked to socio-political events, including migration crises and political campaigns, showing its real-world relevance. Reactions to the dataset on social media were analysed to explore how content was received, adapted, and promoted by audiences. This comprehensive approach provides valuable insights into the interplay between the creators' intentions and the audience's reception of visual content.

### Theory-Driven Frameworks and Narrative Framing

A significant focus of the literature on extremist imagery has been the development and application of theoretical frameworks to explore specific ideologies and communication strategies within propaganda. These frameworks, which often involve testing theoretical hypotheses, are instrumental in unpacking the visual and narrative strategies that underpin extremist messaging. Baele et al. (2020a) provide a comprehensive and structured framework for analysing ISIS visual propaganda, categorising it into four central themes: in-group identity, out-group vilification, symbolic representation, and gruesomeness. This framework reveals the structural consistency of ISIS's visual messaging, which seeks to promote a strong sense of group belonging while dehumanising perceived adversaries. The use of symbols, such as religious iconography or historical allusions, reinforces claims of legitimacy and continuity, while gruesomeness amplifies the emotional impact of the imagery, aiming to provoke an emotional response such as fear or admiration. The framework is effective at analysing the structural design of Islamist extremist propaganda, which increases the interpretability of the content.

Social movement theory has also been a pivotal framework for examining extremist imagery. Neuman and Rogers (2007) and Rasoulikolamaki and Kaur (2022) explored how visuals are used for collective action and recruitment by using content to form narratives that resonate with the audience. For instance, Rasoulikolamaki and Kaur (2022) identified dualistic imagery in ISIS propaganda, where fighters are portrayed as heroic warriors and as oppressed victims. This portrayal is used to garner support and legitimise the group's actions by framing them as a necessary response to external threats and injustices. Similarly, Neuman and Rogers (2007) highlight how the careful curation of visual narratives fosters a sense of collective identity and shared purpose, which is essential for extremist mobilisation.

Innovation in extremist visuals is another key area of exploration. Ingram (2014, 2015) and Milton (2016) examined the strategic use of high-quality imagery in ISIS propaganda, emphasising its role in ideological dissemination and coordination. In using cinematic techniques, dramatic compositions, and key visual themes, ISIS increased the palatability of its content to a global audience. These studies underscore the group's deliberate efforts to

innovate its visuals, using sophisticated production methods to increase the appeal and impact of its propaganda. These innovations also enabled the group to control its reputation as a modern and formidable entity.

While these studies provide valuable insights into Islamist extremist propaganda, a gap is present in the application of similar frameworks to extreme right imagery. Extreme right propaganda often operates through fragmented and decentralised networks, utilising implicit messaging and cultural symbols to communicate. Unlike Islamist propaganda, which frequently integrates text and visuals, extreme right content often uses a range of types of imagery, including the use of standalone visuals. The underrepresentation of comparable analyses of extreme right imagery highlights the need for frameworks to capture the unique characteristics of this type of content. In this vein, studies are emerging but are still limited in number. For example, despite the methodological contributions of Caiani's study, the subjective nature of analysis poses challenges, as researchers' perspectives may influence the coding and interpretation of themes. Learning from this, the framework being developed will focus on the classification and identification of categories and symbols which are less vulnerable to interpretation and bias. An iterative approach would ensure the framework remains adaptable, enabling the analysis to evolve alongside broader socio-political developments or crises. These enhancements aim to build on Caiani's work while addressing its identified gaps, ensuring a more robust and dynamic analytical tool.

By adapting existing frameworks such as Baele et al.'s (2020a) and incorporating insights from social movement theory and narrative innovation, this thesis seeks to bridge this gap. The focus on both the structural and narrative elements of propaganda ensures a holistic approach, capturing the nuanced strategies used by extreme right groups to disseminate their ideologies and garner support. This expanded framework will improve the understanding of extreme right visuals but also provide a basis for developing tools to detect and counteract its spread across online platforms.

#### Data-Driven Approaches: Grounded Theory and Inductive Analysis

Designing methodologies to analyse extreme right imagery online necessitates careful attention to ethical considerations to ensure both the integrity of the research and the well-being of those conducting it. The chapter "The Methodological and Ethical Implications of

Using Imagery as the Primary Method of Researching Online Extremism" examines the challenges and complexities associated with using visual data to study extremist activities online (Kingdon, 2024). One key recommendation that emerged from this research is to pair visual analysis with other methodologies, such as textual or network analysis, to enhance the depth and accuracy of findings and maximise the use of data. By combining methods, researchers can achieve a more comprehensive understanding of the context, dissemination, and impact of visual materials within extremist networks, whilst extracting the most out of their dataset.

Grounded theory presents an inductive approach to analysing complex and evolving content, including extremist visual propaganda. Unlike hypothesis-driven methodologies, grounded theory allows researchers to generate frameworks directly from the data. This data-driven exploratory process is well-suited to the complex nature of extremist right imagery. Windisch and Simi (2022) and Mattheis (2022) used grounded theory to analyse extreme right visual propaganda and, in doing so, found consistent themes and patterns within these images. The work highlights how extreme right imagery uses historical and cultural symbols to communicate key narratives. For example, symbols of national identity, such as flags or references to historical conflicts, are often used to appeal to promote in-group solidarity and simultaneously vilify perceived out-groups. The decentralised nature of extreme right networks, where imagery is often user-generated, highlights the necessity of an inductive approach that can adapt to the diversity of content within a dataset.

A key feature of grounded theory is its iterative refinement process, as emphasised by Charmaz (2006) and Berg (2017). This process involves several rounds of revisiting and adapting frameworks based on new findings, enabling the framework to evolve with each iteration. For example, initial analyses can be used to discover broad themes, such as nationalism, which can then be refined into specific subcategories, such as references to significant and relevant historical events. This adaptability is maximised when applied to a large dataset, where more variety and subtlety are likely to be present. However, the grounded theory approach has limitations. One major challenge is that as the framework is built from the data, there is a risk of overfitting findings to the specific dataset, limiting the generalisability of the results. This limitation is a greater concern in large-scale analyses,

where the findings can be further embedded. To offset these limitations, hybrid methodologies can be used to combine the strengths of inductive and deductive approaches and offset their limitations. Hybrid frameworks leverage the exploratory power of grounded theory while incorporating existing knowledge for the literature. For example, the iterative refinement process can be paired with a structured coding framework to allow the analysis to be flexible and methodologically rigorous. By integrating inductive and deductive elements, researchers can balance the need for data-driven discovery with the requirement for consistency and scalability.

In the context of extreme right visual propaganda, this hybrid approach is particularly relevant. The complexity of extreme right imagery means that purely deductive frameworks may be less suited to capture new and emerging patterns unique to specific datasets. At the same time, the implicit and symbolic nature of many visuals requires some interpretation that strictly inductive methods are less suited to carry out. By combining grounded theory's flexibility with deductive methodologies, researchers can develop robust frameworks that can capture the complexity and diversity of extreme right content. This thesis adopts such an approach, building on the inductive principles of grounded theory while incorporating deductive insights from established frameworks such as Baele et al. (2020a). In doing so, this study aims to contribute novel insights into the visual strategies of extreme right imagery, using this knowledge to create practical policy recommendations for online platforms to consider in their pursuit of combating harmful content online.

### Innovative Approaches and the Unique Challenges of Extreme Right Content

Extreme right imagery is uniquely challenging due to its reliance on implicit messaging and minimally text-based visuals. Unlike Islamist propaganda, which often combines text and imagery to communicate ideological narratives, extreme right content frequently uses visuals without lengthy text that relies on implicit messaging interpreted by the viewer through cultural, historical, or symbolic references. This allows extreme right groups to communicate ideologies in ways that are less immediately identifiable as extremist, enabling their content to persist online and evade content moderation efforts. As highlighted by TAT (2022), the persistence of such material underscores the need for novel approaches, such as analytical frameworks, to capture and characterise these unique characteristics.

Extreme right imagery often uses historical and nationalistic references to reinforce the in-group. For example, symbols such as flags, historical events and monuments to strengthen a sense of cultural identity and pride. These visuals also make a stark comparison to the representation of cultural decay associated with the out-group. The implicit nature of this visual communication is designed to speak to the pre-existing knowledge of the in-group, making the hidden messaging difficult to decipher for those outside the ideological community. Unlike the more direct approaches seen in Islamist propaganda, where textual narratives often accompany images to frame the message explicitly, extreme right imagery and propaganda can commonly rely on the viewer's ability to interpret subtle messaging. This strategy widens the reach of the content and makes it more difficult for online platforms to detect using content moderation tools.

Research by Ingram (2014, 2015) and Milton (2016) highlights the strategies employed by extremist groups for their online content to evade detection, a trend that is also evident in extreme right content. These strategies include the use of coded language, the re-appropriation of mainstream symbols, and the dissemination of content through decentralised networks. For example, extreme right groups adopt symbols or memes with ambiguous meanings, allowing their content to circulate widely without being flagged as explicitly extremist. The decentralised nature of these groups further complicates detection efforts, as new symbols and themes can emerge organically from online communities, making it challenging for analysts to keep up to date on recent trends. For example, the increased production quality of extreme right imagery likely comes as a result of the high-quality visuals used by groups like ISIS. The extreme right imagery can contain professionally edited graphics, styled cinematically and visually appealing. These traits work to increase the reach and impact of the content while seeking to blur the lines between extremist propaganda and mainstream media, further complicating content moderation efforts.

Given these challenges, it is clear that traditional analytical frameworks, which often rely on explicit textual cues or predefined symbolic meanings, are insufficient for analysing extreme right content. As noted by Baele et al. (2020a), to effectively analyse extremist imagery, research needs to capture the ideological nuances present within imagery and visuals. It is necessary for frameworks to focus on detecting implicit and symbolism, historical and cultural

contexts, and being adaptable to the evolving nature of the extreme right. There is reason to believe that frameworks would benefit from iterative refinement, allowing researchers to identify emerging patterns that are unique to the dataset being used. This thesis builds on these insights by developing a framework specifically for extreme right imagery. By incorporating elements of inductive analysis from grounded theory and the thematic categorisation proposed by Baele et al. (2020a), the framework aims to identify the many traits and characteristics of extreme right imagery. This framework aims to tackle the unique challenges posed by implicit messaging and evolving visual strategies. In doing so, this research contributes to an understanding of how extreme right groups use imagery to advance their ideological goals.

#### The Adaptation of Baele et al.'s Framework

Baele et al.'s (2020a) framework, originally designed to analyse ISIS visual propaganda, is well-placed to serve as the foundation for this thesis, which seeks to build a framework for detecting extreme right imagery. This framework is uniquely suited for adaptation due to its methodological robustness, its capacity to categorise key visual themes, and its ability to identify ideologies of extremist content. The following paragraphs outline why Baele's methodology provides an ideal starting point for this thesis, focusing on its theoretical foundation, adaptability, and ability to address the specific challenges of extreme right imagery.

Baele et al. (2020a) used a structured approach to analyse ISIS propaganda, categorising imagery into four key themes: in-group identity, out-group vilification, symbolic representation, and gruesomeness. This framework was developed to categorise and frame the mechanisms that underpin visual propaganda. Extreme right imagery shares similarities in their approach to imagery to groups like ISIS, for example, the promotion of in-group cohesion and demonising out-groups and the use of symbolism, such as historical or nationalistic references. The "gruesomeness" category present in Baele et al.'s framework, which includes the use of shocking or dehumanising imagery, is also relevant to the extreme right, where the extreme right's use of violence-glorifying content to provoke emotional responses. Baele's categories allow these elements, among others, to be identified and

analysed, making it an ideal adaptable foundation for analysing imagery across different extremist ideologies.

Baele's framework also provides a balanced foundation for integrating both inductive and deductive methodologies. The deductive aspect leverages established categories, ensuring that the analysis is grounded in prior research. The inductive element, inspired by grounded theory approaches such as those advocated by Charmaz (2006) and Berg (2017), enables the framework to evolve based on emerging patterns within the dataset. This hybrid approach addresses a key limitation in existing studies. Grounded theory, while effective for exploratory analysis, often lacks the deductive structure needed for systematic validation. By combining the strengths of both approaches, Baele's framework ensures methodological rigour and flexibility. Existing methodologies for analysing extremist imagery, such as Systemic Functional Multimodal Discourse Analysis (SF-MDA), are highly effective for multimodal content but less applicable to text-free datasets like those often found in extreme right imagery. Baele's framework offsets this limitation by focusing on the visual cues to make it more relevant for instances where images contain no text. Additionally, the innovative use of digital platforms by extreme right groups, as highlighted by TAT (2022), underscores the need for methodologies that can adapt to the evolving strategies of these groups. Baele's framework, with its emphasis on visual categorisation and its flexibility for adaptation, is uniquely positioned to address these challenges.

## Conclusion

This chapter has synthesised key insights from existing studies on extremist imagery, with a focus on the unique characteristics and challenges posed by extreme right visual propaganda. The analysis shows how the extreme right strategically uses imagery to promote in-group cohesion and demonise out-groups, leveraging cultural, historical, and nationalistic symbols and narratives to communicate these ideologies. While these visuals often evade traditional detection methods, they simultaneously can resonate with their target audience. The chapter also examined the methodologies utilised in existing research, including grounded theory and multimodal analysis, to identify their strengths and limitations in analysing extreme right imagery, drawing on the insights presented in the first section of the chapter. Although these approaches have provided valuable insights, they exhibit limitations, particularly when

applied to standalone, text-free visuals typical of extreme right propaganda. To address these gaps, this thesis proposes the adaptation of Baele et al.'s framework, integrating both inductive and deductive methods to develop a robust extreme right image analysis framework. This adapted framework aims to capture implicit messaging while addressing the format of the visual content. The framework will adopt an iterative process, allowing for continual refinement based on initial findings to accommodate the trends in initially drawn from the literature. This allows for the analysis of visuals to be adapted and contextualised in line with a broader understanding of the literature as well as the unique knowledge found within the dataset.

The findings presented here set the stage for the next chapter, which will focus on translating these insights into a concrete methodology. Integrating the methodological advancements discussed in this chapter ensures a more comprehensive and responsible approach to studying extreme right imagery. From having critically examined the strengths and limitations of existing methods, this next chapter will outline the systematic approach that will underpin the framework development for detecting and analysing extreme right imagery, ensuring it is both theoretically grounded and practically effective in addressing the challenges outlined in this chapter.

## Chapter 4: Framework Development

### Introduction

This chapter details each step that has occurred in the development process of creating the extreme right image analysis framework. Having used the literature to frame the importance and relevance of Baele's framework, this chapter will detail how the framework was developed through a three-part iterative design process. This process ultimately led to repurposing Baele's framework into a novel extreme right image analysis framework.

### Building the Framework

The framework developed by Baele et al. (2020a) has provided sufficient grounds to be repurposed in this study to analyse extreme right imagery. However, because this framework was created to capture the visual style of the Islamist Extremist group ISIS, modifications needed to be made to the original approach to repurpose the framework to analyse the

extreme right. Taking an existing theoretical framework and repurposing is a fruitful approach to conducting research. Whittaker (2022) adapted Gill's coding variables (Gill et al., 2017, Gill and Corner, 2015) to apply them to an empirical dataset of terrorist legal cases to better understand the phenomenon of online radicalisation. Therefore, this section will outline the three-part framework development process used to repurpose and redesign a new extreme right image analysis framework. In doing so, the first and final iteration of the framework will be presented.

To develop the framework, this thesis adopts a modified grounded theory approach across a three-part iterative process. Step 1 begins by identifying the overarching narratives presented within the extremist movement. Inspired by the original study, this stage starts by identifying the overarching crisis and solution of Islamist extremism. These 'master narratives' as they are referred to are then supplemented by several sub narratives which fall under the remit of the previously identified crisis and solution narratives. Identifying these narratives and sub-narratives will be drawn from findings in the terrorism and extremism literature. In doing so this first step prepares the researcher to draw connections and make distinctions between imagery and narratives used by the extreme right. Step 2 refers to the establishment of the main narratives and sub-narratives found in the extreme right movement, which is essential in developing a thorough and robust protocol for coding images. Following the same broad structure as Baele et al., this protocol is initially comprised of four primary categories. These are as follows: 1) master frames, 2) narrative connection, 3) shock value of an image, and 4) symbolic valence of an image. Step 3 refers to the iterative refinement of the framework. This process will use the information, and the structure developed in the first two steps and then expose the framework to the dataset through iterative analysis to refine the framework further. At the end of this three-part iterative process, the result will be an extreme right image analysis framework created from findings in the literature in combination with a dataset of extreme right images.

### **Step 1**

The first step centres on identifying the master crisis and solution frames (later referred to as master frames) from the existing literature, in addition to any specific narratives which fall within the remit of these master frames. When analysing these frames in the context of the

extreme right, this paper refers to the concept of an in-group, and an out-group, in conjunction with several additional narratives expressed in the Anders Breivik manifesto. As a result of the spectrum of extreme right individuals, groups, and collectives these master frames are subject to a high degree of variability depending on where in the extreme right spectrum these frames are drawn from. Therefore, it is unavoidable that the master frames and narratives will fall short of fully encompassing the extreme right collective as a whole. However, by drawing this understanding from Breivik's manifesto this thesis draws key concepts from an influential and recent extreme right document. For this purpose, this study looks to adopt the frames present in the manifesto of Anders Breivik. The decision to look at this document stems from several reasons, including that it is well noted in the literature that Breivik (despite being denounced by some extreme right groups/individuals) is notoriously recognised by the extreme right community (Cai and Landon, 2019), at to develop the framework from extremist data in addition to academic framings and understandings. Whilst this does risk skewing the framework to a European context, the benefits of drawing from extremist data are valued higher than the risk of potential bias.

This notoriety comes as a result of a terrorist attack carried out by Breivik in Utøya, Norway in 2011. This attack resulted in the death of 77 people in what has become one of the most well-known attacks carried out by the extreme right in recent years (BBC, 2022). Particularly as the attack was carried out by Breivik to great devastation in the absence of any accomplices or support (Cai and Landon, 2019). Before carrying out this attack Breivik published a 1,500-page manifesto composed of both his own words and articles by other right-wing authors. The content of Breivik's manifesto served to communicate the ideology which ultimately fuelled his motivation to carry out the attack. The lasting impact that the document has had in the years since its release is also a reason to use the manifesto to gather the master frames of the extreme right. Whereby, in this time Breivik's manifesto has been directly linked to at least five other extreme right terrorists, including most notably Brenton Tarrant (also referred to as the Christchurch shooter) (Ware, 2020; Cai and Landon, 2019). Tarrant carried out a terrorist attack motivated by his extreme right ideology in Christchurch, New Zealand in 2019, resulting in the death of 50 Muslims (BBC, 2020).

In his manifesto, Breivik outlines clear master frames of both an in-group and an out-group. As previously noted, these views do not capture the master frames consistent throughout the entire extreme right collective (Tony Blair Institute, 2019). Rather, these master frames offer a proposed definition of the largest point of conflict between the extreme right collectives, who is the in-group and who is this out-group. With the out-group or primary enemy of the extreme right ranging from Muslims, Jews, women, non-whites, or all these groups, for example. On the other hand, the inclusion of the in-group can range from race, culture, religion, gender, or all these factors.

Breivik understands the first master frame, the in-group, as the majority of Ethnic Europeans (white in complexion) who are actively adhering to or desiring to return to a society sharing conservative values. This in-group is disdained and persecuted by a small and powerful minority. Unlike the in-group, the second master frame, the out-group is separated into two collectives unified by their antagonistic attitude to the in-group. Firstly, there are Muslims who are perceived as violent and dehumanized barbarians. Secondly, there are Cultural Marxists who are perceived as intolerant and hysterical ultra-liberals who suppress the truths in pursuit of ideological ends. Where Muslims desire the violent death of the in-group, Cultural Marxists strive to end the in-group's ideals and values (Johnsen, 2014).

Continuing to draw from Breivik's manifesto, in combination with the two master frames, work carried out by the Tony Blair Institute (2019) and Johnsen (2014) identifies four key narratives which align with the broader master frames. These themes consist of the West vs. Islam, in-group victimisation, anti-establishment, the justification of violence, and major news events. These themes are not limited to Breivik's manifesto, rather, they form an ideological framework of mainstream and extreme ideas that can be applied to other collectives of the extreme right (Tony Blair Institute, 2019). Choosing to use master frames drawn by researchers from the Breivik manifesto is twofold. Firstly, these themes were aligned with the understanding of the master frames and secondly, the themes were sufficiently broad to potentially capture all themes present in the images whilst still providing scope to refine and build upon. The definitions of each of these four narrative themes found in Breivik's manifesto are covered below, as per the findings drawn from the 2019 study by the Tony Blair Institute. The West vs. Islam narrative is defined as 'the belief that the Christian West should unite

against Islam'. The in-group victimisation theme is understood as the 'belief that there is a global conspiracy to replace the white race'. The anti-establishment narrative's definition details 'advocating the complete overhaul of the political system and rejection'. The justification of violence narrative is defined as 'urging people to fight to defend their religion or culture with force'. The final narrative, major news events, refers to any large-scale, internationally pertinent event communicated by a news outlet.

## **Step 2**

The second step of the framework was to develop a thorough and robust protocol for coding images using the knowledge gained from Step 1. Akin to the structure of the framework as per Baele, et al.'s original framework, this framework is built around four key categories. The following paragraphs detail what specifically was encompassed in each of the four categories based on the research carried out, before initiating the first stages of framework development.

1) The in-group and out-group category refers to individuals represented within the image. The definition of the in-group and out-group can be found in step one. The type of individuals represented in the image can be coded categorically depending on whether individuals or messages portrayed in the image belong to the in-group or the out-group. The absence of an individual or message falling into the in-group, or the out-group (or instances where it is unclear to the coder) constitutes entering the 'other' category.

2) The narrative connection refers to whether any of the five narratives identified in step one is present within an image. These consist of the West vs. Islam, in-group victimisation, out-group demonisation, anti-establishment, justification of violence, and major news events. In cases where none of these narratives are present, images will be entered into a category labelled 'other'.

3) The shock value of the image can be better understood as whether the image does or does not depict gruesome content. The complexity of this category is that gruesomeness is a broad and inherently subjective term particularly when dealing with a topic as sensitive as extremism. The original framework adopted the definition of gruesomeness used by

Halfmann and Young (2010). This category and definition will remain in place for the framework development. Said definition will only consider content as being gruesome on the basis that it crosses one of three conditions. The first of these conditions is whether an image triggers an instinctive reaction of disgust by the coder, the second criterion is whether the image displays corpses, and the third refers to dismembered or beheaded bodies.

4) The symbolic valence of an image refers to the symbols present within the image. The symbolic valence of the image can be coded categorically based on the presence of specific symbols. Similarly, to the extreme right master frame and narrative, finding themes of symbols in extreme right imagery is complex. However, compiling the categories of images by the work of the Anti-Defamation League (2022), the Centre for Analysis of the Radical Right (Richardson and Wasser, 2020), and Farennet (2016), this study initially used five symbol categories. These five categories consist of common signs, runes, acronyms & abbreviations, extremist organisations, clothing brands & prints, and hate slogans.

Common signs include depictions of the swastika, SS, SA, black sun, eagle of the SS skull, Celtic Cross, and Pepe the frog. The runes category refers to letters from pre-Roman alphabets. Acronyms and abbreviations include combinations of letters and/or numbers that hide a hidden meaning. Extremist organisation images include flags, emblems and names depicting an extremist organisation. Clothing brands and prints include neo-Nazi clothing brands and clothes with racist slogans and signs printed on them. Hate slogans include those identified in the Anti-Defamation League Hate on Display hate symbol database (2022).

As a result of this step, an initial framework was developed. The format of this framework can be found below:

- **Master Frames:**
  - In-group
  - Out-Group
  - Other
- **Narratives:**
  - West vs. Islam

- In-group victimisation
- Anti-Establishment
- Justification of Violence
- Major News Event
- Other
- **Gruesomeness:**
  - Yes
  - No
- **Symbols:**
  - Common Signs
  - Runes
  - Acronyms/Abbreviations
  - Extremist Organisations
  - Slogans
  - Clothing Brands & Prints

### **Step 3**

Following the initial development of the framework, the next phase entailed three iterative rounds of framework development to tailor and refine the framework to the extreme right dataset. The role of the framework development was to evolve the framework from a research-driven codebook to a hybrid research and data-driven codebook. Each phase of the framework development increasingly subjected the framework to a large sample of the dataset.

Each round of the framework development incorporates an increasingly large sample from the dataset. Phase 1 of the framework development utilised 250 images from two extremist group Telegram groups totalling 500 images and two Telegram groups. The dataset for Phase 2 consisted of 250 images from three extremist Telegram groups. In the third and final phase of the framework development, a sample of 250 images was taken for each extremist group Telegram group, excluding cases where there are less than 250 images, then the entire Telegram group's number of images is used.

After being exposed to three rounds of the framework development the framework was substantially refined. With the addition of one new category and significant subcategory changes. The final codebook can be found in Appendix I with all the categories, subcategories, and definitions provided. A list of the final framework can be found below:

- **Master Frames:**
  - In-group Neutral/Positive
  - Out-group Negative
  - Emphasising Division
  - None,
  - Unknown
- **Narratives:**
  - Anti-identity
  - Anti-cultural Marxism
  - Conspiracy Theory
  - In-group Victimisation
  - Anti-establishment/Global Elite
  - Global Politics and Military
  - Call to Action
  - Activism
  - Humour
  - Other.
- **Shock Value:**
  - Yes
  - No.
- **Image Format:**
  - Single Image:
    - Undoctored Photograph
    - Cartoon/Meme
    - Doctored Photograph
    - Screenshot.

- Image & Text:
  - Undoctoring image
  - Graphic Design
  - Meme
  - Screenshot.
- **Symbolic Valence:**
  - Common Extremist Symbols
  - Acronyms/Abbreviations,
  - Extremist Organisations
  - Slogans
  - Weapon/Gore
  - Religious Symbols
  - Extremist Clothing
  - Key figures
  - News Outlet
  - Other

Whilst there were numerous changes made to the framework through category and sub-category names and definitions, it is necessary to draw attention to the image format category. The inclusion of this category and its subcategories were included to gather insight beyond the content of imagery into the formatting. The ambition for considering the format of imagery was to create more opportunities to uncover insights on extreme right posting behaviours to ultimately inform novel policy recommendations for online platforms to consider. The creation of this category stemmed from presenting the framework to academics and stakeholders and receiving feedback.

The approach to developing this framework has been rigorous; it is necessary to identify that it is not without limitations. The creation and utilisation of the dataset, for example, is a strength of the findings of this research. However, there are characteristics and limitations within this dataset that are necessary to identify. Whilst drawing from the manifesto of Anders Breivik is a strength of this framework, it cannot possibly give a holistic representation of the entire extreme right. As a European extremist, his views are less representative of a

U.S.-based extreme right figure. For example, identifying a quality of the in-group as “Ethnic Europeans”. However, providing a robust framework with identified limitations still provides significant utility, including opportunities for future refinement by online platforms and academe.

### Conclusion

To conclude, this chapter fulfils the explanation of how the extreme right image analysis framework was created for this thesis. This has been achieved by providing the means to begin a three-step framework development process of exploring the academic work which informed the initial design followed by three phases of framework development using a modified grounded theory. The final step was using the results to refine the framework, which in its final form can be found in Appendix I.

## Chapter 5: The Method

### Introduction

Following the establishment of the extreme right image analysis framework, the Method chapter of this thesis outlines the research design, data, and research strategy used. This thesis undertakes an image analysis of primary extreme right imagery extracted from the encrypted messaging platform Telegram. From 17 Telegram groups belonging to six different extremist groups. This research design and data are used in conjunction with the image analysis framework to develop an understanding of the visual style of the extreme right. Designing this framework to assess the feasibility of providing clarity and structure to the extreme right visual style is ultimately used to advise data-driven policy recommendations to help online platforms moderate against extreme right imagery content.

This chapter will be split up into several sections including, data and sample, measures, and analytic strategy. In doing so it will be made apparent the decision behind the data source and quality, dataset characteristics, and how these dataset traits relate to the research questions of the thesis.

### Data and Sample

The key characteristics of this dataset encompass six designated extreme right groups, 17 distinct Telegram groups, and a total of 25,547 images. By exploring the dataset more

thoroughly and identifying its key characteristics a clearer understanding will be provided. Consequently, enhancing transparency and further contextualising the subsequent results derived from the bespoke framework. Although the focus of this study is on extreme right images, the dataset's attributes allow for a broader scope of understanding, which will support and contextualise the results of the study more broadly. To further establish transparency, context, and insight into the broader findings, it is necessary to provide an overview of the dataset's characteristics.

To create a framework to analyse extreme right imagery, over 25,000 images were collected from the encrypted messaging platform Telegram. Images were collected from this platform due to its categorisation as a beacon platform. As previously state, this means that the data collected on Telegram is relevant to a range of online platforms dealing with extreme right imagery beyond Telegram. Consequently, in this research, Telegram functions as a data source for extreme right imagery populated across multiple platforms. This means that conclusions derived from this data are not solely relevant to Telegram but the boarder spectrum of content moderation practices of online platforms. Lastly, at the point of data collection, extremism research was less commonly carried out on Telegram when compared to other platforms like Gab, Facebook and X (X Company, 2024). As such, the dataset was collected to fill a gap in the literature and to use data relevant to multiple online platforms.

Telegram is a cloud-based encrypted messaging platform that allows users to communicate by sending other users text, voice, and video messages, sharing media and files, and making voice or video calls across devices. Communication is carried out through individual chats, group interactions, and broadcast-style messaging. Where more than two users are involved in communication Telegram has groups and channels. Telegram channels are designed for broadcasting messages to a large audience, with only admins able to post content, while subscribers can view posts, share them, and engage via comments (if enabled). Conversely, Telegram groups function as private or public spaces for discussion, allowing up to 200,000 members to interact via messages, media, and polls, with features like replies and reactions to foster conversation. Similarly to other online messaging platforms (like WhatsApp and Facebook Messenger), users can join several groups (conversations) with either an individual or a group, and users can use a search bar to find a group or a channel by searching its name

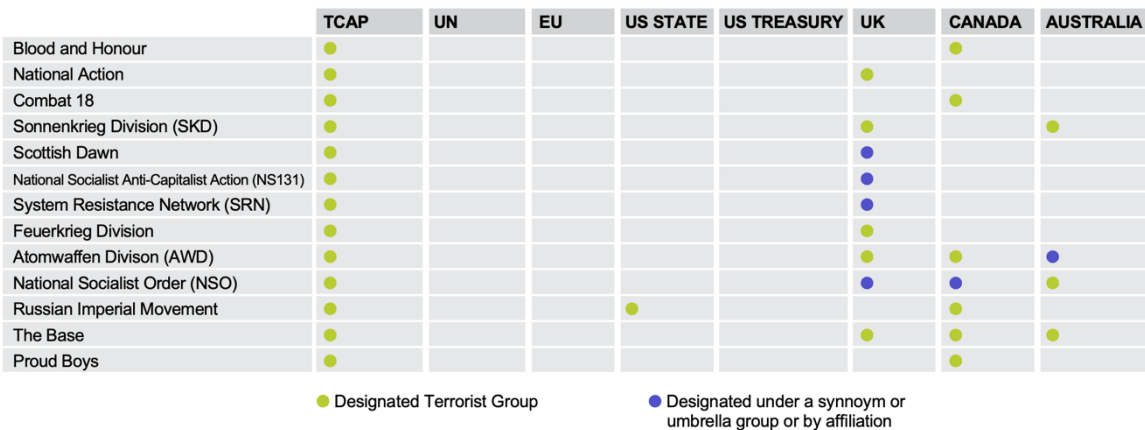
(in this case searches involved entering the name of an extreme right group into the search bar). For Telegram groups and channels, the respective admin/s for that chat have the option to approve every new joining member.

In this study, Telegram groups that did not require an admin to approve joining the group as a member were collected and used. The decision to collect data from Telegram groups was based on the desire to understand the imagery of the extreme right collective, including content created by extreme right groups and content created and repurposed by their followers and members. Collecting data from Telegram channels would have only allowed data created by the extreme right group to be collected and not by their followers, which would have been to the detriment of the aims of the study.

Having established where the data would be collected from the next step was to consider which extreme right groups data to collect. The priority when making this decision was to establish a dataset that could offer insight into the complexities and various fragments of collective that come under the extreme right covered in the literature review. This fragmentation refers to far-right and nationalist groups, accelerationist groups, alt-right groups and a sample of online organisations which have been known to represent the extreme right. By collecting a data sample which explores some of the fragments of the extreme right collective the dataset is diverse and representative of multiple extreme right collectives.

As a result of this project being in collaboration with TAT, deciding on group inclusion began with addressing TAT's TCAP group inclusion found in Figure 1 and Figure 2 below (Figure 1 shows the policy and time of collection and Figure 2 shows the policy at the time of submission). The groups included in TAT's TCAP are all designated by at least one nation in the global north. Using data from designated extreme right groups was decided on to strengthen the findings from the data analysis. By analysing the data of designated groups, the findings and recommendations are based on content that should be removed from online platforms due to their origin from extremist groups.

To find the data, the extreme right group names were searched for on Telegram and where official groups were present (or groups claiming to be official and not supporter groups), they were scraped for their images. This led to the collection of six groups (The Proud Boys, The



Base, National Socialist Order, Atomwaffen Division, Feuerkrieg Division, and Blood & Honour) and 17 distinct Telegram groups. While all of the designated groups in the TAT TCAP inclusion policy were searched for, only these six designated groups were identified. Of the groups found in Figure 1 many from TAT’s TCAP inclusion policy are missing. In these cases, the group names were searched for and either the groups did not have any presence on Telegram, or where Telegram groups were found, they were Telegram channels and not groups, or they were supporter-created and there were no claims of the group being an official outlet for the extremist organisation. From each of the Telegram groups, all the images posted on that group were scraped and stored for analysis. These data characteristics ultimately aim to best position this thesis in answering the four research questions

Through an infographic, Figure 1 contains the extreme right groups in the scope of TAT’s TCAP at the time of data collection, found in their Year 1 Transparency Report (TAT, 2022). Figure 2 does the same using the information provided in their Year 2 transparency report (TAT, 2023).

Figure 1: TAT's Designation of Extreme Right Entities Captured by the TCAP in 2021

	UN	EU	US State	US Treasury	UK	Canada	Australia	New Zealand
<b>Atomwaffen Division</b>					●	●		
<i>National Socialist Order</i>					●	●	●	
<b>Blood and Honour</b>						●		
<b>Combat 18</b>						●		
<b>Feuerkrieg Division</b>					●			
<b>National Action</b>					●			
<i>National Socialist Anti-Capitalist Action</i>					●			
<i>Scottish Dawn</i>					●			
<i>System Resistance Network</i>					●			
<b>Proud Boys</b>						●		●
<b>Russian Imperial Movement</b>			●	●			●	
<b>Sonnenkrieg Division</b>					●		●	
<b>The Base</b>					●	●	●	●
<b>James Mason</b>						●		

● Designated terrorist entity     
● Designated under a synonym or umbrella group or by affiliation     
● Content banned by the New Zealand Classification Office

Figure 1: TAT's Designation of Extreme Right Entities Captured by the TCAP in 2022

### Extreme Right Groups

As shown above, this dataset is comprised of six different designated extreme right groups. A short overview of these groups is provided below to give content to these groups and their positioning within the extreme right.

To explore the content of the dataset, it is necessary to frame the coding scheme used to characterise each unique extremist group's Telegram group. It is necessary to identify the meaning behind each of these codes as they are used to identify each of the Telegram groups individually through the following chapters. To create a code for each extremist group Telegram groups two pieces of information are necessary. Firstly, there is the name of the extremist group Telegram groups, as it appeared on Telegram at the point of collection. Secondly, there is the extremist group or organisation to which this Telegram group officially belongs. Once these two categories are identified, a unique extremist Telegram group code can be created. The relationship between extremist Telegram groups, extremist groups and unique code is found in Table 1 below. This process resulted in 17 unique codes being created, as will appear in the data visualisation in the following sections.

Table 1: Extreme Right Dataset's Groups, Telegram Groups, and Unique Telegram Groups Code.

<b>Extremist Group</b>	<b>Extremist Group Telegram Group Name</b>	<b>Extremist Group Telegram Groups Code</b>
Atomwaffen Division	☢️ AWD Chat ☢️	AWD
Blood & Honour	Blood and Honour Ukraine Chat	B&H1
Blood & Honour	Blood and Honour Ukraine	B&H2
Blood & Honour	Blood and Honour Hungary	B&H3
Feuerkrieg Division	Feuerkrieg Division <b>**Official**</b>	FKD
National Socialist Order	National Socialist Order	NSO
The Proud Boys	Proud Boys	PB1
The Proud Boys	Proud Boys Britannia	PB2
The Proud Boys	Official Proud Boys Ireland	PB3
The Proud Boys	Official Wisconsin Proud Boys	PB4
The Proud Boys	PROUD BOYS FROM ILLINOIS, BUT WE CANT SAY WHAT CHAPTER . LOL	PB5
The Proud Boys	Proud Boys Victoria	PB6
The Proud Boys	Seattle Proud Boys	PB7
The Proud Boys	Proud Boys New South Wales	PB8
The Base	The Base	TB1
The Base	Roman Wolf	TB2
The Base	The Base	TB3

The Proud Boys, founded in 2016 by Gavin McInnes, are a far-right group blending internet-based satire, trolling culture, and violent offline behaviour (Widdicombe, 2013). With a claimed membership of 3,000–8,000 (Coutts, 2017; Kenes, 2021; and SPLC, 2021), their recruitment exploits online algorithms to radicalise individuals while presenting themselves as a playful, men-only fraternity. Membership degrees range from reciting a creed to engaging in politically motivated violence. Despite disassociating himself in 2018 (Kriner and Lewis, 2021), McInnes' xenophobic and violent rhetoric heavily influences the group's identity, serving as a gateway for members to more extreme ideologies. The Proud Boys exemplify the intersection of alt-right culture, extremist narratives, and the evolving role of the internet in modern extremism.

The Base, a neo-Nazi and accelerationist group, was recently designated as a terrorist entity on the 26<sup>th</sup> of July 2024 by the European Union (European Union, 2024). The group was founded in the U.S. in 2018, and advocates for white supremacy, militant accelerationism, and the establishment of a white ethnostate achieved through the means of violence and

societal collapse (Renard and Rekawek, 2024). The group exists and is networked across numerous countries and has been involved in numerous violent plots, including targeting minorities, inciting terrorism, and spreading extremist propaganda globally via online platforms and has a growing presence in Europe.

Atomwaffen Division (AWD) was founded in 2015 by Brandon Russell as an accelerationist neo-Nazi organisation desiring societal collapse and the establishment of white ethno-states through acts of terrorism and violence (Ware, 2019). Its ideology is heavily inspired by James Mason's book *Siege*, which promotes leaderless, cell-structured terrorism (Johnson, 2023). AWD rejects political engagement, instead opting to destabilise governments and hasten societal collapse. AWD has been linked to numerous murders, hate crimes, and explosives offences in the U.S. (Fulgoni and Menis, 2024). The group inspired the creation of international chapters, including the Sonnenkrieg Division and was officially disbanded in 2020.

The National Socialist Order (NSO) emerged in 2020 as a rebranding of the Atomwaffen Division following its official disbandment in 2020 (Fulgoni and Menis, 2024). The group re-emerged through AWD's remaining members. However, its birth was faced with rapid decline by September 2022, when the NSO dissolved, with former members founding the National Socialist Resistance Front (Counter Extremism Project, 2022). The NSO's brief existence highlights the fragility and fluidity of extremist groups and networks, whose ideologies and internal conflicts often lead to fragmentation.

Feuerkrieg Division (FKD) was founded in late 2018 by a 13-year-old Estonian known as the "Commander". The neo-fascist, accelerationist group is inspired by *Siege Culture* and AWD despite rejecting direct affiliation with the group (Newhouse, 2021). Its influence expanded across Europe and North America, recruiting members through decentralised networks, propaganda, and online platforms. FKD members have been involved in numerous violent plots, including a failed bombing in Lithuania, plans for synagogue arson in the UK, and thwarted terrorist attacks in the U.S. (United States Attorney's Office, 2020). After the "Commander" was discovered by Estonian authorities in 2020, FKD's activity declined, with many members transitioning to other groups (BBC, 2020a).

Blood & Honour (B&H) is an extremist group originally founded in 1987 by Ian Stuart Donaldson from England's neo-Nazi skinhead music scene. The group sought to unite white supremacist groups through hate music and propaganda (Counter Extremism Project, 2024). The group organises up to 15 concerts annually in the UK, where it funds its activities through event fees and merchandise sales. Despite being banned in countries like Germany, Spain, and Russia, B&H continues to operate internationally, often evading authorities by announcing concert details mere hours in advance (Keatinge et al., 2019). Linked to the violent extremist group Combat 18 (C18) and U.S.-based sub-factions including the America Division and Council 28. Though its influence has waned since its inception, B&H remains present in the global white nationalist movement, leveraging music and media to recruit and radicalise followers.

### Measures

The Measures section outlines how key constructs, and variables of interest are used to analyse extreme right imagery collected from Telegram and their relevance to the research questions (RQ).

In this study, a bespoke coding framework is developed through an iterative, modified grounded theory approach. This framework combines deductive elements, informed by prior literature, and inductive insights derived from repeated engagement with the dataset. The coding categories and subcategories reflect the complexity of extreme right communication strategies and were refined through three rounds of iterative testing. The framework identifies the following primary variables for analysis:

**Master Frames:** The overarching communicative intent of the image is categorised into in-group neutral/positive, out-group negative, division emphasis, none, and unknown. This measure directly relates to RQ 1 and 3, as it identifies the communicative role of imagery within extreme right groups and helps determine whether a consistent visual style exists.

**Narratives:** Specific messages or themes conveyed by the image, such as anti-identity, anti-cultural Marxism, conspiracy theories, in-group victimisation, anti-establishment rhetoric,

global politics, and others. By analysing narratives, this measure addresses RQ1, RQ3, and RQ4 by identifying the themes that drive visual communication strategies.

**Shock Value:** The presence or absence of provocative or emotionally charged imagery that can elicit strong reactions. This measure is critical for RQ2, focusing on the use of graphic or provocative content. To ensure objectivity, shock value was cross-referenced with the “Weapons/Gore” subcategory within symbolic valence.

**Image Format:** The structural form of the image, including categories like single image versus image and text, and further subcategories such as undoctored photographs, memes, doctored images, and screenshots. Subcategories further specify the type of imagery, such as cartoons, doctored photographs, or memes. This measure provides insight into the stylistic choices employed by extreme right groups, contributing to understandings relevant to RQ3 and RQ4.

**Symbolic Valence:** The presence of identifiable symbols, including extremist slogans, acronyms, weapons, gore, religious symbols, clothing, and figures associated with extremist ideologies. This measure is pivotal for identifying both consistent visual styles (RQ3) and group-specific differences (RQ4). By quantifying the prevalence and type of symbols, this measure offers actionable insights for content moderation.

### [Analytic Strategy](#)

The process of analysing the extreme right imagery data initially began by extracting key themes from the literature to establish a collection of categories and subcategories as a starting point. This background knowledge formed a basis to carry out a modified grounded theory approach to forming the framework and analysing the data. The second step constituted three iterative rounds of image coding, whereby an increasingly growing sample of the image dataset was exposed to the framework to refine the presence and definitions of the categories and subcategories. From this process, a structure was formed for the framework comprising five categories and 35 subcategories that can be seen as follows:

- 1) **Master Frames:** in-group neutral/positive, out-group negative, emphasising division, none, and unknown.
- 2) **Narratives:** anti-identity, anti-cultural Marxism, conspiracy theory, in-group victimisation, anti-establishment/global elite, global politics and military, call to action, activism, humour, and others.
- 3) **Shock Value:** yes and no.
- 4) **Image Format:** single image and image (undoctored photograph, cartoon/meme, doctored photograph, and screenshot) and image and text (undoctored image, graphic design, meme, screenshot).
- 5) **Symbolic Valence:** common extremist symbols, acronyms/ abbreviations, extremist organisations, slogans, weapon/gore, religious symbols, extremist clothing, key figures, news outlet, and other.

The modified grounded theory approach is used to create these categories both inductive and deductive processes. By using previous research to inform the categories a deductive approach was taken. Additionally, by using iterative rounds of coding to refine the framework, an inductive approach was taken. The process of manually coding each image in the dataset followed a six-step process. These steps include Familiarisation with data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and writing up the findings.

After developing the framework through a combination of qualitative and quantitative analysis, a strictly deductive approach was employed to analyse the dataset of images. Each image was manually coded using the established framework, enabling a systematic evaluation of specific traits to address and explore the four research questions. This approach was essential because, while forming the framework relied on a level of interpretation necessary to understand the complexity of extreme right communication strategies. The analysis of images against the framework allowed the research to strictly measure and validate patterns quantitatively. By combining qualitative and quantitative methods, this mixed-methods approach enhances the accuracy, quality, and applicability of the findings to online platforms, where visual communication is particularly influential.

The four research questions were carefully designed to ensure the rigour, clarity, and relevance of the study, particularly in the context of extreme right communication on online platforms through imagery. Each question addresses a key dimension of visual communication in extreme right imagery.

RQ1: To what extent is imagery used as a medium of communication by extreme right groups? This question aims to establish the foundational role of imagery in extreme right communication strategies. Quantifying the prevalence of visual content in extreme right communication determines whether imagery serves as a primary mode of communication or a supplementary tool (knowledge useful to content moderators). This question will be answered by analysing the characteristics of the dataset, i.e. how many images are posted per extremist group and Telegram group, posting frequency and other key traits from the dataset.

RQ2: To what extent is graphic imagery used in the communicative strategies within extreme right imagery? In this RQ the focus shifts to the nature of the visual content, specifically exploring the use of graphic or provocative imagery. This question helps to identify whether extreme right groups rely on visually striking or emotionally charged content to capture attention, provoke reactions, and reinforce ideological messages. This knowledge can help moderators decide whether to prioritise looking for these traits when looking for extreme right imagery. This will be explored by analysing the presence of 'shocking images. The Shock Value category will be explored in combination with the 'weapons/gore' subcategory within the symbols category. This approach is taken as the shock value category is subjective due to it being determined based on the interpretation of a single coder. This necessitates the presence of the 'weapons/gore' subcategory within the symbols category to provide a non-subjective value.

RQ3: Is there an extreme right visual style that is communicated and reinforced through the use of imagery? This question investigates whether a consistent, identifiable visual style exists across extreme right groups. Understanding this helps to determine how a shared "visual style" may contribute to group identity, coherence, and recognizability within online platforms. Understanding this visual style can directly feed new knowledge to content

moderators for targeted content removal. This is carried out by holistically considering the framework and dataset as a whole to gather a big-picture understanding.

RQ4: Do extreme right groups have differing visual styles that are communicated and reinforced through the use of imagery? While the previous question examines a unified style, this question explores the diversity of visual communication strategies among different extreme right groups. By identifying group-specific styles, this analysis highlights how visual content might reflect differing ideologies, goals, or target audiences. Discoveries related to this question can promote the removal of specific groups whilst preserving content that falls outside of its remit, preventing inaccurate and unintended content removal. This question necessitates a similar approach to the previous whilst expanding on the differences and variability across groups and within extremist groups that have multiple Telegram groups.

These questions form a comprehensive framework to analyse extreme right imagery, ensuring that both the commonalities and differences in their visual strategies are explored. This approach not only characterises the role of imagery within extreme right communication but also provides insights into how such content might be understood in a way that develops targeted content moderation practices.

## Conclusion

To conclude, this chapter fulfils the methodological and data contributions of the thesis. The purpose of providing these contributions in this chapter has been to capture the most pronounced characteristics of extreme right imagery. In doing so, it has been determined what data was collected (the six extreme right groups) and from which location/platform (the Telegram messaging platform). Then by exploring the key characteristics of said dataset, these were positioned against the research questions, showing the relationship between the design of the study and the aims of the study.

## Chapter 6: Results

### Introduction

The extreme right image analysis results are presented in this chapter. The bespoke methodology, detailed in Chapter 4, was designed to provide insights into the visual style, and

the visual communication strategies employed by various extreme right groups. The results come from applying the framework to the aforementioned dataset sourced from multiple extremist communication groups found on the encrypted messaging platform Telegram. Thereby, this chapter will assume the structure of first outlining several key statistical qualities of the dataset before moving to the quantitative results drawn from applying the methodology to said dataset.

Having covered the traits of the dataset, including its strengths and limitations, the results of the analysis will now be addressed. The results of the study will be covered by first considering the d=framework development before using the same format of the framework itself. The framework consists of five key components, Master Frames, Narratives, Shock Value, Image Format, and Symbols. Each of these categories and their results will be presented in this section of the chapter. Both the Narratives and Symbols categories allowed for more than one subcategory to be selected subcategories to be chosen (unlike the other categories) and as a result, the percentage totals at the bottom of the categories' table do not add up to 100%. Due to multiple selections being made, the total figure exceeds 100%.

### Dataset Characteristics

Before exploring the results stemming from the dataset being applied to the framework, it is first necessary to outline dataset characteristics that have an impact on the study results. These dataset characteristics have been characterised into three groupings, the number of unique images, the duration of a Telegram group's activity, and the average posting frequency based on the previous two metrics. To best comprehend these dataset characteristics, each will be examined independently in this section.

The first dataset characteristic to be identified is the dataset's number of unique images. As previously stated, this dataset comprises a total of 25,547 images. Figure 3 visualises the total number of unique images per extremist group Telegram group.

From this figure, there are several points to note. Firstly, there is significant variability in image count between groups. The NSO group Telegram group for example only accounts for just 3 unique images, whereas the PB2 Telegram group has 10,654 unique images. Although

this provides useful insight into extremist group behaviours it is a characteristic of the dataset necessary to highlight.

Furthermore, there is the noteworthy presence of PB images more broadly throughout the dataset. One Telegram group alone accounts for 41.7% of the dataset (PB2). The significance of this contribution to the total dataset is only emphasised when all the PB Telegram groups are combined. The result of this combination is a total of 22,656 images, which equates to 88.7% of the dataset's total images being accounted for by a single extremist group made up of 8 unique Telegram groups.

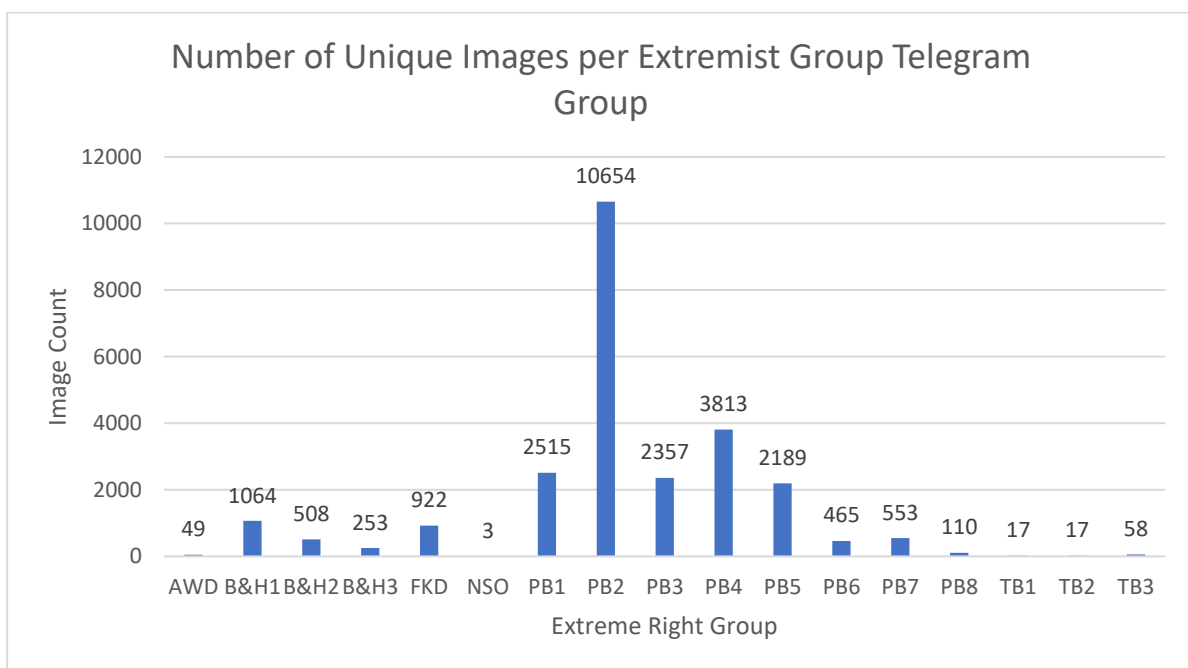
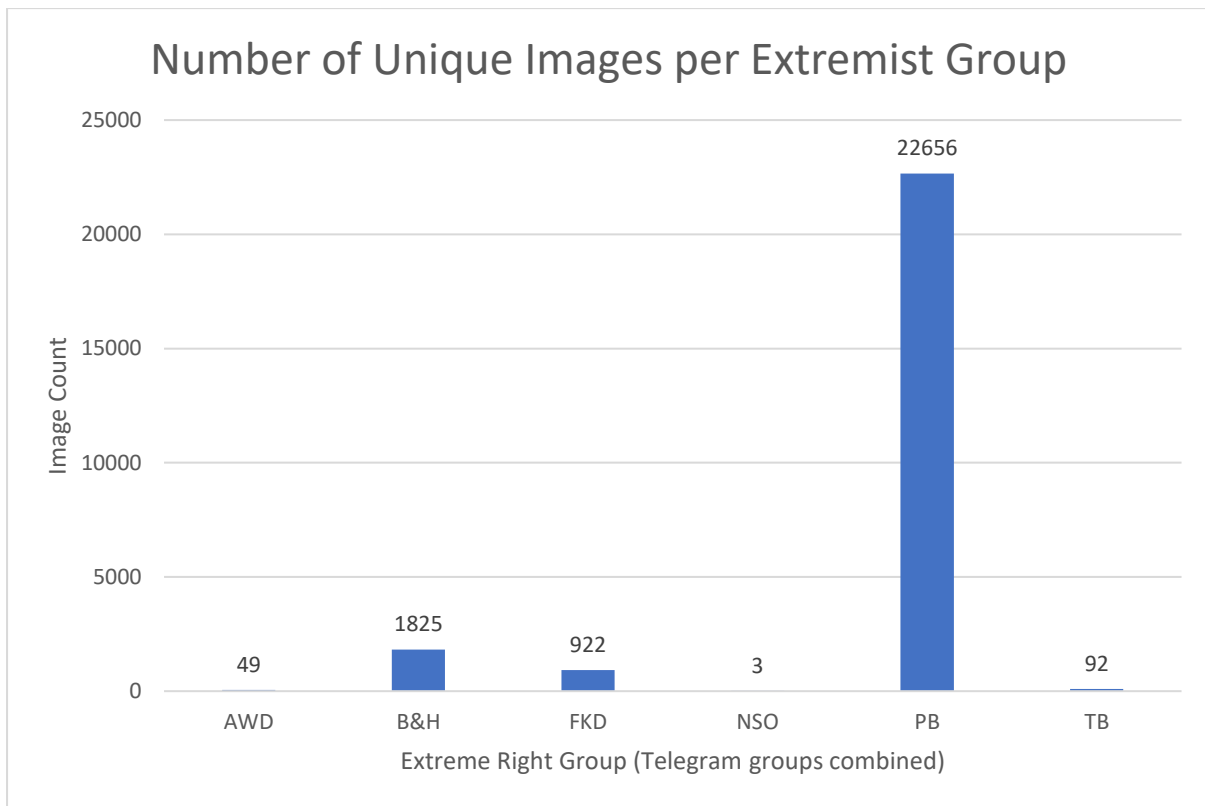


Figure 2: The Number of Unique Images per Extremist Right Group Telegram Group

This leads to Figure 4, which presents the same data of total image count as seen in the previous figure, however, in the cases where a single extremist group has multiple Telegram groups, the image count of those Telegram groups has been combined (for example in the case of The Base, TB1, TB2, and TB3 are condensed to become TB). This gives a total figure of images posted per group as opposed to per Telegram group.

This visualisation reflects a stark difference in the use of images between extremist groups. The PB group showcases an intentional use of images as a means of communication. Whereas, in the cases of all other groups, there are signs of image use but to a far lesser extent. This,

therefore, has a significant impact on the results and findings of this study, as the results are affected by the significant representation of PB imagery in the dataset.



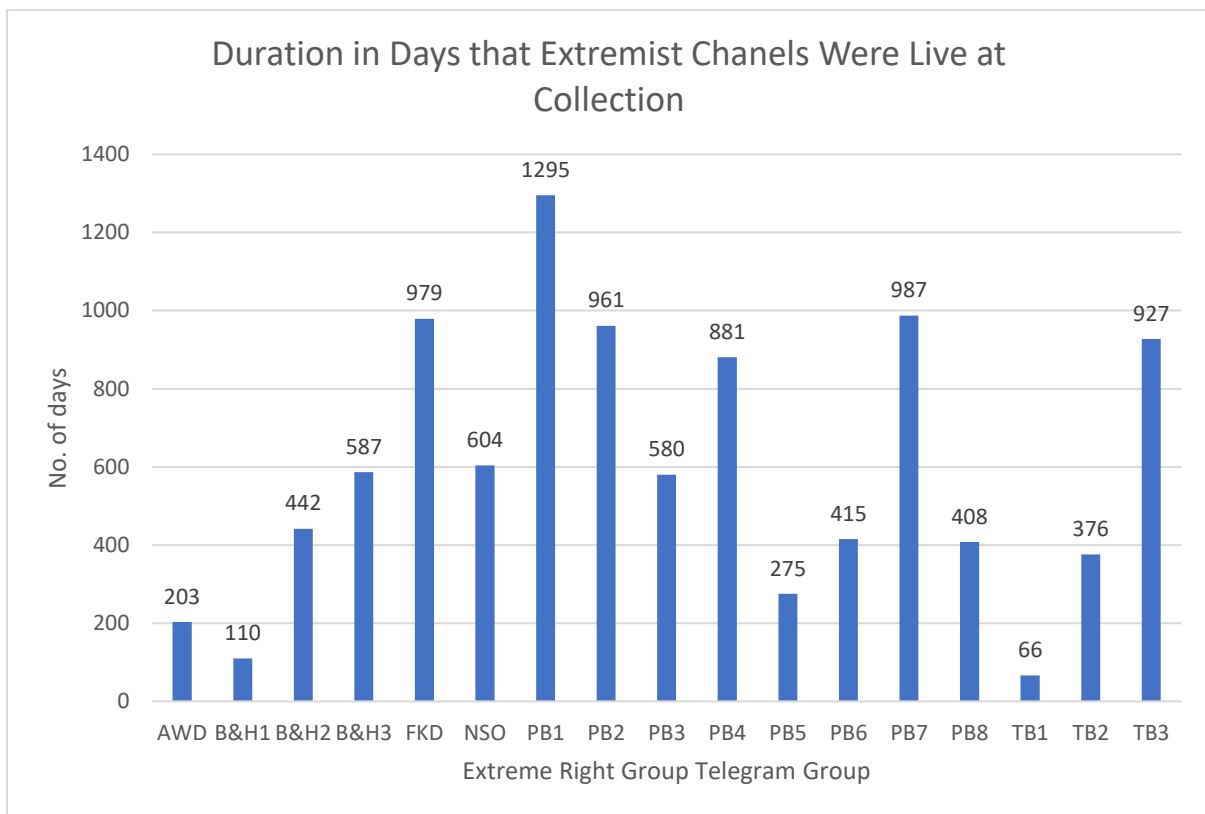
*Figure 3: The Number of Images Posted by the Combined Telegram groups of Each Extreme Right Group*

### Duration

The previous figures showcased many images posted by extremist groups on their respective Telegram group and revealed significant dataset traits as a result. Exploring the duration each Telegram group was live at the point of collection offers further insight into these results. Said results are visualised in Figure 5 found below.

The data used to form this figure was curated by calculating the difference between the date that the data from the group was scraped from Telegram, and the date the group was created. Clarifying the duration of each Telegram group as this trait could be a contributing factor to the disproportionate representation of images from the PB Telegram group in the dataset. However, Figure 5 dispels the notion that the number of images per Telegram group can be explained and correlated to the number of days a given Telegram group has been active. In the cases where the total image count is highest (i.e., Telegram groups belonging to PB, B&H,

and FKD) this explanation would show that the length of time a Telegram group has been active would also be high. However, this is not universally true. This logic can also be applied to the cases where the Telegram group has the lowest image counts (i.e., Telegram groups belonging to AWD, NSO, TB). None of these groups had the group with the smallest live duration.



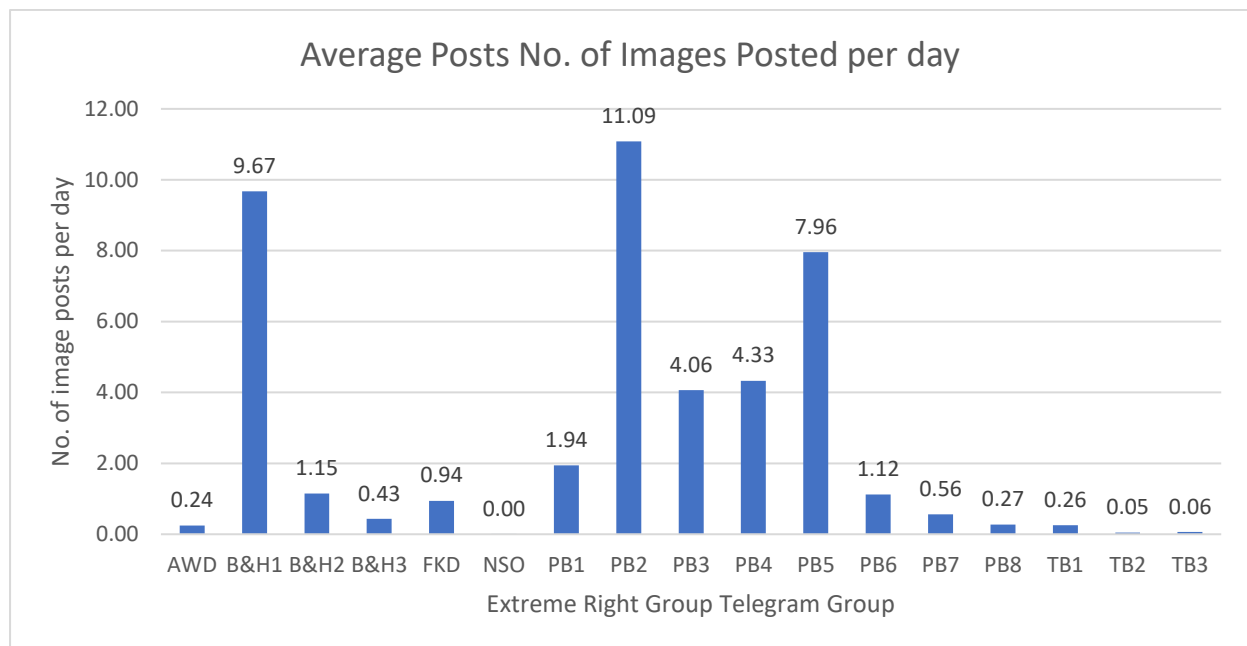
*Figure 4: The Number of Days Each Extremist Group Telegram Groups Was Live at Data Collection.*

### Frequency

Figure 4 further reinforces the results found in Figure 5 that explore the relationship between the number of posts per Telegram group and the number of days that the group has been live.

In Figure 6, three main points are communicated. Firstly, there is the variability in the average number of posts per day. This leads to the second point which is that even with an extremist organisation, Telegram groups showcase significant variability in image posting frequency. Finally, the ratio between images posted and days the Telegram group has been live presents

notable variability between extremist groups and Telegram groups. These results not only reinforce the notion of variability in posting behaviours between extremist groups but also between extremist group Telegram groups.



*Figure 5: The Average Frequency of Posts Per Day Per Telegram Group.*

Figure 7 consolidates the results of the dataset into a single illustration. Combining, the image count, duration of Telegram group activity and average number of daily posts. In doing so it is apparent that the PB2 Telegram group accounts for most of the dataset. Furthermore, the PB group accounts for a significant portion of the dataset. It is also clear that the duration of each Telegram group’s activity does not explain this disproportionate representation. This provides essential context to be considered before moving on to developments of this bespoke framework, which characterises extreme right images.

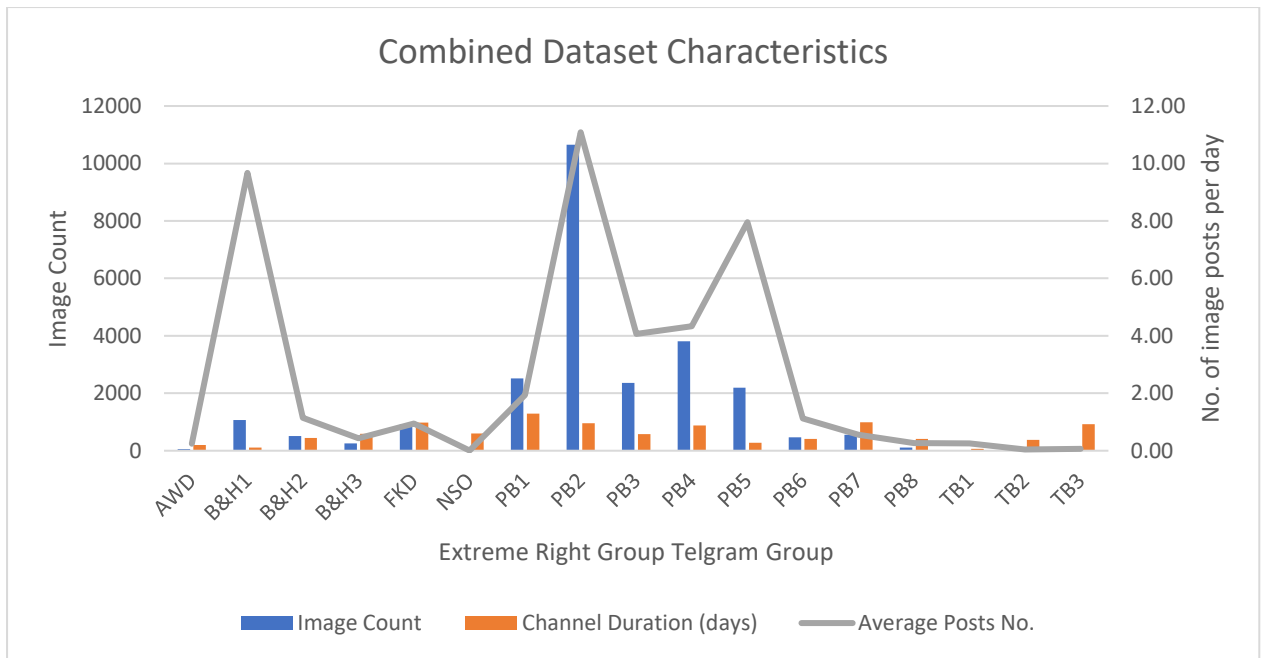


Figure 6: The Number of Images, the Duration a Telegram Group was Live, and the Frequency of Posts Per Day.

### Master Frames

The first component of the framework identifies the ‘Master Frames’. This section is made up of five subcategories. The results from this section of the framework can be seen below in Table 2, where a percentage value can be found for each of the ‘Master Frames’ components from each extremist Telegram group, in addition to the average result of all Telegram groups.

When calculating the average percentage of Master Frames for the entire dataset, there are three key points to be made from the results. Firstly, the total percentage of the ‘Unknown’ category across the dataset is at 3.1%, meaning that the use of images to target the in or the out-group is clear and easily identified. Secondly, with almost 50% of the results being marked as ‘In-group Positive’, there is a clear focus on reinforcing the in-group instead of demonising the out-group. Thirdly, over 10% of the results do not discuss the in-group, out-group, or emphasising division, meaning that over 1 in 10 images require further information to be effectively categorised by this framework to interpret their intention.

Table 2: Master Frame Results (%)

Group Name	Master Frames				
	In-Group Positive	Out-Group Negative	Emphasising Division	None	Unknown
AWD	34.69	14.29	24.49	16.33	10.20
B&H1	78.01	7.33	8.93	3.85	1.88
B&H2	50.98	29.72	12.60	4.33	2.36
B&H3	69.57	15.02	12.65	1.19	1.58
FKD	69.31	16.27	1.84	8.79	3.80
NSO	100.00	0.00	0.00	0.00	0.00
PB1	46.88	34.75	9.62	5.57	3.18
PB2	56.87	24.83	5.78	8.00	4.52
PB3	24.31	51.59	9.50	12.98	1.61
PB4	35.09	40.26	7.13	10.75	6.77
PB5	36.73	40.43	4.98	15.44	2.42
PB6	36.77	38.92	4.52	18.28	1.51
PB7	33.27	41.41	2.17	11.93	11.21
PB8	30.91	37.27	5.45	26.36	0.00
TB1	100.00	0.00	0.00	0.00	0.00
TB2	11.76	17.65	5.88	64.71	0.00
TB3	94.83	1.72	0.00	1.72	1.72
<b>Totals</b>	<b>53.53</b>	<b>24.20</b>	<b>6.80</b>	<b>12.37</b>	<b>3.10</b>

Table 2: This table presents the five subcategories within the Master Frame category and provides the percentage distribution of extremist right groups' Telegram group for each subcategory. The table also includes a row at the bottom that displays the combined percentage for all extremist groups and Telegram groups.

### Narratives

Following the Master Frames category is the 'Narratives' section of the framework. This section is made up of 10 subcategories which capture the story presented within each image. The results from this section of the framework can be seen below in Table 3, where a percentage value can be found for each of the 10 Narrative components from each extremist Telegram group, in addition to the average result of all Telegram groups.

From this figure there is scope for significant interpretation, however, there are three main narratives to be drawn from the average result of all Telegram groups. Firstly, activism is a more present narrative than the call to action narrative. Humour also has a significant

presence in the dataset. Lastly, over 10% of the dataset fulfils the *other* category. Therefore, there may be additional categories outside of this framework which may be added in future iterations of this study.

*Table 3: Narratives Results (%)*

Group Name	Narratives									
	Anti-Identity	Anti-Cultural Marxism	Conspiracy Theory	In-group Victimisation	Anti-Establishment/Global elite	Global Politics and Military	Call to action	Activism	Humour	Other
AWD	22.45	12.24	0.00	2.04	2.04	2.04	0.00	53.06	2.04	22.45
B&H1	0.00	0.00	0.00	0.00	5.92	10.43	6.95	64.66	1.60	10.53
B&H2	5.91	15.55	0.20	0.79	5.71	9.45	9.84	46.46	1.77	18.70
B&H3	0.79	1.98	0.79	0.79	2.77	7.51	15.02	43.08	1.19	20.16
FKD	15.29	2.39	2.71	2.60	4.01	3.04	10.63	67.25	8.13	10.74
NSO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
PB1	3.54	5.61	2.62	15.98	18.41	14.55	3.78	20.83	32.52	12.88
PB2	3.19	6.91	0.76	8.84	9.35	6.98	14.86	44.58	12.17	11.34
PB3	1.65	18.71	15.95	9.59	31.44	17.73	3.39	8.57	62.79	6.28
PB4	0.79	33.36	3.23	11.83	7.32	10.78	4.20	20.95	34.96	13.03
PB5	0.23	13.57	8.22	13.84	33.30	14.80	5.39	27.91	39.47	7.81
PB6	3.87	11.83	1.29	4.52	18.28	10.54	1.08	37.20	23.87	0.00
PB7	1.27	15.55	2.35	12.66	15.91	3.80	1.45	49.01	14.47	22.06
PB8	1.82	12.73	3.64	10.00	18.18	10.91	9.09	14.55	44.55	13.64
TB1	0.00	0.00	23.53	0.00	0.00	41.18	35.29	100.00	0.00	0.00
TB2	0.00	0.00	0.00	17.65	17.65	5.88	0.00	5.88	0.00	70.59
TB3	0.00	0.00	25.86	3.45	0.00	34.48	46.55	51.72	1.72	0.00
<b>Totals</b>	<b>3.58</b>	<b>8.85</b>	<b>5.36</b>	<b>6.74</b>	<b>11.19</b>	<b>12.01</b>	<b>9.85</b>	<b>44.45</b>	<b>16.54</b>	<b>14.13</b>

*Table 3: This table presents the ten subcategories within the Narratives category and provides the percentage distribution of the extreme right Telegram group for each subcategory. The table also includes a row at the bottom that displays the combined percentage for all extremist groups and Telegram groups.*

### Shock Value

The results of the 'Shock Value' category can be found in Table 4 using the same format seen in previous figures. Here, a percentage value can be found for each of the two Shock Value subcategories (Yes and No) from each extremist Telegram group, in addition to the average result of all Telegram groups.

Although there is notable cross-group variability, there is one key finding from this category. This is because the visual style of the extreme right does not place an emphasis on shocking the viewers of its images. With almost 98% of the dataset not being composed of shocking imagery, this finding is robust in its claim.

*Table 4: Shock Value Results (%)*

Group Name	Shock Value	
	Yes	No
AWD	2.04	97.96
B&H1	7.61	92.39
B&H2	10.83	89.17
B&H3	7.11	92.89
FKD	6.18	93.82
NSO	0.00	100.00
PB1	0.08	99.92
PB2	0.22	99.78
PB3	0.00	100.00
PB4	0.29	99.71
PB5	0.00	100.00
PB6	0.00	100.00
PB7	0.54	99.46
PB8	0.00	100.00
TB1	0.00	100.00
TB2	0.00	100.00
TB3	0.00	100.00
<b>Totals</b>	<b>2.05</b>	<b>97.95</b>

*Table 4: This table presents the two subcategories within the Shock Value category and provides the percentage distribution of extremist right groups' Telegram group for each subcategory. The table also includes a row at the bottom that displays the combined percentage for all extremist groups and Telegram groups.*

#### Image Format

The results stemming from the 'Image Format' category can be found in Table 5 below, whereby a percentage value can be found for each of the five Master Frame components from each extremist Telegram group in addition to the average result of all Telegram groups.

From these results, the total percentages offer three key points. Firstly, undoctored photographs were the most common subcategory. This points toward a desire by the extreme right to use simple and possibly low-quality imagery as opposed to high-quality edited imagery. This is reinforced by unedited screenshots accounting for over 16% of the dataset, pushing a narrative of low-quality, easily shared imagery being favoured by the extreme right. Finally, the two meme categories in the ‘Single Image’ and ‘Image & Text’ categories account for over 22% of the dataset, which points to a hybrid approach of low-quality simple imagery in combination with higher-quality edited imagery.

Table 5: Image Format Results (%)

Image Format								
Group Name	Single Image				Image & Text			
	Undoctored photograph	Cartoon/Meme	Doctored photograph	Screenshot	Doctored Image	Graphic Design	Meme	Screenshot
AWD	34.69	0.00	6.12	0.00	2.04	32.65	24.49	0.00
B&H1	26.41	0.85	21.80	6.30	23.50	0.47	1.79	18.70
B&H2	41.73	2.36	0.00	5.31	5.91	35.24	3.94	5.51
B&H3	29.64	0.00	13.04	0.00	56.13	1.19	0.00	0.00
FKD	45.88	3.25	4.23	19.09	3.04	18.55	5.75	0.22
NSO	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00
PB1	17.50	12.01	5.65	29.07	3.10	19.88	12.49	0.32
PB2	30.39	3.74	5.01	13.59	3.94	18.98	23.72	0.54
PB3	8.91	16.21	6.28	21.43	6.62	23.59	16.89	0.08
PB4	26.83	4.14	4.38	30.32	1.99	12.64	19.33	0.37
PB5	27.64	0.55	16.99	18.36	2.19	15.40	18.64	0.23
PB6	20.00	2.15	2.37	18.06	2.58	24.95	29.89	0.00
PB7	15.73	5.97	3.80	44.48	2.89	15.01	11.93	0.18
PB8	8.18	3.64	0.00	19.09	0.91	19.09	49.09	1.82
TB1	0.00	0.00	0.00	0.00	82.35	17.65	0.00	0.00
TB2	52.94	0.00	0.00	47.06	0.00	0.00	0.00	0.00
TB3	44.83	1.72	1.72	3.45	24.14	24.14	0.00	0.00
<b>Totals</b>	<b>25.37</b>	<b>9.21</b>	<b>5.38</b>	<b>16.21</b>	<b>13.02</b>	<b>16.44</b>	<b>12.82</b>	<b>1.65</b>

Table 5: This table presents the two subcategories (each with four of their own categories) within the Image Format category and provides the percentage distribution of extremist right groups' Telegram groups for each subcategory. The table also includes a row at the bottom that displays the combined percentage for all extremist groups and Telegram groups.

### Symbols

The fifth and final category of the framework is the ‘Symbols’ category. The results from this section of the framework can be seen below in Table 6, where a percentage value can be

found for each of the five Master Frame components from each extremist Telegram group, in addition to the average result of all Telegram groups.

The Symbols category has yielded a notable array of results, however, there are again three key findings. Firstly, over 34% of the dataset presents no definitive symbols. Therefore, this category has been useful in uncovering that symbols alone are not a substantial enough tool to be used in identifying extreme right content. However, there are symbols favoured by the extreme right, including the presence of weapons/ gore which accounted for almost 20% of the dataset. Extremist organisations (via logos and text) are the second most prominent feature in this category by being present in under 30% of extremist images. This leaves common extremist symbols and the most present symbols in extreme right imagery at over 30% prevalence. This in turn leaves significant scope for content removal tools to be applied using this information of extreme right symbols.

Table 6: Symbol Results (%)

Group Name	Symbols									
	Common Extremist Symbols	Acronyms/ Abbreviations	Extremist Organisations	Slogans	Weapons / Gore	Religious Symbol	Clothing Brands and Print	Key Figures	News Outlet	No Definitive Symbols
AWD	38.78	0.00	6.12	2.04	44.90	10.20	0.00	6.12	0.00	26.53
B&H1	37.78	0.66	9.96	0.75	39.66	0.47	3.85	2.16	0.38	25.56
B&H2	36.81	4.72	24.21	1.18	7.09	1.57	1.77	5.91	3.35	48.43
B&H3	30.83	6.72	15.81	0.00	7.51	0.40	3.95	16.21	0.00	45.45
FKD	53.58	2.71	46.85	37.20	42.62	4.01	0.65	17.46	4.34	18.55
NSO	100.00	0.00	33.33	0.00	33.33	0.00	0.00	0.00	0.00	0.00
PB1	1.59	2.43	15.23	1.19	7.91	0.20	15.55	37.97	5.69	30.97
PB2	1.20	2.82	35.72	3.60	12.43	2.04	13.15	17.73	3.19	36.59
PB3	2.33	4.62	11.46	1.19	1.23	0.55	3.61	33.56	11.84	46.25
PB4	3.59	0.58	11.96	1.26	5.95	0.47	7.58	19.15	5.59	51.90
PB5	1.69	0.37	6.58	1.32	9.50	0.82	1.42	17.27	4.16	65.92
PB6	20.86	11.40	28.17	3.87	4.30	0.43	4.09	22.58	2.58	37.85
PB7	14.29	5.97	28.57	3.44	4.52	0.36	8.14	47.74	10.13	20.25
PB8	11.82	1.82	17.27	0.00	7.27	0.00	1.82	12.73	3.64	56.36
TB1	100.00	0.00	100.00	0.00	41.18	0.00	0.00	0.00	0.00	0.00
TB2	0.00	0.00	11.76	0.00	11.76	0.00	0.00	5.88	0.00	70.59
TB3	84.48	0.00	87.93	51.72	46.55	0.00	0.00	1.72	0.00	3.45
<b>Totals</b>	<b>31.74</b>	<b>2.64</b>	<b>28.88</b>	<b>6.40</b>	<b>19.28</b>	<b>1.27</b>	<b>3.86</b>	<b>15.54</b>	<b>3.23</b>	<b>34.39</b>

Table 6: This table presents the ten subcategories within the Symbols category and provides the percentage distribution of extremist right groups' Telegram group for each subcategory. The table also includes a row at the bottom that displays the combined percentage for all extremist groups and Telegram groups.

Having established the results of the analysis achieved by this study, there is a significant opportunity to discuss these results and their implications in the following Findings chapter. The results reveal narratives, behaviours, and key findings of the variability of the extreme right. These include distinct extreme right priorities such as referring positively to the in-group, using humour, and avoiding shocking imagery. This reinforces the applicability of the framework in capturing the visual style of the extreme right across several groups. These results will supplement the statements and findings that will be made in the following chapter of this research.

## Conclusion

These results have been extracted through the application of a bespoke extreme right image analysis framework to a novel dataset of extreme right imagery. From this analysis, valuable insights are offered into the communication strategies utilised by extreme right groups on Telegram and online communication platforms more broadly. These results and insights have been formulated not only through the conception of the categories and subcategories of the bespoke framework but also through quantifying and exploring the qualities of the dataset itself. In examining the quantitative metrics of the core results, this thesis provides a comprehensive understanding of the dataset characteristics, and the visual tactics utilised by different extremist groups. These findings lay the groundwork for further discussion and analysis in subsequent chapters, offering novel findings on the nuanced tactics of the extreme right's use of images.

## Chapter 7: Discussion

### Introduction

Having established the results of the study, this chapter will follow the format of answering each research question, referring to the tables and charts found in the results section to evidence any arguments made.

Four research questions were used to form the basis of this study. Thereby the discussion looks to directly link the results and findings of this research to these questions. To re-establish the research questions being referred to, they can be seen as follows:

1. To what extent is imagery used as a medium of communication by extreme right groups?
2. To what extent is graphic imagery used in the communicative strategies within extreme right imagery?
3. Is there an extreme right visual style that is communicated and reinforced through the use of imagery?
4. Do extreme right groups have differing visual styles that are communicated and reinforced through the use of imagery?

Each research question constitutes one of the four sections, which ultimately form the structure of the Discussion chapter, starting with RQ1 through to RQ4. This is followed by policy recommendations aimed centrally at online platforms whilst also applying to related stakeholders (Governments, Non-Governmental Organisations and Researchers) to help implement a holistic approach to helping platforms more effectively respond to extreme right imagery. The answers to these four research questions will be based on the data drawn from the dataset and the extreme right image framework. While this creates a useful insight into prominent extreme right groups and collectives on the platform Telegram, which is well-established as being the central platform of use by the extreme right (Crawford, 2020), answers are subject to this dataset alone. Therefore, while this discussion asserts that the extreme right exhibits certain characteristics as indicated by the data analysis, the limitations of the data must be acknowledged. However, it is the opinion of this thesis that the dataset and its findings do give insight into the wider extreme right because of this inclusion of a range of group types. These insights are strengthened by findings in the wider literature, as discussed in Chapters 2 and 3.

RQ1: To what extent is imagery used as a medium of communication by extreme right groups?

The first research question looks at the extent to which extreme right groups utilise imagery. In order to approach this question, there are two avenues to explore. The first is looking at the total number of posts per group across the entire dataset to get a sense of how widespread imagery is used between extreme right groups. The second is analysing the average number of posts per group per day to explore how prolific specific extreme right groups can be in the context of posting behaviour and frequency. This first research question is directed toward the behaviour of extreme right groups (in the form of image posting) and not at specific image characteristics. Consequently, the answer to this question can be found in the characteristics of the dataset, whereas the answers to future questions will be found in the results of the framework.

In order to address this research question, a pre-emptive question must be asked. If the question of how widely imagery is utilised by the extreme right is to be posed, it is first necessary to ask whether imagery is used at all. As shown in Figures 3 and 4 in Chapter 5 (The

Method), evidence is clear that images are used by groups in this dataset to varying extents, and the literature review shows that this is a theme evident among other extreme right groups.

Exploring the preliminary question of whether imagery is used by the extreme right, it is necessary to consult Figure 3. From this figure, the data shows that the extreme right utilises imagery. All six of the dataset's extremist groups and each of the 17 Telegram groups have used imagery to varying degrees. At least three images were found in a single Telegram group (as seen in the NSO Telegram group), and the total number of images posted in a Telegram group ranges as high as 10,654 images (as seen in the PB2 Telegram group). Therefore, the data finds that the extreme right does use images in their digital communication.

This leads to the initial question: to what extent are extreme right groups utilising imagery? Referring to Figure 3, it is made clear that the answer varies. It varies not only from one extremist group to another but from one Telegram group to another, even in the cases where those Telegram groups belong to the same extreme right parent organisation. When looking at extremist groups, there are only three that posted over 500 images. These are B&H, FKD, and the PB. From this finding, some use of imagery is consistent with extreme right behaviour online, but this behaviour varies to a significant extent. From this data, the use of imagery is present within extreme right behaviour online. However, it would be a misrepresentation to claim that image use is widespread amongst all extreme right groups. Rather, an adapted and more accurate claim is that specific extreme right groups appear to favour imagery as a means of communication in comparison to others. This encourages a more detailed and nuanced perspective on claims made about extreme right imagery, which can often be very general, such as various types of imagery being a widely adopted communication tool for the extreme right (ISD, 2024). This can particularly be seen in the case of B&H, FKD, and PB.

In the cases of FKD (922 images), the results can be taken to represent the group's imagery posting as a whole as these are spread over their one and only Telegram group. Whereas, in the cases of B&H (images 1825) and The Proud Boys (22656 images), these results can be explored further by analysing internal group differences in posting behaviours, i.e. the differences between imagery posted in multiple Telegram groups belonging to a single

extremist group. Looking first at the FKD, 922 images found on a single Telegram group indicate that there is an intentional use and preference toward utilising imagery. This is indicated by the fact that there is only one Telegram group, which has a relatively medium-large concentration of images in comparison to other groups in the dataset.

Moving to the groups with multiple Telegram groups, the B&H group comprises 1825 images spread over three distinct Telegram groups. From these three Telegram groups, there is substantial variability in the use of images. B&H1 has 1064, B&H2 has 508 and B&H3 has 253. This suggests that there is a behavioural and stylistic communication difference within groups, between Telegram groups. Finally, the Proud Boys accumulated 22,656 images, spread over eight distinct Telegram groups. The results from these Telegram groups can be separated into three distinct categories. Category 1 is around 500 images (in the cases of PB6, PB7, and PB8), Category 2 are the Telegram groups, which total around 2,500 images (in the cases of PB1, PB3, PB4, and PB5), and Category 3, 10,000 images (in the case of PB2). Therefore, this group proves to be a distinct outlier when compared to the rest of the dataset in the high overall image and Telegram group count. However, the variability between all the Telegram groups is consistent with what is found in the context of B&H. The data communicates that where an extreme right group utilises imagery over multiple Telegram groups, there is significant variability between those Telegram groups. Therefore, analysing the scale of the use of images by individual extremist groups may be a simplification of more complex behaviour. The posting behaviour of a group as a whole can provide useful context but exploring multiple unique Telegram groups underneath the banner of an extreme right group can provide more insight into the behaviours of extreme right groups online.

This variability between extreme right groups, with multiple Telegram groups belonging to a single extreme right group, is reinforced and explored in more detail in Figure 6 whereby the figure visualises the average number of images posted per day from each Telegram group. From this figure, posting behaviour (in the form of average posting frequency) varies significantly not only between extremist groups but within the numerous Telegram groups under the same extremist group banner. From this figure, two groups showcase a high posting frequency, B&H and PB. In the case of B&H, the B&H1 Telegram group has a significantly higher posting frequency when compared to the other two Telegram groups. The B&H1

Telegram group averages almost 10 image posts per day, whereas B&H2 and B&H3 average one or fewer image posts per day. This is an indication that posting behaviour again is specific to the Telegram group as opposed to the group. In the context of the Proud Boys, there is also notable in-group variability. PB2 approaches 11 image posts per day, and PB5 averages eight image posts per day. This notably drops in the cases of PB4 and PB5 which average around four image posts per day. And the remaining four Proud Boys Telegram groups average around one post per day. This further reinforces the understanding that posting behaviour differs not only between extreme right groups but also within those groups.

To consider the extent to which imagery is adopted as a medium by extreme right groups, this section has looked at the number of image posts, and average posting frequency per day. In doing so, the data indicates that images are present throughout all six of the groups and all 17 of the Telegram groups. However, between groups, there is significant variability in posting behaviour. Furthermore, in the cases where groups have multiple Telegram groups, there is also significant variability in posting behaviour. Therefore, it can be claimed that imagery is a medium utilised by the extreme right, however, the Telegram group belonging to a given extreme right group accounts for posting frequency more than the group to which they belong. In that, the members and culture formed in specific Telegram groups appear to determine image posting frequency above the extreme right group or collective that they belong to.

[RQ2: To what extent is graphic imagery used in the communicative strategies within extreme right imagery?](#)

The second research question is whether graphic imagery is present within extreme right imagery and the extent to which it is prevalent. To address this, reference will be made to Table 4. This table refers to the third component of the framework titled 'Shock Value'. The criteria for categorising an image in this category must meet one of the following three conditions: The first of these conditions is whether an image triggers an instinctive reaction of disgust by the coder, the second criterion is whether the image displays corpses, and the third refers to the presence of dismembered or beheaded bodies. To further expand these findings, Table 6, which contains the fifth component of the framework titled 'Symbols', will also be consulted. Within this table, the fifth subcategory titled 'Weapons/Gore' will be

referred to. This subcategory is defined as ‘the presence of weaponry within images, in addition to the presence of gruesome and graphic content including displays of blood or nudity’, which, due to the incorporation of gruesomeness, can also be taken into consideration when considering the full scope of graphic imagery in the dataset. In summary, to understand whether graphic imagery is present within the dataset, two characteristics will be interrogated there are the shock value category and the weapons/gore subcategory within the symbols category.

Before analysing the extent to which graphic imagery is prevalent within extreme right imagery, it is first necessary to look at whether graphic imagery is commonplace within the dataset. First, consulting the ‘Shock Value’ category in Table 4, which captures the entire dataset, the data shows that 2.05% of the images contain graphic imagery. To reframe this point, when analysing this dataset, for every 100 images analysed, on average, over two of those will contain graphic imagery. Thus, while according to the Shock Value category as a whole, there is a small quantity of graphic content in the dataset, there is noticeable variability between groups.

Continuing to analyse the ‘Shock Value’ category found in Table 4, out of the six extremist groups and 17 Telegram groups, roughly half of the extreme right groups and Telegram groups have at least one instance of graphic content. This is a significant proportion of the data, particularly when one considers the potential prevalence of this across numerous platforms where the total number of images is significantly scaled up. Specifically, six Telegram groups did not contain any content that constituted as having a shock value. In the cases of TB and NSO, both groups did not showcase any graphic imagery, including in the cases where there were multiple Telegram groups. This points toward a group-wide choice in visual style, which will be covered in more detail in the coming research question sections.

Four extremist groups did contain shocking (graphic) content, i.e. AWD, B&H, FKD and PB. In the cases where graphic content was present, there is variability in the extent to which this content is present. For instance, three groups contained over 2% of shocking content, these groups were AWD (2.04%), B&H (7.61%, 10.83%, and 7.11%) and FKD (6.18%). In the case of the PB, four Telegram groups contained graphic imagery, and another four did not. Where

shocking content was present, its prevalence never exceeded more than 1% of all the images. This data firstly demonstrates that one extremist group can showcase multiple visual styles, as is the case with the PB. Secondly, that shocking and graphic imagery appears to be intentionally used and not used by groups. This is highlighted in cases where an extreme right group has multiple Telegram groups and maintains a similar percentage of shocking and graphic imagery across those Telegram groups. For example, TB Telegram groups do not contain any shocking imagery, B&H contains between 7-11% across three Telegram groups, and the PB does not exceed 0.54% across eight Telegram groups.

From the data, it is clear that groups such as AWD, B&H, and FKD are more willing to use shocking context, whereas NSO and TB avoid it entirely, with PB including a small quantity. Furthermore, although understanding extremist behaviour according to groups can be useful, there is the capacity for inter-group variability in several contexts, including shocking and graphic imagery. Furthermore, because graphic content is minimal in this dataset, existing tools, such as those developed for detecting ISIS's highly graphic propaganda, cannot be effectively utilised or repurposed to moderate extreme-right content. This highlights the need for policy recommendations that prioritise research into automated systems and collaboration on developing an understanding of distinct traits of extreme right imagery. Such efforts are essential to developing more nuanced and accurate moderation strategies in the future.

Building on the understanding gained from the 'Shock Value' category, the 'Weapons/Gore' subcategory found in the 'Symbols' category within Table 6 provides more insight into the findings. The instances of images containing weapons and gore are present in over 19% of the data across all groups. Furthermore, nine Telegram groups had less than 10% of total images containing Weapons/gore images in the 'Symbols' category. This includes two out of three B&H groups (7.09-7.51%) and seven out of eight PB groups (1.23-9.50%). A total of eight Telegram groups exceeded the 10%. This includes AWD (44.90%), one of three B&H groups (39.66-44.90%), FKD (42.62%), NSO (33.33), one of eight PB groups (12.43%), and all three TB groups (11.76-46.55%). Similarly to the case of Shock Value, Weapons/Gore appears to be intentionally used and not used by groups. This notion is reinforced by the similar percentages across extremist groups with multiple Telegram groups. For example, the three Telegram TB groups contain over 10% of Weapons/Gore within their imagery and the PB Telegram groups,

excluding PB6, do not exceed 10%. In B&H, PB, and TB, there is one channel that is an outlier to the other channels. Despite stating that graphic imagery is present in extreme right imagery, this data indicates that it is more of an intentional visual style preference by certain groups as opposed to the entire extreme right collective.

Whilst weapons and gore are not necessarily indicative of graphic content, there is room to interpret them as tools commonly used by such groups. Results from this subcategory add strength to the claims of the previous paragraph on across-group differences, where stating the presence of categories across the extreme right as a whole can provide useful but simplified insights. To reinforce this point, in the cases of groups with multiple Telegram groups, i.e., B&H, PB, and TB, these all showcase significant variability in their prevalence of weapons and gore. The PB Telegram groups, in the case of B&H Telegram groups, B&H2 and B&H3 fail to exceed 8% of images containing weapons and gore. Whereas, in the case of B&H1, this percentage reaches 39.66%. This finding is replicated in the case of TB, whereby, of TB1 and TB3's images, over 40% contained weapons and gore, whereas, in the case of TB2, this figure falls short of 12%. Further evidencing both across-group and within-group variability.

This section has approached the questions of whether graphic imagery is present within extreme right imagery and the extent to which it is prevalent. In doing so, numerous findings have been developed. Firstly, graphic content can be found across all extremist groups from this dataset. However, in exploring the extent to which this is the case, the claim is true to varying extents. In some cases, it appears to be a choice and preference, such as in the case of B&H, due to the high posting frequency of graphic imagery compared to other extreme right groups. In the case of other groups, avoiding this category of content also appears to be intentional, particularly in the case of the PB, due to the low posting frequency of graphic imagery. Exploring this research question reveals that while graphic imagery is present in the extreme right content, it's prevalent only when the number of images is significantly increased, raising the observation that the proportion of graphic imagery is substantial enough to warrant concern consideration for targeted moderation.

RQ3: Is there an extreme right visual style that is communicated and reinforced through the use of imagery?

Having already established that the extreme right utilises imagery to varying extents and that graphic imagery is present within extreme right imagery to an extent, we are led to the third research question. This question poses whether there is an extreme right visual style within the imagery. In other words, this question explores whether there are consistent themes across all the extreme right groups and Telegram groups in the dataset, which gives a confident picture of visual choices and preferences across the extreme right. To approach this question, Tables 2 through 6 will be addressed. Each of these five tables accounts for each of the five components that form the framework. These consist of the following five framework categories: Master Frame, Narratives, Shock Value, Image Format, and Symbols. By addressing the total results for each of these categories, the question will be addressed.

The first of the five figures and framework categories is the 'Master Frame' category found in Table 2. The 'Master Frame' category captures crisis and solution frames in the images through five distinct subcategories. This table presents that over half of the dataset's imagery (53.53%) refers to the in-group positively. Additionally, almost a quarter of the dataset's imagery (24.20%) refers to the out-group negatively. The remaining (just over) 22.27% is accounted for by the remaining three categories, which capture emphasising division, the absence of a master frame and inconclusive cases. From these findings alone, several conclusions can be drawn. Firstly, over 80% of the dataset of images targets the in-group or the out-group in their messaging, uncovering that messaging is clear, targeted, and intentional. Using the definition established at the beginning of this thesis, referring to the in-group and the out-group of the extreme right (as inspired by Berger, 2015), it can be inferred that images are not used for a separate purpose to other forms of propaganda (namely text and speech), rather, they are another medium to communicate sentiments shared within a group and its members. Furthermore, with almost half of the dataset meeting the 'In-Group Positive' criteria, there is an intention behind using images to reinforce and strengthen the in-group and its accompanying ideology. Beyond this, portraying the out-group negatively is a second priority after the in-group has been appealed to.

The second category found in Table 3 identifies the 'Narratives' category of the framework. This category of the framework comprises 10 subcategories. To best understand and frame the results from this section, the subcategories can be grouped into three groups. The first grouping combines the following four narrative subcategories: 'Anti-Identity', 'Anti-Cultural Marxism', 'Conspiracy Theory', 'In-Group Victimization', and 'Call to Action'. These subcategories are grouped as they each individually account for less than 10% of the dataset results, and therefore, these narratives are not understood as a priority by the extreme right. Most notably, 'Call to Action' is in this grouping at 9.85%, which indicates that the role of imagery in extreme right communication is not largely aimed at mobilising members. The second grouping also contains four narrative subcategories, which consist of: 'Anti-Establishment/Global Elite', 'Global Politics and Military', and 'Other'. The criterion for this grouping is a prevalence in the imagery of between 10-15% of the narrative dataset. Out of these three, the most prevalent subcategory is 'Global Politics and Military' at 12.01%, highlighting an interest in reference to historical and current events. However, the fact that even the 'Other' subcategory is made up of images communicated in an ideology-focused Telegram group means that, although not clearly targeted, they are likely intended to build community, and so all images could be said to indirectly build the in-group identity. The third and final grouping of narrative subcategories is made up of the 'Activism' category at 44.45% and 'Humour' at 16.54%. The prevalence of these two narratives aligns with the overall results of the master-frame results. This is because the majority of the dataset refers to the in-group positively. 'Humour' and 'Activism' are approaches to reinforce and strengthen the in-group by strengthening the group's identity through the use of visual communication and a sense of belonging and reinforcing it through humour. The use of humour also presents a secondary effect of reducing the severity of topics being discussed and the optics of a given group. These two narratives further develop an understanding of the visual style of the extreme right.

Revisiting the third category that was discussed in RQ2, the 'Shock Value' category in Table 4 provides useful insight into the visual style of the extreme right. With only 2.05% of the dataset being composed of graphic imagery, this leaves 97.95% of the dataset to be defined as non-graphic imagery. As previously surmised, when analysing the images which form this dataset, for every 100 images encountered, on average, just over two of those will contain graphic imagery. This tells an important story about the use of graphic imagery by the extreme

right. In that, whilst it can be stated that graphic imagery is present within extreme right imagery, it is not a defining feature of the extreme right visual style. The infrequent presence of graphic imagery does appear to be reflective of the visual style of the extreme right more broadly.

The fourth category is the 'Image Format' category found in Table 5. This category is split into two halves, with one half categorised as 'Single Image' (an image without text added in post-production) and the other half categorised as 'Image & Text' (an image with text added in post-production). The single imagery half comprises the majority of the dataset, accounting for 55% of the dataset's images. From this, the 'Undoctored Photograph' (25.37%) and 'Screenshot' (16.21%) subcategories are the most common of the four subcategories. Indicating that there is a preference toward simplicity over sophistication, in that, undoctored imagery can be uploaded and posted quickly and with ease. Of the four subcategories under the 'Image & Text' half, the 'Graphic Design' (16.44%) & 'Meme' (12.82%) subcategories are most present, thus, in cases where editing occurs, there appears to be a preference toward creative and edited imagery as opposed to minimally edited imagery. The graphic design element could be demonstrating an inspiration from groups like ISIS, which, as discussed in the literature review, used high production quality to blend into mainstream media and enhance credibility (Ingram 2014, 2015 and Milton 2016). The use of memes showcases the building of an in-group community, as well as the use of seemingly innocent formats to leverage platform affordances to evade moderation (Tech Against Terrorism, 2022; Doerr, 2021). These findings communicate a visual style of the extreme right in the context of an image's format. There is a high presence of unedited photographs and screenshots and a less frequent presence of highly doctored, sophisticated imagery in the form of graphic designs and memes.

The fifth and final category is the 'Symbols' category, which is located in Table 6. This category comprises ten subcategories which capture the broad spectrum of symbols adopted and utilised by the extreme right. This category can also be split into three groupings to best understand the results. The first and most populated grouping is the less than 15% grouping, which accounts for five of the ten symbol subcategories: Acronym/ Abbreviations (2.64%), Slogans (6.40%), Religious Symbol (1.27%), Clothing Brands and Print (3.86%), News Outlet

(3.23%). The second grouping is between 15-25% of the dataset, which accounts for two of the three remaining subcategories. These are the Weapons/ Gore (19.28%) and Key Figures (15.54%) subcategories. The third and final grouping consists of categories with a greater than 25% presence in the dataset. These subcategories include the Common Extremist Symbols (31.74%), Extremist Organisations – (28.88%), and No Definitive Symbols (34.39%). This reinforces the findings of the previous 'Image Format' category, which found that extreme right imagery trends toward unsophisticated imagery, as evidenced by the presence of unedited images being the most frequent image type. The absence of symbols in most cases communicates that although these modes of communication are available and occasionally used, it does not account for most images. Thereby, the visual style of the extreme right frequently showcases an absence of symbols. However, where symbols are present, they mostly identify the extremist organisation posting the content, in addition to identifying key figures. However, as evidenced in the literature review, there may be a greater importance placed on symbols as certain symbols may have been unidentifiable to the author of this thesis. As Mattheis (2022) points out, certain symbols depend on shared cultural or historical knowledge, emphasising the audience's role in interpreting imagery (see also IRTL, 2022). This highlights that analysis often depends on researchers' cultural and historical expertise, introducing potential bias (Forchtner and Kølvråa, 2017), pointing to a need to expand our understanding of potential symbols of extreme right groups so that they can be accurately identified and incorporated into the framework.

To summarise, the extreme right visual style addresses both the in-group and the out-group of the extreme right. Activism and humour are the most prevalent narratives in the imagery, graphic content is present yet infrequent, and there is an emphasis on unedited and highly edited imagery with few cases of slightly manipulated imagery, with infrequent use of symbols beyond the identification of extremist organisations and key figures. Whilst this analysis has drawn this visual style from the data. It is necessary to reiterate that 88.7% of the data is for just one of the groups (Proud Boys). Whilst these findings give essential insight into empirical understandings of the extreme right, the disparity between image count across groups in the dataset limits the utility of this finding to online platform moderation efforts.

This section has addressed the third research question, which poses whether there is an extreme right visual style within the imagery. In doing so, each of the five categories which formed the extreme right image analysis frame was individually addressed. This drew from the most populated subcategories from the entire dataset and resulted in the establishment of the following categorisation of the visual style of the extreme right.

RQ4: Do extreme right groups have differing visual styles that are communicated and reinforced through the use of imagery?

The fourth and final research question builds on the previous, having established core elements of an extreme right visual style, this question is aimed at expanding this understanding further. In doing so, the question posed asks whether specific groups belonging to the extreme right collective have different visual styles. To explore this answer, each of the ten extremist groups will be looked at through each of the categories of the framework. In the cases of the extreme right groups with multiple Telegram groups, the same approach will be taken by looking at each of the Telegram groups simultaneously. Before this, it is necessary to reiterate that the dataset for this thesis was collected to achieve a dataset that reflected the extreme right by including a range of RWE groups. Without claiming to be a perfect representation of the entire extreme right, it is the opinion of this thesis that the dataset and its findings do give insight into the wider extreme right because of this inclusion of a range of designated extreme right groups.

Starting with the AWD Telegram group, the data contained 49 images from a Telegram group that was live for 203 days at the point of collection. Equating to roughly one image posted every four days, AWD showcases a low posting frequency. Looking beyond posting behaviour to visual style, AWD showcases significant similarities to the extreme right visual style. For master frames, there is a notable high occurrence of emphasising division, however, referring to the in-group positively is the most common master frame. Activism is the most common narrative. However, there is also a distinctly high occurrence of anti-identity images compared to the average of the extreme right as a collective. There is a notable occurrence of graphic images at 2%, which aligns with the key themes of the accelerationist ideology on which the group is based, namely by achieving societal collapse through displays of violence (Ware, 2020). The image format highly resembles the extreme right, whereas symbols also

differ, with a high occurrence of common extremist symbols and weapons/gore which confirms the violent reputation of the group (Fulgoni and Menis, 2024). As a result, this represents a minor variance in the visual style of AWD compared to the extreme right. These findings require more data to be strengthened. However, useful insights can be drawn from the trends in data extracted from this Telegram group.

The B&H group is the first of three groups to contain multiple Telegram groups. In order of B&H1 – B&H3, the Telegram groups have 1064, 508, and 253 images spread over 110, 442 and 587 days. It is immediately clear that these three Telegram groups have notably differing posting frequencies, with the Telegram group with the most posts being live for the fewest number of days and vice versa. Each of the Telegram groups is predominantly populated by referring to the in-group positively and using the activism narrative. This finding aligns with the modus operandi of B&H which has been known to media and technology to recruit and radicalise followers (Southern Poverty Law Centre, 2024). All three Telegram groups exceed a 7% presence of graphic imagery. Akin to the extreme right more broadly, there is a focus on undoctored photographs and doctored image formats, with B&H2 containing a high occurrence (35.24%) of graphic designs. Lastly, common extremist symbols and not definitive symbols were the most common symbol subcategories, with B&H1 almost reaching 40% of weapons/gore. There is a clear visual style by B&H, this style is consistent with the broader style of the extreme right and is mostly shared across all three Telegram groups. There are examples where this consistency across Telegram groups is interrupted, however, not so drastically as to claim that one Telegram group has an entirely different visual style to another, but enough to account for inter-group visual style variability.

Moving on to the next extreme right group, the FKD accumulated a total of 922 images from a Telegram group, which was live for 979 days at the point of data collection. This equates to around one image posted per day for approaching three years. The group actively refers to the in-group positively, with this master frame accounting for almost 70% of the dataset. In line with this finding, Activism is the most populated narrative at above 65% of the dataset and aligns closely with the values of accelerationism that FKD is founded on (Newhouse, 2021). Notably, graphic imagery is at 6%, which differs substantially from the extreme right as a whole and more closely aligns with the AWD, and NSO, the other accelerationist groups.

Undoctored images and unedited screenshots account for most of the image format and there is a high occurrence of common extremist symbols, extremist organisations and slogans. FKD presents a visual style which proactively appeals to the in-group and infrequently refers to the out-group graphically.

The next group is NSO, with only three images posted by the Telegram group. The apparent lack of images belonging to this group is itself a finding and confirms the short lifecycle of attempted revival and rebranded AWD (Counter Extremism Project, 2022; Fulgoni and Menis, 2024). However, this finding pertains to posting behaviour rather than the visual style of the group. Despite having a Telegram group for over 600 days, there are only three images, equating to one image posted for every 200 days the Telegram group was live. This indicates that images are not proactively posted and shared by the NSO. Beyond this finding, it would provide little benefit to try and use these three images to make claims on NSO's visual style. As a result, this research will refrain from this and determine the visual style of the NSO as inconclusive.

The largest group in both total image count and number of distinct Telegram groups is the PB with 22656 images spread across 8 different Telegram groups. In chronological order of PB1-PB8, the Telegram groups have 2515, 10654, 2357, 3813, 2189, 465, 553, and 110 images spread over 1295, 961, 580, 881, 275, 415, 987, and 408 days. What is necessary to draw from these data is that there is not a definitive correlation between the number of days a Telegram group has been live and the number of images it produced. In the case of master frames, all Telegram groups prioritise the out-group negatively and the in-group positively. In all cases, barring PB1, out-group negatives exceed the count of in-group positives. In the case of narratives, humour and activism are most present. However, the extent to which these narratives are present varies significantly. Depending on the group, all the narratives are populated throughout, with the anti-identity and anti-establishment/global elite narratives being highly populated by certain Telegram groups. The high presence of content commenting on identity, politics and key figures is aligned with existing literature and understanding of the PB (McBain, 2020; Nagle, 2017). Consistently across all Telegram groups, there are few to no instances of graphic content. Similarly, across the extreme right, there is a high frequency of undoctored photographs and screenshots in conjunction with graphic

designs and memes. The absence of graphic content and memes are common characteristics of alt-right groups like the PB (Kriner and Lewis, 2021). Looking lastly at symbols, key figures and no definitive symbols are most present across all Telegram groups, with infrequent spikes in common extremist symbols, extremist organisations and clothing brands and prints, another example of characteristics of the PB found in the literature (ADL, 2021; Kenes, 2020). The inherent high posting frequency demonstrated by the PB is an outlier to the rest of the extreme right groups on Telegram.

The group TB is the last of the extreme right groups with multiple Telegram groups, with three in total. In chronological order of TB1-TB3, the Telegram groups have 17, 17, and 58 images spread over 66, 376, and 927 days. Between all the Telegram groups, there is an indication that imagery is an infrequently adopted medium of communication. Looking at the master frames TB1 and TB3 both target the in-group positive category over 90% of the time. Contrastingly, TB2 does not target any master frame over 60% of the time, showcasing distinct differences between Telegram groups. This is reflected in the narrative category, where TB2 uses no narratives over 70% of the time, whereas this category is not present in either of the other two Telegram groups. Instead, TB1 and TB3 utilise activism, calls to action, global politics and military narratives. These findings mirror sentiments in the literature which frame TB as a militant accelerationist group (Renard and Rekawek, 2024). However, reinforced by the presence of graphic imagery found in the dataset. Furthermore, there is a mix of image formats, primarily including undoctored photographs, doctored images, and graphic design. Continuing the theme between Telegram groups, 70% of the data in TB2 has no definitive symbols, whereas TB1 and TB3 are populated by common extremist symbols, extremist organisations and slogans. From these findings, TB1 and TB2 reflect many of the visual style characteristics found in the extreme right as a whole. In the case of TB3, there are stark differences which reinforce the sentiment, which is reinforced by this dataset, that it is not possible to confidently state that every extreme right group has a definitive visual style. Rather, interpreting the visual style of an extreme right group requires a level of nuance.

Having established in RQ3 that there is a visual style presented by the extreme right, this research question looked to understand this visual style further. In doing so, an understanding of cross-group and intra-group differences was further uncovered. Firstly, across the dataset,

it became apparent that the time a Telegram group was live did not have a significant impact on the number of images posted, which means posting may be a result of other factors such as engagement levels of Telegram group users, the number of Telegram group users or the intentions of Telegram group users. Image sharing is more present in certain extreme right groups over others and differs from one telegram group to another, even if the telegram groups belong to a single extreme right group. A crucial finding to extract is that this framework can discern the visual style of the extreme right. Though there were outliers and variations, results in evidenced trends and patterns could be drawn to form a definitive style. Despite this, due to the overrepresentation of one group in the dataset (the PB), this thesis cannot claim to have uncovered this visual style of the extreme right that online platforms can use to inform their content moderation practices. However, the framework has shown to be a useful tool and should be further explored with a more diverse correct dataset.

Across extreme right groups, there are distinct differences in the use of graphic content. Groups such as AWD, NSO, and B&H showcase a higher occurrence of graphic imagery compared to the extreme right collective average. AWD's posting aligns with their accelerationist ideology groups in the dataset due to graphic content playing a significant role in conveying messages of violence and societal collapse. Similarly, B&H, known for its active use of media for recruitment (Counter Extremism Project, 2024), has a consistent visual style across its multiple Telegram groups, with all exceeding 7% graphic content. Conversely, the PB Telegram group avoids and minimises the use of graphic imagery, focusing instead on undoctored photographs, doctored images, and memes. This disparity between groups evidences the variability in content strategies within the extreme right, suggesting that graphic content is not a universal characteristic across the extreme right, but a tool leveraged on a group-to-group basis.

Memes and humour emerge as prominent features in the visual style of certain extreme right groups, most notably the PB. The PB heavily use humour narratives and memes, seemingly appealing to broader audiences, normalising their ideology, and evading content moderation. Such content aligns with existing literature on alt-right strategies, emphasising identity, politics, and key figures through approachable and satirical imagery (Hawley, 2017; Greene, 2019; Doerr, 2021; Yiannopoulos and Bokhari, 2016). The PB contrasts with groups like TB and

AWD, where humour is absent, and the focus shifts toward activism and reinforcing the in-group. While specific image characteristics are integral to some groups, their absence in others suggests a fragmented visual strategy. Whilst this fragmentation can be explained as a behavioural choice by groups that tailor their content to the goals and preferences of each group, it is also necessary to consider that there may also be a lack of strategy or leadership which is reflected in the data.

The visual style of the extreme right is drawn from the collected results of the six extreme right groups and 17 distinct Telegram groups that form the dataset. Novel insights were drawn from specific groups and cases of multiple Telegram groups that diverge from the visual style of the extreme right. Certainly, specific extreme right groups have a visual style that appears unique to a specific extreme right group, such as the PB, which posts more memes and B&H, which posts more graphic content than other groups in the dataset. However, the visual styles of extreme right groups are informed by the location from where data is extracted. In certain cases, they can differ from one Telegram group to another, and it is also feasible to consider that they may differ from one platform to the next.

### Policy Recommendations

The growing prevalence of extreme right content online poses significant challenges for platforms, policymakers, and researchers. This thesis, through its analysis of visual propaganda, underscores the need for targeted and collaborative efforts to address these challenges effectively. The findings presented have revealed multiple strategies employed which complicate detection and moderation, requiring a multifaceted approach that incorporates advanced technologies, human oversight, and public engagement. This section outlines policy recommendations aimed at improving the detection, understanding, and moderation of extreme right content by platforms. To achieve this, they involve multiple stakeholders, which work with platforms to moderate extreme right material, including Governments, Non-governmental Organisations, and Researchers. The recommendations are structured around three core objectives: developing knowledge through research, utilising symbols as a cornerstone of content moderation, and increasing understanding of extreme right group dynamics. Each objective builds on insights from the dataset analysis and

literature review, offering practical, evidence-based solutions to address the evolving nature of extreme right propaganda.

The recommendations emphasise the importance of fostering collaboration between stakeholders in this arena. Transparency in moderation systems, the development of adaptive frameworks, and public education initiatives are highlighted as foundational measures for creating a proactive and sustainable approach to countering extremism online. By aligning these efforts with human rights principles and cultural sensitivities, the proposed policies seek to balance the need for effective moderation with the protection of legitimate expression. In presenting these recommendations, this section provides a roadmap for advancing and supporting the content moderation practices of platforms by fostering a coordinated, informed response to the proliferation of extreme right content. The aim is to empower platforms and stakeholders with the tools and frameworks necessary to navigate the complexities of moderating extreme right imagery online while maintaining ethical integrity.

#### Policy Recommendation 1: Developing Knowledge Through Research

The effective moderation of extreme right content, particularly visual propaganda, requires a foundation of robust research and knowledge-sharing. The dataset analysis and literature review demonstrate the need for collaboration between stakeholders to improve understanding and detection of the symbolic and implicit strategies employed by extreme right groups (e.g. European Commission, 2022). To achieve this, the central aim of this recommendation is to foster transparency in automated moderation systems (1a), supported by the development of adaptive analytical frameworks (1b) and public education initiatives (1c). These supplementary measures underpin the primary goal of creating an evidence-based, cooperative approach to countering extreme right content.

#### **1a. Sharing Structures of Automated Content Moderation**

##### **Finding:**

The literature review and research findings underscore significant challenges in accessing the underlying structures, principles, and operational mechanisms of automated content moderation systems. These systems are critical for detecting and mitigating harmful content,

yet their effectiveness is hampered by a lack of transparency, with the criteria for flagging and categorising content often opaque (Duarte et al., 2017; Koetsier, 2020). Platforms often prioritise the protection of proprietary information and user privacy, leading to restricted access for external researchers and civil society stakeholders. This lack of openness has created a significant trust gap between platforms and external actors, stalling collaborative innovation and impeding the development of adaptive solutions to emerging moderation challenges. Researchers have repeatedly highlighted that this opacity limits the ability to understand the nuances of these systems, including their handling of implicit extremist messaging, such as symbolic and visual propaganda employed by extreme right groups (Hendrycks and Dietterich, 2019). Without adequate access, researchers struggle to provide data-driven, targeted recommendations to address the specific challenges posed by such content. The need for transparency is particularly acute given the adaptability of extreme right groups, who exploit gaps in moderation frameworks to evade detection (Tech Against Terrorism 2022). As demonstrated in this study's dataset analysis, implicit visual strategies and symbolic communications is seen in extreme right imagery, enabling these groups to circulate their messages without triggering traditional moderation systems. Platforms' reluctance to share non-sensitive insights into their systems not only hampers innovation but also undermines the accountability needed to build public trust and confidence in their moderation practices (Facebook, 2024).

**Recommendation:**

Online platforms must establish structured and secure mechanisms to provide researchers with access to non-sensitive insights into their automated moderation systems. These mechanisms should focus on enabling meaningful collaboration while safeguarding user privacy. Access to anonymised, structured data that reflects content flagged and processed by moderation systems will allow researchers to identify patterns, gaps, and weaknesses in current practices. Additionally, platforms should support controlled testing environments where researchers can evaluate and refine algorithms without affecting live systems. Regular forums for dialogue between platforms, researchers, and related stakeholders, such as civil society organisations, should also be established to facilitate knowledge-sharing and collaborative problem-solving. To ensure fairness and accountability, platforms should

implement third-party audits of their moderation systems to assess their effectiveness in detecting implicit and evolving extremist content.

**Impact:**

Enhancing transparency will enable researchers to conduct targeted investigations into the limitations and strengths of existing moderation systems, providing platforms with actionable insights for refinement. This will be particularly effective in addressing nuanced challenges, such as implicit visual propaganda and symbolic messaging, that current systems often fail to detect. By leveraging academic expertise, platforms can develop more robust algorithms that are adaptive to the evolving tactics of extremist groups. Collaboration between platforms, researchers, and civil society organisations will not only strengthen trust and accountability but also ensure that moderation practices are informed by a diverse range of perspectives and expertise. This recommendation aligns with calls in the literature for increased openness and innovation in moderation frameworks. The proposed measures will help bridge the gap between academic research and platform operations, enabling a more effective, evidence-based approach to countering online extremism.

**1b. Research and Adaptive Analytical Frameworks for Extreme Right Imagery**

**Finding:**

Extreme right imagery uses implicit messaging and symbolic strategies to convey ideological narratives, as demonstrated in the dataset. These visuals often operate without accompanying text (55% of this dataset), complicating detection through existing analytical methodologies. Frameworks like SF-MDA (Jewitt et al. 2016) are effective for multimodal content but struggle with standalone visuals, while grounded theory approaches alone (e.g. Charmaz 2006; Berg 2017), although flexible, can lack systematic validation for large-scale datasets. The decentralised and evolving nature of extreme right networks, as highlighted in the literature review, further complicates detection efforts, demanding adaptive and iterative tools (Lorenzo-Dus and Nouri 2021; Rapoport 2004; Suttmoeller et al. 2018; Hawley 2018; Taylor 2020; Bliuc et al., 2020; Caiani et al., 2012; Schain, 2006). Additionally, there is limited research into the multiple functions of extreme right imagery (Caiani 2024), such as mobilising

supporters (Halfmann and Young 2010; Corrigan-Brown and Wilkes 2014) reinforcing group identity (Adami 2020), and vilifying out-groups (Baele et al. 2020a), and how these functions interact with cultural and contextual factors (Milman and Doerr 2021; Windish and Simi 2022). Without a deeper understanding of these dynamics, the development of effective moderation strategies remains incomplete. Moreover, the transparency fostered through the implementation of Recommendation 1a (sharing the structure and underpinnings of automated content moderation systems) has the potential to significantly fuel this endeavour. Insights gained from enhanced access to moderation frameworks, data structures, and operational principles will offer researchers the foundation to refine and validate analytical methodologies. This will enable more accurate categorisation and understanding of extremist imagery. This framework is argued to be a proof of concept. New insights about extreme right imagery have been gained by developing a hybrid framework which could be strengthened by larger datasets and continued refinement.

**Recommendation:**

Platforms and researchers should build on collaborations suggested in Recommendation 1a to develop hybrid analytical frameworks that integrate inductive and deductive methodologies. By extending sharing agreements from automated structures to data that feeds these, the research could become better aligned with the real-world threat of extreme right imagery and be used to develop more efficient methodologies. The insights from such research can then feed back into automated classification. Building on existing models like Baele et al. (2020a), these frameworks should incorporate new categories that capture narrative-specific motifs and implicit symbolic messaging, particularly in standalone visuals. Research initiatives should prioritise testing frameworks for the analysis of text-free datasets to refine methods for interpreting content that relies solely on visuals. Further studies must explore the ideological narratives, emotional appeals, and communicative functions of extreme right imagery, investigating how these elements operate within cultural, historical, and political contexts. Insights gained through transparent collaboration with platforms should directly inform these methodologies, offering an empirical basis for developing adaptive frameworks that respond to evolving extremist strategies.

**Impact:**

Experiences in developing and using more robust and adaptive analytical frameworks will enable researchers and platforms to decode the implicit and symbolic strategies of extreme right imagery with greater accuracy. By leveraging insights obtained through enhanced platform collaboration, these methodologies can be refined to reflect real-world applications, resulting in more precise identification and categorisation of extremist visuals. This approach will also expand understanding of the ideological and emotional dimensions of extremist imagery, facilitating the development of targeted, culturally sensitive interventions that address its societal and psychological impact. Ultimately, the iterative feedback loop created by integrating findings from Recommendation 1a into analytical frameworks will provide a more dynamic and responsive approach to addressing extreme right imagery. This synergy will not only enhance research outcomes but also strengthen the overall efficacy of content moderation systems, ensuring they remain adaptive to new and evolving extremist tactics.

**1c. Enhancing Media Literacy and Public Awareness****Finding:**

The dataset and literature review reveals the sophistication of extreme right visual propaganda, which leverages implicit messaging to foster ideological alignment (e.g. Caiani 2024; Windish and Simi 2022; Mattheis 2022). These strategies often involve cultural nostalgia, historical motifs, and subtle semiotics, making them difficult for the public to identify (Baele et al. 2020a; Forchtner and Kølvråa 2017; Mattheis 2022). A lack of public awareness about these tactics not only facilitates their spread due to a limited understanding of their origin and purpose (Backhaus, 2023; Pantucci, 2022) but also perpetuates a cycle of ineffective pressure on platforms. Instead of demanding nuanced, threat-based action, public discourse and even Government criticism can default to calls for a blanket or quick content eradication (e.g. Network Enforcement Act, 2017; Australia's Criminal Code Amendment 2024; Conseil Constitutionnel, 2020), which can lead to overzealous censorship, legal challenges, and the further decentralisation of extremist networks (Tech Against Terrorism, 2023b).

Education plays a critical role in fostering informed public engagement with these issues. By understanding the complexities of extremist visual strategies, stakeholders such as the public

can advocate for the right type of pressure on platforms, one that prioritises proportionate action based on the specific threat posed by different forms of content. This includes urging platforms to adopt greater transparency in their algorithmic decision-making and moderation practices (as recommendation 1a discusses), enabling researchers and policymakers to better assess the effectiveness of these measures. Such an approach recognises that eradicating all extremist content online is neither feasible nor always desirable, as it risks pushing such material into harder-to-monitor spaces. Instead, a balanced strategy, rooted in fairness and transparency, can strengthen both public trust and the efficacy of counter-extremism efforts.

**Recommendation:**

Governments and civil society organisations should collaborate to launch targeted educational campaigns to equip the public with tools to identify and critically assess extremist imagery. These campaigns should include multimedia resources, workshops, and online modules designed to help individuals decode implicit messaging strategies such as coded symbols, historical references, and ideological motifs. They should also increase public understanding of the broader societal threat posed by extreme right content online, highlighting how it undermines democratic values and fuels polarisation.

**Impact:**

Enhanced media literacy will empower individuals to recognise and report extremist content, reducing its dissemination and influence. By fostering a more informed public, these campaigns will not only mitigate the normalisation of extremist narratives but also exert pressure on platforms to share their content moderation practices and algorithms. This feedback loop will create a demand for greater transparency and accountability, supporting Recommendation 1a by encouraging platforms to open their systems to scrutiny and collaboration. By demystifying the implicit strategies of extreme right propaganda, this recommendation heightens the likelihood of proactive responses of users by flagging and disengaging and reduces the ability of this imagery to infiltrate mainstream discourse.

## Policy Recommendation 2: Utilising Extreme Right Symbols for Content Moderation

As seen in the dataset analysis and literature review, symbols are a central feature of extreme right propaganda, functioning as key markers of ideological alignment, group identity, and mobilisation strategies, highlighting the important role that symbols play in extreme right imagery. To effectively leverage this insight, platforms must prioritise symbol recognition in their content moderation. Central to this recommendation is the development and maintenance of a global database of extremist symbols (2a), supported by cross-platform collaboration to ensure consistency and scalability in implementation (2b). Together, these measures will enhance the ability of platforms to detect and mitigate the spread of extremist content effectively.

### **2a. Utilising Symbols to Flag Extreme Right Content**

#### **Finding:**

The dataset analysis highlighted that symbols are a pervasive and critical component of extreme right imagery, with nearly 65% of the analysed content containing at least one symbol and nearly 30% linked to identifiable extremist organisations. These symbols act as visual shorthand, encapsulating ideological messages, signalling group identity, and fostering cohesion within extremist networks. The literature review corroborates this finding, demonstrating that symbols evoke shared historical and cultural narratives, enhancing their resonance with target audiences (Yiannopoulos and Bokhari, 2016; IRTL, 2022; Windish and Simi 2022; Caiani's 2024; Baele et al. 2020a; Mattheis 2022). Moreover, their decentralised and replicable nature makes symbols effective tools for disseminating extremist messaging across digital platforms. However, their adaptability allows extremist groups to obscure meanings or modify designs to evade detection, presenting a significant challenge for content moderation systems (McKelvey et al. 2015; Milner 2014). Automated tools often fail to recognise emerging symbols or implicit variations effectively, the uncertainty of which can lead to over or under removal (Alexander, 2020; Morada, 2023).

#### **Recommendation:**

To address this issue, online platforms should develop and maintain a dynamic, globally comprehensive database of extremist symbols, logos, and branding. This repository should

integrate input from regional experts to ensure cultural and contextual accuracy, reflecting the diverse and evolving nature of symbols used by extreme right groups. Collaboration with civil society organisations, law enforcement, and academic institutions is essential to identify obscure and emerging symbols. Additionally, this database must be embedded into machine learning algorithms for automated content moderation, with periodic updates to account for newly identified symbols and their variations. The database should also form the foundation for knowledge-sharing initiatives with smaller platforms and independent researchers, as recommended in earlier policy proposals, to ensure consistency and scalability across the digital ecosystem.

**Impact:**

An updated and culturally informed database will enable platforms to detect and flag extremist symbols more effectively, mitigating the proliferation of harmful material. By proactively addressing both established and emerging symbols, platforms can disrupt extremist communication strategies and prevent the normalisation of their ideologies. Integration into machine learning algorithms will enhance the precision of automated moderation, reducing errors and improving public trust in moderation practices. The database's collaborative nature will help standardise moderation approaches across platforms, creating a unified and coordinated response to evolving extremist tactics.

**2b. Cross-Platform Collaboration**

**Finding:**

Extreme right groups exploit the fragmented nature of online platforms to amplify their reach and evade detection (Zannettou et al. (2018a); Nouri et al., 2020; McKelvey et al. (2015); Hawley, 2017; Greene, 2019; Doerr, 2021). As detailed in Chapter 1, decentralised strategies allow extremist content to migrate seamlessly between large networks and smaller, fringe platforms (Hawley, 2018; Taylor, 2020; Ohlheiser and Dewey, 2016; Powers, 2022; Lorenzo-Dus and Nouri, 2021). Exploiting platform-specific affordances such as hashtags and memes, these groups evade moderation systems that operate in isolation. The literature emphasises the critical need for collaborative approaches to counter these tactics, particularly given the centrality of symbolic communication in extreme right content dissemination.

**Recommendation:**

To support the implementation of the extremist symbol database proposed in Recommendation 2a, platforms should be encouraged to engage in collaborative data-sharing agreements across the digital ecosystem. This includes both major networks and smaller, decentralised platforms. Initiatives like the GIFCT and TAT, discussed in Chapter 1 of the literature review, should consider expanding their remit to share implicit visual motifs, coded symbols, and platform-specific dissemination strategies, targeting platforms outside of their mentorship programmes as not all platforms will wish to, or be able to, participate in the wider programme. This collaboration could build confidence and encourage platforms to recognise the benefits of not only sharing flagged content and metadata but also gaining these expertise, guidance, and knowledge-sharing opportunities to address any potential challenges in integrating and utilising the shared database. These organisations can use their status to coordinate with governments and civil society organisations to ensure that their databases are updated and culturally sensitive, to avoid the suppression of legitimate discourse (Llansó, 2020). By aiming to foster a culture of cooperation and shared expertise, GIFCT and TAT would further enhance their efforts to enhance overall content moderation practices, leading to a more unified and proactive approach to countering extremist material online.

**Impact:**

By fostering cross-platform collaboration, this recommendation will provide practical mechanisms for platforms to implement and enforce the symbol database outlined in Recommendation 2a. Smaller platforms, often exploited by extremist groups, will gain access to shared resources and expertise, enabling them to address extremist content more effectively. This unified framework will standardise moderation practices, ensuring that extremist content is tracked and mitigated holistically across the digital ecosystem. The increase in collaboration will strengthen trust and accountability among stakeholders, particularly towards platforms that generally do not participate in collaborative efforts, reinforcing the collective response to online extremism. By reducing gaps in moderation practices, this recommendation will help prevent the migration of extremist content and disrupt the decentralised strategies of extreme right networks. By linking symbol recognition efforts with cross-platform collaboration, this policy ensures that platforms can effectively

enforce moderation practices, even in a fragmented digital environment. The practical implementation of the symbol database through collaborative frameworks enhances its scalability, adaptability, and impact in combating the evolving tactics of extreme right groups.

### Policy Recommendation 3: Increasing Understanding of Extreme Right Groups

To counter extreme right propaganda effectively, platforms must enhance their understanding of the structural and behavioural characteristics of these groups. Central to this effort is the adoption of comprehensive definitions that encompass group-specific dynamics, particularly the in-group/out-group dichotomy, which underpins a significant proportion of their content and strategies. This core recommendation (3a) is supported by supplementary measures (3b and 3c) that address the practical implementation of these definitions through moderator training and alignment with international human rights principles. Together, these steps aim to create a nuanced, context-sensitive framework for moderating extreme right content, as informed by data analysis and literature review.

#### **3a. Comprehensive Definitions of Extremist Groups**

##### **Finding:**

As discussed in Recommendation 2a, symbols are a useful tool to capture extreme right imagery, with 65% of the data analysed containing at least one symbol. However, 30% was linked to identifiable extremist organisations, leaving most of the imagery without clearly identifiable symbols to capture. The analysis also revealed that over 80% of the dataset of images targets the in-group or the out-group in their messaging. The in-group/out-group dynamic is central to extreme right strategies (Berger 2015), fostering internal cohesion and vilifying perceived external threats (Abrams & Hogg, 2006; Tajfel, 1981; Tajfel & Turner, 2004). This theme transcends cultural and geographical boundaries, making it integral to understanding the identification of extreme right material. Indeed, how the in/group out/group theme is materialised in imagery can, however, involve cultural nuance (Rasoulkolamaki et al. 2023). Whilst definitions of extremism that focus on clear group identifiers like emblems or slogans are crucial, these alone may fail to account for broader symbolic and visual elements linked to these movements, leading to imagery being incorrectly

labelled as non-extremist or grey-zone imagery. Colours, composition, and iconography play crucial roles in eliciting emotional and ideological engagement, as noted by Wignell et al. (2017). The data analysis shows that certain visual styles can be attributed to certain extreme right groups, for example, whilst roughly half of the extreme right and Telegram groups had at least one instance of graphic imagery, TB and NSO had no graphic imagery across any of their multiple Telegram channels. These results show that certain elements of imagery may be useful beyond traditional symbols to identify the attribution of extreme right groups. To achieve this, further research into how these groups choose to represent their graphic nature, for example, whether the colour red or the icon of flames is uniquely tied to a specific group, could refine these attributions. This also applies to narrative subcategories such as 'Activism' (found in nearly 45% of the data) and 'Humour' (at nearly 17%). Understanding the modern iterations of characters such as Pepe the Frog, for example, could help identification (Yiannopoulos and Bokhari, 2016).

**Recommendation:**

Platforms should expand their definitions of extremism to include not only explicit, group-specific identifiers but also symbolic and associated imagery used by extreme right groups. These expanded definitions should incorporate a broader range of visual and contextual markers, such as colours, specific characters, or design elements that have become associated with these groups. By recognising the communicative function of such imagery in the in-group/out-group dynamic, content moderation systems can better address grey-zone content. The symbol database should be expanded collaboratively with experts and researchers, reflecting cultural, linguistic, and visual patterns tied to extreme right movements. Continuous updates should ensure adaptability to emerging trends and regional variations. By adding this to the symbol database suggested in Recommendation 2a, platforms could access information to increase the accuracy of their classifiers.

**Impact:**

By including a broader range of visual elements within the definition of extremism, platforms can enhance their ability to detect and address implicit messaging that may not be captured by clearer identifiers like symbols. This refined approach will reduce the risk of both under- and over-enforcement, striking a balance between protecting legitimate expression and

moderating extremist content. Additionally, incorporating these nuanced criteria will empower moderators and automated systems to evaluate grey zone content with greater contextual sensitivity. A globally coherent and comprehensive framework will position platforms to address the evolving strategies of extreme right movements effectively.

### **3b. Enhancing and Contextual Human Moderation**

#### **Finding:**

Extreme right imagery frequently employs subtle semiotics, such as cultural nostalgia, historical references, and gender role symbolism, which are challenging to detect without specialised training (Forchtner and Kølvråa 2017). Regional variations in extreme right imagery require localised expertise to ensure accurate evaluation and culturally sensitive moderation (Feamster, 2018; IRTL, 2022). The psychological toll on moderators working with harmful content is well-documented, yet platforms often fail to provide adequate support for their wellbeing (Newton, 2019; Steiger et al., 2021). These challenges are further amplified for third-party moderators, who often operate under different conditions and may lack access to the same resources and training as in-house teams (Financial Times, 2023a).

#### **Recommendation:**

Platforms and cross-platform organisations such as the GIFCT and TAT should implement comprehensive training programmes led by experts in cultural studies, extremist symbolism, and digital communication, extending these initiatives to third-party moderation teams. Training should equip all moderators with the knowledge to identify subtle and context-specific extremist content, including semiotics and the in-group/out-group dynamics central to extreme right strategies. Third-party moderators should receive equivalent resources, including access to mental health support, mandatory debriefing sessions, and secure work environments. Additionally, regional expertise networks should be established to guide on culturally specific content, with consistent integration between in-house and third-party teams to ensure uniform application of moderation standards, particularly during crises or events with heightened extremist activity. Certification pathways for moderators should ensure consistent competency across all teams and centralised training hubs (containing modules and case studies) and shared resource portals, such as that discussed in Recommendation 2a and 3a, should maintain online libraries of visual, linguistic, and

behavioural markers associated with extremist groups, updated regularly with contributions from regional experts. Mental health support frameworks should include partnerships with professionals to provide counselling and debriefing services, with equal access to third-party moderators. Flexible scheduling and workload rotations should help reduce burnout, particularly for moderators handling graphic or high-stress content. Integrated feedback loops should establish communication channels between in-house and third-party teams to share learnings, report emerging trends, and align practices, reinforced by regular cross-team meetings.

**Impact:**

Specialised and inclusive training will empower both internal and third-party moderators to identify implicit extremist content more accurately, reducing the misclassification of benign material and improving overall moderation precision. Equipping all moderators with equitable access to mental health and well-being resources will mitigate the psychological strain of these roles, fostering a sustainable and effective workforce. Regional expertise networks and structured feedback mechanisms will enhance cultural sensitivity and ensure consistent application of moderation standards. By addressing disparities in training and resources and embedding practical tools and processes, platforms can create a robust, scalable, and globally coherent human moderation framework that aligns with the expanded definitions proposed in 3a.

**3c. Regulatory Gaps and Human Rights Support**

**Finding:**

The transnational and decentralised nature of extreme right networks creates significant inconsistencies in content moderation and enforcement across jurisdictions (Tech Against Terrorism, 2023b). As discussed in Chapter 1 of the literature review, platforms often face challenges in aligning their moderation practices with international human rights obligations, complicating the delicate balance between curbing harmful content and protecting freedom of expression (Reed and Henschke, 2021). This challenge is particularly acute in the context of extreme right imagery, which is difficult to define and distinguish from grey material. Such ambiguity risks either over-moderation, which can infringe upon freedom of expression, or under-moderation, which may jeopardise public security. Platforms are thus under mounting

pressure to clearly understand their role in upholding human rights responsibilities (as a unique type of non-state actor with responsibilities of regulating public discourse) and to develop effective and transparent strategies for implementing these principles in their moderation practices. This is made to be more complicated when considering the other national, regional and international legislative obligations and responsibilities platforms are placed under.

**Recommendation:**

Platforms should be supported to align their policies with globally recognised human rights frameworks, such as the ICCPR (ICCPR, 1966), to ensure moderation practices respect both freedom of expression and the need to curb harmful content (Dvoskin, 2023; Benesch 2020; Sander 2020). Governments and international organisations must harmonise national legislation through cross-border regulatory frameworks, addressing enforcement gaps and legal ambiguities. To support platforms, structured guidance, such as training, toolkits, and expert consultations, should be provided to every level of platform staff (and tailored to the nature of any given team’s work) to facilitate the practical implementation of human rights principles across all platforms, including smaller ones which may lack the resources to implement this themselves (Weimann and Vellante, 2021). To help implement their own human rights obligations, Governments can provide funding and resources for smaller platforms within their remit to access this training. Meanwhile, organisations like GIFCT and TAT could facilitate more cross-border collaboration, sharing best practices, tools, and research on how to integrate human rights principles into moderation.

**Impact:**

Helping platforms align policies with globally recognised human rights principles would improve their ethical governance, offering a clearer framework for navigating the tension between curbing extreme right imagery content and protecting freedom of expression. This alignment helps platforms mitigate risks of both over- and under-moderation. It would also foster greater trust among users by demonstrating a commitment to transparency, accountability, and fairness in moderation practices. By embedding human rights principles into everyday operations through targeted training and tools, platforms, especially smaller

ones, can achieve consistency in enforcement without being overwhelmed by resource constraints.

This approach strengthens the legitimacy of moderation practices, reducing opportunities for extremist groups to exploit perceived governmental or institutional hypocrisy regarding human rights. By undermining these in/out-group dynamics, platforms can disrupt the narrative frameworks that extremists use to radicalise and mobilise supporters. In conjunction with the expanded definitions proposed in Recommendation 3a and the training initiatives outlined in 3b, this rights-based framework ensures a comprehensive, enforceable, and ethically sound moderation strategy that contributes to long-term security and social cohesion.

### Conclusion

In conclusion, from this chapter, many new findings have been discovered about the extreme right from the analysis of this dataset. These findings were uncovered as a means to answer each of the four research questions this research looked to answer. In RQ1, the extent to which imagery is widely adopted as a medium by extreme right groups was addressed. In discussion of the findings, it was uncovered that imagery is widely used by the extreme right. In RQ2, whether graphic imagery is prevalent within extreme right imagery was explored. As a result, it was found that graphic imagery is present, but not a defining feature of extreme right communication through imagery. In RQ3, it was considered whether there is an extreme right visual style within the imagery. According to the dataset and the extreme right imagery analysis framework, a visual style can be determined. This led to RQ4 where it was explored whether extreme right groups have differing visual styles, to which it was discovered that this is the case.

These findings contribute to the understanding of extremist use of the internet in several ways. These can be condensed into three key points. Firstly, despite the complexity around analysing extremism, with issues around definitions and categorisations of groups, it is possible to establish a broad understanding of these complex topic matters, in this case leading to the establishment of a defined visual style of the extreme right. Secondly, it is necessary to define new insights into what this style consists of and acknowledge its

limitations to ensure accurate policy implications. A defined visual style of the extreme right does not perfectly capture the visual style of all extreme right groups or one specific group; rather, it is an amalgamation of a particular set of extreme right groups. These distinctions are essential to make. Finally, whilst discussion around similarities and differences usually focuses on these between extremist groups, this research has highlighted that not only are there cross-group differences in the extreme right, but there are substantial intra-group differences across different chapters and Telegram groups belonging to a single group. These findings have significant implications for the academic and policy fields alike.

The insights from this research extend beyond theoretical contributions, offering practical applications for platforms and policymakers. Firstly, the research demonstrates the feasibility of defining and categorising extreme right visual styles, despite the complexities inherent in extremist content. Secondly, it underscores the importance of acknowledging the diversity within extreme right groups, ensuring that policy and moderation strategies remain nuanced and avoid over-generalisation. Lastly, the findings highlight the adaptability of extreme right groups, necessitating dynamic and context-aware approaches to counter their evolving tactics.

## Chapter 8: Conclusion

This thesis set out to understand the characteristics of extreme right visual communication to provide new knowledge and new policy recommendations to improve the accuracy of moderating extreme right content on online platforms. In doing so, this research has produced a novel dataset, a bespoke extreme right image analysis framework, policy recommendations and contributions of new knowledge to the field. At the time when this project was first being developed, the wider body of online terrorism literature was primarily populated with research that focused on Islamist extremist groups, leaving significant scope for more research to be conducted on the extreme right. Furthermore, whilst this thesis was in its early stages, social media companies were still in the wake of widespread extremist adoption of their platforms. As a result, this thesis aimed to fill the gap and contribute to the underrepresented body of extreme right literature in addition to creating recommendations for online platforms to improve their content moderation practices.

To achieve the aims of this research, the thesis took the following structure. Beginning with a literature review comprised of the following two chapters: Extreme Right (Chapter 2) and Challenges of Extreme Right Content Moderation (Chapter 3). These chapters introduce topics and define keywords that are core to the thesis. The literature review also offers context and framing as to why it is necessary to moderate extremist content, why it is necessary to support this practice through academic input and why the extreme right is a subset of extremism necessary to understand further. Using this context the Methodology chapters Framework Development (Chapter 4) and The Method (Chapter 5) describe the process of data collection and framework design. Following this is the Results (Chapter 6), which communicates the key characteristics of the results drawn from the dataset and framework described in the previous chapter. The Discussion (Chapter 6) provides more specific content to the results and directly applies them to each of the four research questions before offering three research and data-driven policy recommendations.

#### [A New Method for Understanding Extreme Right Imagery](#)

This thesis offers four areas of contribution, the methodological contribution, which presents a framework for understanding extreme right images, is central to the thesis. This framework was designed to improve understanding of the use of images by the extreme right groups. This is an important topic because as online platforms have improved their capabilities in removing extremist content, extremist collectives like the extreme right have found new ways to circumvent these strategies like communicating through imagery. Whilst there is research that analyses such imagery, a gap has been left in the literature of developing a framework which pinpoints the key characteristics used by the extreme right. This framework can be used to embed into tools for online platforms to use and in content moderation systems and policies more widely.

This framework was designed through a modified grounded theory. This process initially began by extracting key themes from the literature to establish a collection of categories and subcategories as a starting point. This was followed by three iterative rounds of coding, whereby, an increasingly growing sample of the image dataset was exposed to the framework to refine the presence and definitions of the categories and subcategories. From this process,

a structure was formed for the framework comprising five categories and 35 subcategories that can be seen as follows:

- 1) **Master Frames:** in-group neutral/positive, out-group negative, emphasising division, none, and unknown.
- 2) **Narratives:** anti-identity, anti-cultural Marxism, conspiracy theory, in-group victimisation, anti-establishment/global elite, global politics and military, call to action, activism, humour, and other.
- 3) **Shock Value:** yes and no.
- 4) **Image Format:** single image and image (undoctored photograph, cartoon/meme, doctored photograph, and screenshot) and image and text (undoctored image, graphic design, meme, screenshot).
- 5) **Symbolic Valence:** common extremist symbols, acronyms/ abbreviations, extremist organisations, slogans, weapon/gore, religious symbols, extremist clothing, key figures, news outlet, and other.

### Key Findings

This thesis is comprised of 8 chapters, each of these chapters were designed to progress through answering each of the research questions set out by this thesis. These research questions are as follows:

- RQ1: To what extent is imagery used as a medium of communication by extreme right groups?
- RQ2: To what extent is graphic imagery used in the communicative strategies within extreme right imagery?
- RQ3: Is there an extreme right visual style that is communicated and reinforced through the use of imagery?
- RQ4: Do extreme right groups have differing visual styles that are communicated and reinforced through the use of imagery?

Following the Introduction, Chapters 2 and 3 form the literature review for this thesis. Chapter 2 titled Extreme Right Online Propaganda and Content Moderation covers the context and

background of extreme right online propaganda, and why images are important to understand. The contents of this chapter provide clarity and understanding around the concept of extremism as understood by this thesis. Addressing this term led to the discussion of definitions and the proposal of a new definition. Inspired by the work of Berger (2018), a new definition was formed using the following: extremism is the belief that an in-group's success or survival can never be separated from the need for hostile action against an out-group. The hostile action must be part of the in-group's definition of success. Hostile acts can range from verbal attacks to diminishment. This definition was used then to create further understanding of the differences between extremism, violent extremism, online extremism, terrorism and radicalisation. Creating clarity around the term extremism is central to the core of this thesis and the research questions within its remit.

Chapter 3 expanded on the previous chapter, which is titled Existing Studies Analysing Extremist Images. Following a similar structure to the previous chapter, an initial focus was placed on understanding the context and definitions of the extreme right. This chapter characterises the widespread adoption of the internet by the extreme right. The use of the internet by the extreme right has led to the ongoing formation of new sub-collectives, such as the Alt-Right. This is an extreme right collective, born out of internet culture, which shares extreme right ideology embedded into humour and satire communicated on the internet. The extreme right and factions within it require online platforms to create a unique approach to moderating extreme right content due to its stark juxtaposition to other types of extremist group's images. To define the concept of online extremism, this thesis has understood it as "internet activism that is related to, engaged in, or perpetrated by groups or individuals that hold views considered to be doctrinally extremist".

Having conceptualised all the key concepts essential to this study, Chapters 4 and 5 entail this study's Methodology. These chapters outline the process of creating a bespoke extreme right image analysis framework frame within researching existing analytic frameworks. Through this, Baele, et al. (2020) were deemed to have created a relevant and useful framework for this thesis, which captures the visual style of ISIS. This provided a basis to create something in a similar vein to a different context, i.e., the extreme right. This was followed by a two-step process of creating the first framework iteration. The process incorporated analysing works

done which captured key themes found in influential extreme right individuals' manifestos, specifically looking at Anders Breivik. This included work by The Tony Blair Institute (2019) and Johnsen (2014) which captured the major narratives or master frames present within Breivik's manifesto. Following this, the second step followed a modified grounded approach. Using the master frames identified, increasingly large portions of the dataset were analysed to let themes in the data present themselves and form new categories and subcategories beyond the master frames formed in the first step. After three iterations of step 2, the finalised dataset was formed, and a novel framework was designed to analyse extreme right images.

Following the methodology was the Results (Chapter 6), which provided an overview of the data collected from the final study. The structure of this chapter followed the format of the framework itself. The framework consists of five key components: Master Frames, Narratives, Shock Value, Image Format, and Symbols. Each of these categories and their results were presented individually using a table to display and describe the results. The Master Frame results revealed significant use of in-group positive talk. The Narratives category revealed significant use of the activism narrative. Shock Value results revealed a minimal presence of shocking imagery. The Image Format category presented a high prevalence of single image formats, primarily through undoctored photography. Finally, Symbols were present in most of the dataset, with key figures and extremist organisation symbols being the most present categories in the dataset.

The last chapter before the Conclusion was Chapter 7, the Discussion. This chapter was organised into key sections, one section for each of the four research questions and the final section which outlines the application of the research question to policy recommendations for online platforms. Although the answers to the research questions could be found in the results generally, the Discussion chapter explicitly stated the answer to each of the four questions. As a result, this determined the extent to which imagery is a widely adopted medium by extreme right groups, whether graphic imagery is prevalent within extreme right imagery, whether there is an extreme right visual style within images and how extreme right groups have differing visual styles, as per the data set. These findings, along with the wider context found in the literature, were used to produce several policy recommendations which

can be presented to online platforms for their consideration and implementation to ultimately achieve the purpose of this thesis, which is to capture and characterise the visual style of the extreme right. These can be found in the following contributions section of this conclusion chapter.

### Contributions of Study to Future Applications

This study has produced four key contributions to the field of study. These contributions have been allocated to the following categories: theoretical, empirical, academic, and policy. Each of these will be addressed in combination with exploring their potential for future application.

The first contribution of the study is its methodological contribution. This pertains to the creation of a theoretical framework for extreme right image analysis. The development of this framework was done so in a robust and data-driven manner, allowing for future adoption and adaptation to be carried out with confidence in its design. The robust nature of this bespoke extreme right imagery analysis framework has been developed in a way that allows for future work to create further understandings of extreme right communication through visuals. Passing a new dataset through the framework is the most logical approach to adapting the framework. This could include analysing variables such as changing the extreme right groups, applying a similar future dataset to look at changes in visual styles over time or changing the image format to a GIF, Sticker, or thumbnail for example, to analyse the impact, if any, on the framework results from this. The possibilities surrounding future applications of this framework are limited only to the resources of the researchers who use it.

The second contribution offered by this thesis is its data contribution. To create the aforementioned theoretical framework for extreme right image analysis, this study collected a novel dataset. Said novel dataset is the result of collecting of over 25,000 extreme right images spread over six extreme right groups and 17 distinct Telegram groups. As a consequence of collecting a novel dataset, there is increased understanding created of extreme right groups and their use of images. As shown in the Results chapter, the dataset alone, before its categorisation and analysis, offers insight into extreme right group behaviour from the number of Telegram groups, the duration a group has had a presence on a platform and the image posting frequency. Given the significance of entering a new dataset into the

academic field, there is significant scope to utilise this data to explore alternative topics. Future use of this dataset could entail a comparison of image posting between different extremist collectives (e.g. extreme right imagery compared to Islamist extremist imagery). Alternatively, the qualities of the dataset could be used to facilitate a temporal comparison study, whereby a new dataset could be collected to compare differences in datasets collected with the same method years apart.

The third contribution this thesis offers is that of an empirical contribution. Building on the first two contributions, through creating a framework and collecting a novel dataset, the finding of this research contributes data-driven, novel, and empirical insights into how the extreme right communicates online and contributes to defining an extreme right visual style. As a result, this new insight and understanding, created around the existing body of extreme right and online extremism, offers a notable contribution. Similarly to the first two contributions, the scope for future work building on this offers a logical follow-up step. There is significant scope for future studies that expand the findings and explore new topics such as similarities and differences across different collectives and ideologies (e.g. QAnon or Incel).

The fourth and final contribution offered by this research is its policy contributions. This contribution refers to the process of using the findings drawn from this thesis to inform data-driven policy recommendations. These recommendations are designed to advise on content moderation approaches for a multitude of online platforms to consider. Said recommendations are based on results and findings extracted from the dataset, in addition to the context taken from the wider body of literature. The policy recommendations offered by this study offer a logical end point to the thesis. This is because they offer the intended subsequent recommendations for practice and under the current scope of this study, the author sees no further need to expand on these, instead, the focus should be on sharing these with stakeholders such as online platforms in a meaningful format to consider their implementation.

#### Limitations and Future Work

Whilst this thesis has been robust in design it is not without limitation, and where limitations are present there is scope for future work to develop.

The first limitation of this study is its source data. This study used a dataset of designated extreme right groups from Telegram. The result of this collection led to 88.7% of the dataset being sourced from one group (the Proud Boys), and several groups with very few images (namely AWD, NSO, and TB). The overrepresentation of one group and the underrepresentation of other groups despite having provided useful insight leave significant room to stress test the framework against a different dataset of images. In doing so this work would challenge the strength and relevance of each of the categories and subcategories that make the framework, and where applicable add and remove categories that are biased or tailored toward PB data.

The second limitation of this study is the limited use of metadata. This study has explored image counts per extremist group and Telegram group, how long a Telegram group has been active, and the average number of posts per day. This leaves significant room for future work to run a similar study whilst collecting additional metadata. Such data could include timestamps of when images were posted, user IDs relating to each image post, how images were interacted with (views/likes/reactions), the number of users in each Telegram group etc. Depending on the interest of subsequent research, collecting metadata offers future work the opportunity to gather additional insight into topics including the presence of specific categories regarding world events or the presence of highly active users in Telegram groups among other topics.

The third limitation of this study pertains to the construction of the framework. In the early stages of the framework's design Breivik's manifesto was used to inform the design and definitions related to the master frames and certain narratives. Whilst this approach successfully informed the design of the framework it may have caused unintentional consequences. Namely, due to Breivik's European nationality and focus (within his manifesto), it raises into question of whether the master frames and narratives may have a bias toward European views and away from U.S.-based framings and narratives. Whilst this bias has been partially offset by the iterative round of coding the impact of using the manifesto remains. Thereby, future work should firstly consider the presence of this bias and

secondly, consider ways to mitigate this (e.g. by adapting definitions and narratives or by testing the framework against a U.S.-based extreme right manifesto and suggesting changes).

The final limitation and suggestion for future research pertains to the age of the data. At the time of data collection, generative artificial intelligence (AI) generated imagery was not yet widely seen and used by civil society in the same ways that it is today. As such there were little to no instances of AI-generated imagery within the dataset, something that would likely not be the case if the dataset were collected more recently. Thereby, future research should consider both collecting a dataset of AI-generated imagery from extreme right sources and passing it through the framework and revising the framework based on the findings where applicable.

# Appendix I

## Final Codebook

### 1) Master Frames

- The 'in-group neutral/positive' subcategory is understood as images neutrally or positively containing the in-group (the majority of Ethnic Europeans (white in complexion) who are actively adhering to or desiring to return to a society sharing conservative values).
- The 'out-group negative' subcategory is images negatively containing the out-group (non-whites who are perceived as violent and dehumanized barbarians. Secondly, there are Cultural Marxists who are perceived as intolerant and hysterical ultra-liberals who suppress the truths in pursuit of ideological ends).
- The 'emphasising division' subcategory is understood as images which refer to neither group, but its contents explicitly strike differences between each group.
- The 'none' subcategory is understood as images that fall outside of these parameters and images with no individuals represented in them.
- The 'unknown' subcategory is understood as an image which does not fit any of the previous three categories.

### 2) Narratives

- The 'anti-identity' narrative, understood as negative implication toward an individual's race, religion, sex, gender, or appearance.
- The 'anti-cultural Marxism' narrative, understood as a collective of individuals who engage in 'identify politics' and 'political correctness', including the BLM movement and Antifa.
- The 'conspiracy theory' narrative, understood as referring to an explanation for an event or situation that invokes a conspiracy by sinister and powerful groups including the belief of accelerationism.
- The 'in-group victimisation' narrative, is understood as the belief that there is a global conspiracy to replace the white race and that the in-group is being disadvantaged e.g., through censorship.
- The 'anti-establishment/global elite' narrative is understood as advocating the complete overhaul of the political system and rejection of said system of individuals who represent the said system.
- The 'global politics and military' narrative, understood as depictions of armed forces and internationally recognisable political events and figures.
- The 'call to action' narrative, is understood as any encouragement of actions e.g., recruitment, violence, self-improvement.

- The 'activism' narrative is understood as any form of broad pro-in-group rhetoric e.g., in-group empowerment, promoting nationalism, and empowerment.
- The 'humour' narrative is understood as intending to provoke amusement for the in-group, commonly at the expense of the out-group through modes including satire.
- The 'other' narrative, which captures all narratives that do not fall into the previous five allocations, or where a series of images were posted and that image without the context of the previous iterations does not fit another subcategory.

### 3) Shock Value

- Images coded as 'Yes' indicate that an image triggers an instinctive reaction of disgust by the coder, the image displays corpses, or the image displays dismembered or beheaded bodies.
- Images coded as 'No' indicate that an image does not display any of the three qualities listed in the 'Yes' code.

### 4) Image Format

Single Image (an image without accompanying edited text):

- An 'undoctored photograph' refers to a photograph without any identifiable image manipulation.
- A 'cartoon/meme' refers to an image containing an object without accompanying text.
- A 'doctored photograph' refers to a photograph showcasing image manipulation.
- A 'screenshot' refers to images of the data displayed on the screen of a device without accompanying text, including images solely containing multiple screenshots.

Image & Text (an image with accompanying text edited in):

- A 'doctored image' refers to a photograph showcasing image manipulation with accompanying text.
- A 'graphic design' refers to heavily edited visual content with accompanying text.
- A 'meme' refers to an edited image created with humour intentions with accompanying text.
- A 'screenshot' refers to images of the data displayed on the screen of a device with accompanying text, including images solely containing multiple screenshots.

### 5) Symbolic Valence

- The 'common signs' subcategory is understood as those synonymous with the extreme right limited to the swastika, SS, SA, hammer & sickle, runes, black sun, Reichsadler, SS-Totenkopf skull, Celtic Cross, Hammer & sickle, Pepe the frog and the seig heil salute.

- The 'acronyms/abbreviations' subcategory is understood as codes which use combinations of letters and/or numbers that contain a hidden meaning, including those found in the ADL Hate on Display database.
- The 'extremist organisation' subcategory is understood as images including flags, emblems, and names depicting a known extremist organisation.
- The 'slogans' subcategory is understood as memorable phrases adopted by the extreme right, including those found in the ADL Hate on Display database.
- The 'extremist clothing' subcategory is understood as clothing items that appear to be designed by or affiliated with the in-group commonly containing the qualities of the previous 5 categories or clearly belonging to another extremist organisation.
- The 'weapon/ graphic subcategory captures the presence of weaponry within images, in addition to the presence of gruesome and graphic content including displays of blood or nudity.
- The 'religious symbol' subcategory captures any identifiable religious emblem and physical attire, excluding the swastika.
- The 'key figure' subcategory identifies any high-profile individual who holds significance and appears to be a person of interest to the in-group.
- The 'news outlet' subcategory captures images that repurpose material sourced from an identifiable news outlet.
- The 'no defined symbols' subcategory captures all images that do not fall into the previous allocations, including those not containing symbols.

## Bibliography

- Abrams, D., & Hogg, M. A.** (2006). *Social identifications: A social psychology of intergroup relations and group processes*. Routledge.
- Adami, A.** (2020). Visual analysis and right wing (populist) groups. *Partecipazione e conflitto*, 13(1), 315-336.
- ADL.** (2021). Proud Boys. Anti-Defamation League. Retrieved 22 May 2021, from <https://www.adl.org/proudboys>.
- Al-Tamimi, A.J.** (2023). The Islamic State and Its Treatment of 'Out-Groups': A Comparative Analysis. THE CENTER FOR JUSTICE & ACCOUNTABILITY. Retrieved: January 2024, from [https://cja.org/wp-content/uploads/2023/09/Al-Tamimi\\_theislamicstateandoutgroups.pdf](https://cja.org/wp-content/uploads/2023/09/Al-Tamimi_theislamicstateandoutgroups.pdf).
- Alava, S., Frau-Meigs, D., & Hassan, G.** (2017). *Youth and violent extremism on social media: mapping the research*. UNESCO publishing.
- Alexander, J.** (2020). YouTube reinstates monetization on Steven Crowder's channel after a year-long suspension for harassment. *The Verge*. Retrieved: February 25, 2021, from <https://www.theverge.com/2020/8/12/21365601/youtube-steven-crowder-monetization-reinstated-harassment-carlos-maza>.
- Amnesty International.** (2012). *Choice and prejudice: Discrimination against Muslims in Europe*. Amnesty International.
- Amnesty International.** (2021). Europe's fortress approach to migration and asylum: Attacks on the rights of refugees and migrants at the European Union's borders in Greece. Retrieved 3 June 2023, from <https://www.amnesty.org/en/wp-content/uploads/2021/05/EUR1438352021ENGLISH.pdf>.
- Anti-Defamation League** (2022), *Hate on Display: Hate Symbols Database*, Retrieved: January 25, 2021, from <https://www.adl.org/resources/hate-symbols/search?f%5B0%5D=topic%3A1703>.
- Richardson, J. E., & Wasser, M.** (2020). *A guide to online radical-right symbols, slogans and slurs*.
- Anti-Defamation League ADL.** (2016). "Pepe the Frog." Anti-Defamation League (ADL). Retrieved: January 25, 2021, from <https://www.adl.org/education/references/hate-symbols/pepe-the-frog> (Retrieved: January 25, 2021).

**Anti-Defamation League ADL.** (2018) Two Years Ago, They Marched in Charlottesville. Where Are They Now?. Retrieved 27 May 2021, from <https://www.adl.org/blog/two-years-ago-they-marched-in-charlottesville-where-are-they-now>.

**Australia Legislation,** Criminal Code Amendment (Sharing of Abhorrent Violent Material) Bill 2019.

**Baele, S. J., Boyd, K. A., & Coan, T. G.** (2020). Lethal images: Analyzing extremist visual propaganda from ISIS and beyond. *Journal of Global Security Studies*, 5(4), 634-657.

**Baele, S. J., Brace, L., & Coan, T. G.** (2020a). Uncovering the far-right online ecosystem: An analytical framework and research agenda. *Studies in Conflict & Terrorism*, 1-21.

**Batra, K.** (2019, June 5). Introduction to Computer Vision and Building Applications That Can See. Retrieved 22 May 2021, from [https://www.youtube.com/watch?v=L2B6\\_s3UvZA](https://www.youtube.com/watch?v=L2B6_s3UvZA).

**Baudrillard, J.** (1994). *Simulacra and simulation* / by Jean Baudrillard; translated by Sheila Faria Glaser. University of Michigan Press.

**BBC** (2014) 4chan murder pictures: David Kalac arrested in Oregon. BBC, 6 November. Retrieved 10 October 2019, from <https://www.bbc.co.uk/news/world-us-canada-29932087>

**BBC** (2020) Christchurch Mosque attack: Brenton Tarrant sentenced to life without parole. Retrieved 22 June 2023, from <https://www.bbc.co.uk/news/world-asia-53919624>.

**BBC** (2020a) Neo-Nazi group led by 13-year-old boy to be banned. Retrieved 22 June 2024, from <https://www.bbc.co.uk/news/uk-53392036>.

**BBC** (2022) Norway mass killer Anders Breivik ordered to stay in jail. Retrieved 22 June 2023, from <https://www.bbc.co.uk/news/world-europe-60219876>.

**Benesch, S.** (2020). But Facebook's not a country: How to interpret human rights law for social media companies. *JREG Bulletin*, 38, 86.

**Benson, T.** (2016). Inside the "Twitter for racists": Gab – the site where Milo Yiannopoulos goes to troll now. Retrieved 11 October 2019, from <https://goo.gl/Yqv4Ue>.

**Berger, J. M.** (2015) *Extremism*. Massachusetts: Massachusetts Institute of Technology.

**Berger, J. M.** (2018). *Extremism*. MIT Press.

**Berger, J.M.** (2017) "Extremist Construction of Identity: How Escalating Demands for Legitimacy Shape and Define In-Group and Out-Group Dynamics", *The International Centre for Counter-Terrorism – The Hague* 8(7).

**Berger, J.M.** (2021, May 12). The Out-Group in the In-Group. GNET Research. Retrieved 11 May 2022, from <https://gnet-research.org/2021/05/12/the-out-group-in-the-in-group/>.

- Berlet, C.** (2001). When hate went online. In Northeast Sociological Association Spring Conference in April (pp. 1-20).
- Bickert, M., and Fishman, B.** (2018) Hard questions: What are we doing to stay ahead of terrorists? Facebook Newsroom. Retrieved November 2019, from <https://newsroom.fb.com/news/2018/11/staying-ahead-of-terrorists/>.
- Bliuc, A. M., Betts, J., Vergani, M., Iqbal, M., & Dunn, K.** (2020). Growing power of online communities of the extreme-right: deriving strength, meaning, and direction from significant socio-political events' in real life'. International Centre for Counter-Terrorism.
- Blomberg, B., Engel, R. C., & Sawyer, R.** (2010). On the duration and sustainability of transnational terrorist organizations. *Journal of Conflict Resolution*, 54(2), 303-330.
- Botticher, A.** (2017) "Towards Academic Consensus Definitions of Radicalism and Extremism," *Perspectives on Terrorism* 11(4), pp. 73-78.
- Bowman-Grieve, L.** (2009). Exploring "Stormfront": A virtual community of the radical right. *Studies in conflict & terrorism*, 32(11), 989-1007.
- Bright, J.** (2018). Explaining the emergence of political fragmentation on social media: The role of ideology and extremism. *Journal of Computer-Mediated Communication*, 23(1), 17-33.
- Cai, W. and Landon, S.** (2019). "Attacks by White Extremists Are Growing. So Are Their Connections", *New York Times*, Retrieved 20 July 2020, from, <https://www.nytimes.com/interactive/2019/04/03/world/white-extremist-terrorism-christchurch.html>.
- Ware, J.** (2020). Testament to Murder: The Violent Far-Right's Increasing Use of Terrorist Manifestos. International Centre for Counter-Terrorism.
- Caiani, M.** (2024). Visual analysis and the contentious politics of the radical right. *Sociology Compass*, 18(9), e13267.
- Caiani, M., Della Porta, D., & Wagemann, C.** (2012). Mobilizing on the extreme right: Germany, Italy, and the United States. Oxford University Press, 6-9.
- Capellan, J. A.** (2015). Lone wolf terrorist or deranged shooter? A study of ideological active shooter events in the United States, 1970–2014. *Studies in Conflict & Terrorism*, 38(6), 395-413.

**Caren, N., Jowers, K., & Gaby, S.** (2012). A social movement online community: Stormfront and the white nationalist movement. In *Media, movements, and political change*. Emerald Group Publishing Limited.

**Carter, E.** (2018). Right-wing extremism/radicalism: Reconstructing the concept. *Journal of Political ideologies*, 23(2), 157-182.

**Castoriadis, C.** (1987). *The Imaginary Institution of Society*. Polity.

**CCDH.** (2024) Centre for Countering Digital Hate: Our Work. Retrieved 11 April 2024, from <https://counterhate.com/our-work/>.

**Charmaz, K.** (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage.

**Chermak, S., Freilich, J., & Suttmoeller, M.** (2013). The organizational dynamics of far-right hate groups in the United States: Comparing violent to nonviolent organizations. *Studies in Conflict & Terrorism*, 36(3), 193-218.

**Christmann, K.** (2012). *Preventing Religious Radicalisation and Violent Extremism. A Systematic Review of the Research Evidence*. London: Youth Justice Board.

**Coleman, P. T., & Bartoli, A.** (2003). *Addressing extremism. White Paper*. New York: The International Center for Cooperation and Conflict Resolution, Columbia University. [http://www.tc.columbia.edu/i/a/document/9386\\_WhitePaper\\_2\\_Extremism\\_030809.pdf](http://www.tc.columbia.edu/i/a/document/9386_WhitePaper_2_Extremism_030809.pdf).

**Colley, T., & Moore, M.** (2020). The challenges of studying 4chan and the Alt-Right: 'Come on in the water's fine'. *New Media & Society*, 1461444820948803.

**Color Of Change.** (2024). Stop hate for profit. Retrieved 11 March 2024, from <https://colorofchange.org/stop-hate-for-profit/>.

**Conseil Constitutionnel.** (2020). Decision No. 2020-801 DC. Conseil Constitutionnel. Retrieved 11 May 2023, from <https://www.conseil-constitutionnel.fr/decision/2020/2020801DC.htm>

**Conway, M.** (2017). Determining the role of the internet in violent extremism and terrorism: Six suggestions for progressing research. *Studies in Conflict & Terrorism*, 40(1), 77-98.

**Conway, M., & Courtney, M.** (2018). *Violent Extremism and terrorism Online in 2018. The Year in Review: Network of Excellence*.

**Conway, M., & Macdonald, S.** (2023). Introduction to Special Issue: The Practicalities and Complexities of (Regulating) Online Terrorist Content Moderation. *Studies in Conflict & Terrorism*, 1-4.

**Conway, M., Khawaja, M., Lakhani, S., Reffin, J., Robertson, A., & Weir, D.** (2020). *Disrupting Daesh: Measuring takedown of online terrorist material and its impacts. In Islamic State's Online Activity and Responses* (pp. 141-160). Routledge.

**Conway, M., Scrivens, R., & McNair, L.** (2019). Right-wing extremists' persistent online presence: History and contemporary trends.

**Cormode, G., & Krishnamurthy, B.** (2008). Key differences between Web 1.0 and Web 2.0. *First Monday*.

**Corrigall-Brown, C., & Wilkes, R.** (2014). Media exposure and the engaged citizen: How the media shape political participation. *The Social Science Journal*, 51(3), 408-421.

**Council of the European Union.** (2024) Fight against terrorism - Terrorist list. Retrieved 11 January 2024, from <https://www.consilium.europa.eu/en/policies/fight-against-terrorism/terrorist-list/>.

**Counter Extremism Project** (2017) Terrorists on Telegram. Counter Extremism Project. Retrieved 21 May 2019, from <https://www.counterextremism.com/terrorists-on-telegram>.

**Counter Extremism Project.** (2022). 'Extremist content online: White supremacist raising money on GiveSendGo to support campground to be used for training', *The Counter Extremism Project*. (2022) CEP, 12 September. Available from <https://www.counterextremism.com/press/extremist-content-online-white-supremacist-raising-money-givesendgo-support-campground-be>.

**Coutts, S.** (2017). Hate Goes Mainstream: Gavin McInnes and the Proud Boys. Retrieved 8 May 2019, from <https://rewirenewsgroup.com/2017/08/28/hate-goes-mainstream-gavin-mcinnis-proud-boys/>.

**Crawford, B.** (2020). The Proud Boys - ICSR. Retrieved 24 May 2021, from <https://icsr.info/2020/10/01/the-proud-boys/>.

**Crenshaw, M.** (1981) "The Causes of Terrorism", *Comparative Politics*.

**Cronin, A. K.** (2009). *How terrorism ends*. Princeton University Press.

**Dan, P.** (2022, May). The consequences of populism: The inevitable resurgence of antisemitism. In ASN Conference. New York: Columbia University. Retrieved 13 October 2028, from

<https://www.researchgate.net/publication360504548> *The consequences of populism The inevitable resurgence of antisemitism.*

**DeCook, J. R.** (2018). Memes and symbolic violence: proud boys and the use of memes for propaganda and the construction of collective identity. *Learning, Media and Technology*, 43(4), 485-504.

**Doerr, N.** (2021). The Visual Politics of the Alternative for Germany (AfD): Anti-Islam, Ethno-Nationalism, and Gendered Images. *Social Sciences*, 10(1), 20.

**Douek, E.** (2020). The rise of content cartels. Knight First Amendment Institute at Columbia University. <https://knightcolumbia.org/content/the-rise-of-content-cartels>.

**Douek, E.** (2021). Governing online speech: From "posts-as-trumps" to proportionality and probability. *Colum. L. Rev.*, 121, 759.

**Du, L., Ho, A. T., & Cong, R.** (2020). Perceptual hashing for image authentication: A survey. *Signal Processing: Image Communication*, 81, 115713.

**Duarte, N., Llanso, E., & Loup, A.** (2017). Mixed Messages?. Center for Democracy and Technology.

**Dvoskin, B.** (2023). Expert Governance of Online Speech. *Harv. Int'l LJ*, 64, 85.

**Eatwell, R.** (2003) The theories of the extreme right. In: P. Merkl and L. Weinberg (eds.)

**Engstrom, E., and Feamster, N.** (2017) The limits of filtering: A look at the functionality and shortcomings of content detection tools. Retrieved 19 November 2021, from via <https://perma.cc/UV5H-89SK>.

**European Commission.** (2022). Digital Services Act: Commission publishes guidelines on new obligations for online platforms. European Commission. Retrieved 7 March 2023, from [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_22\\_7109](https://ec.europa.eu/commission/presscorner/detail/en/ip_22_7109)

**European Digital Rights.** (2024). French Avia Law declared unconstitutional: What does this teach us at EU level? Retrieved 10 April 2024, from <https://edri.org/our-work/french-avia-law-declared-unconstitutional-what-does-this-teach-us-at-eu-level/>.

**European Platform of Regulatory Authorities** (2021), Online Hate in France - The 'Avia Law': The end of an Intensive Legislation Saga. Retrieved 22 May 2022, from [https://www.epra.org/news\\_items/online-hate-in-france-the-law-avia-an-intensive-legislative-saga-for-a-heavily-censored-law](https://www.epra.org/news_items/online-hate-in-france-the-law-avia-an-intensive-legislative-saga-for-a-heavily-censored-law).

**European Union** (2023) IRU. Retrieved 22 May 2022, from <https://www.europol.europa.eu/about-europol/european-counter-terrorism-centre-etc/eu-internet-referral-unit-eu-iru>.

**European Union.** (2024). Sanctions against terrorism: Council renews the EU Terrorist List and designates a new entity. European Commission of the European Union, Retrieved 26 September 2024, from. URL <https://www.consilium.europa.eu/en/press/press-releases/2024/07/26/sanctions-against-terrorism-council-renews-the-eu-terrorist-list-and-designates-a-new-entity/>.

**Europol** (2019) Europol and Telegram Take on Terrorist Propaganda Online. Europol Newsroom. Retrieved 18 September 2020, from <https://www.europol.europa.eu/newsroom/news/europol-and-telegram-take-terrorist-propaganda-online>.

**Farenet.** (2016) Monitoring discriminatory signs and symbols in European football, Retrieved 18 August 2020, from [https://www.farenet.org/wp-content/uploads/2016/10/Signs-and-Symbols-guide-for-European-football\\_2016-2.pdf](https://www.farenet.org/wp-content/uploads/2016/10/Signs-and-Symbols-guide-for-European-football_2016-2.pdf)

**Feamster, N.** (2018) Artificial intelligence and the future of online content moderation. Retrieved 20 February 2021, from <https://voxpath.eu/artificial-intelligence-and-the-future-of-online-content-moderation/>

**Financial Times.** (2023). "EU tells Elon Musk to hire more staff to moderate Twitter." Retrieved 20 April 2024, from <https://www.ft.com/content/20141fb1-d8f7-4c9e-a0d0-ded1ac8c7947>.

**Financial Times.** (2023a). "'You can't unsee it': the content moderators taking on Facebook." Retrieved 20 April 2024, from <https://www.ft.com/content/afeb56f2-9ba5-4103-890d-91291aea4caa>.

**Finlayson, A.** (2021). Neoliberalism, the alt-right and the intellectual dark web. *Theory, Culture & Society*, 38(6), 167-190.

**Fisher, A., Prucha, N., & Winterbotham, E.** (2019). Mapping the Jihadist Information Ecosystem. *Global Research Network on Terrorism and Technology*, 6.

**Floridi, L.** (2015): "Introduction"; in: Floridi, L. [Ed.] *The Onlife Manifesto: Being Human in a Hyperconnected Era*, London: Springer Open, pp. 1–7.

- Forchtner, B., & Kølvråa, C.** (2017). Extreme right images of radical authenticity: Multimodal aesthetics of history, nature, and gender roles in social media. *European Journal of Cultural and Political Sociology*, 4(3), 252-281.
- Francis, M.** (2012). What causes radicalisation. Main lines of Consensus in Recent Research. *Radicalisation Research*.
- Fulgoni, S., & Menis, S.** (2024). The Atomwaffen Division: the myth of evidence-based policy on the threat of far-right extremism. In *Preventing and Countering Violent Extremism and Radicalisation* (pp. 303-332). Policy Press.
- Galloway, B. and Scrivens, R.** (2018). Hidden Face: Hate Groups Online - Former's Perspective. *Vox Pol.* Retrieved 20 September 2020, from <https://voxpath.eu/hidden-face-hate-groups-online-formers-perspective/>.
- Garrett, R.** (2006). Protest in an information society: A review of literature on social movements and new ICTs. *Information, communication & society*, 9(02), 202-224.
- Gartenstein-Ross, D. & Blackman, M.** (2019) "Fluidity of the Fringes: Prior Extremist Involvement as a Radicalization Pathway," *Studies in Conflict & Terrorism*, DOI: 10.1080/1057610X.2018.1531545
- Gaudette, T., Scrivens, R., & Venkatesh, V.** (2022). The role of the internet in facilitating violent extremism: insights from former right-wing extremists. *Terrorism and Political Violence*, 34(7), 1339-1356.
- Gaudette, T., Scrivens, R., Davies, G., & Frank, R.** (2021). Upvoting extremism: Collective identity formation and the extreme right on Reddit. *New Media & Society*, 23(12), 3491-3508.
- German Federal Ministry of Justice and Consumer Protection**, Act To Improve Enforcement Of The Law In Social Networks (Network Enforcement Act) (Berlin: Germany, 2017). Available at: Roberts, S. T. (2017). Content moderation.
- Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken** [Network Enforcement Act], *Bundesgesetzblatt, Teil I [BGBl I]*, 3352 (2017).
- Gill, P.** (2007) 'A Multi-Dimensional Approach to Suicide Bombing', *International Journal of Conflict and Violence*, 1 (2), pp.142–159.
- Gill, P. & Corner, E.** (2015). Lone Actor Terrorist Use of the Internet and Behavioural.

**Gill, P., Corner, E., Conway, M., Thornton, A., Bloom, M., & Horgan, J.** (2017). Terrorist use of the Internet by the numbers: Quantifying behaviors, patterns, and processes. *Criminology & Public Policy*, 16(1), 99-117.

**Glitch Charity.** (2024). Our campaigns. Retrieved from <https://glitchcharity.co.uk/our-campaigns/>.

**Global Internet Forum to Counter Terrorism (GIFCT).** (2021). GIFCT Transparency Report 2021. Retrieved 20 April 2024, from <https://gifct.org/wp-content/uploads/2021/07/GIFCT-TransparencyReport2021.pdf>.

**Global Internet Forum to Counter Terrorism (GIFCT).** (2024). Content Incident Protocol. Retrieved 20 April 2024, from <https://gifct.org/content-incident-protocol/>.

**Global Internet Forum to Counter Terrorism (GIFCT).** (2024a). Homegrown Terrorism Definitions and Understanding Pathways. Retrieved 20 April 2024, from <https://gifct.org/hbdb/>.

**Global Internet Forum to Counter Terrorism (GIFCT).** (2024b). Membership. Retrieved 20 April 2024, from <https://gifct.org/membership/>.

**Global Internet Forum to Counter Terrorism (GIFCT).** (2024c). Definitional frameworks and principles to counter terrorism and violent extremism conducive to terrorism online. Retrieved 20 April 2024, from <https://def-frameworks.gifct.org/>.

**GNET Research.** (2024). About. Retrieved from <https://gnet-research.org/about/>.

**Greene, V. S.** (2019). "Deplorable" Satire: Alt-Right Memes, White Genocide Tweets, and Redpilling Normies. *Studies in American Humor*, 5(1), 31-69.

**Gross, A. G.** (2009). Toward a theory of verbal–visual interaction: The example of Lavoisier. *Rhetoric Society Quarterly*, 39(2), 147-169.

**Hafez, M., and Mullins, C.** (2015) "The Radicalization Puzzle: A Theoretical Synthesis of Empirical Approaches to Homegrown Extremism," *Studies in Conflict & Terrorism*, 38(11), 958-975, DOI: [10.1080/1057610X.2015.1051375](https://doi.org/10.1080/1057610X.2015.1051375).

**Halfmann, D., & Young, M.** (2010). War pictures: The grotesque as a mobilizing tactic. *Mobilization: An International Quarterly*, 15(1), 1-24.

**Hall, L.** (2021). Proud Boys member named Milkshake accused of shouting 'let's take the f\*\*\*ing Capitol!' is arrested. Retrieved 1 June 2021, from <https://www.independent.co.uk/news/world/americas/crime/capitol-riot-proud-boys-arrested-b1851562.html>.

- Halliday, M. A. K.** (1978). *Language as social semiotic: The social interpretation of language and meaning*. Hodder Education.
- Hassan, H., Farine, L., Kinnish, N., Mejía, D., & Tindale, C.** (2023). What is Extremism? Advancing Definition in Political Argumentation. *Topoi*, 42(2), 573-581.
- Hawley, G.** (2017). *Making sense of the alt-right*. Columbia University Press.
- Hawley, G.** (2018). *The alt-right: what everyone needs to know®*. Oxford University Press
- Helm, B., Scrivens., Holt, T., Chermak, S., and Frank, R.** (2024) Examining incel subculture on Reddit, *Journal of Crime and Justice*, 47:1, 27-45, DOI: [10.1080/0735648X.2022.2074867](https://doi.org/10.1080/0735648X.2022.2074867).
- Hendrycks, D., & Dietterich, T.** (2019). Benchmarking neural network robustness to common corruptions and perturbations. arXiv preprint arXiv:1903.12261.
- Henschke, A., & Reed, A.** (2021). Toward an ethical framework for countering extremist propaganda online. *Studies in Conflict & Terrorism*, 1-18.
- Hiremath, B. K., & Kenchakkanavar, A. Y.** (2016). An alteration of the web 1.0, web 2.0 and web 3.0: a comparative study. *Imperial Journal of Interdisciplinary Research*, 2(4), 705-710.
- HM Government.** (2011) *CONTEST: The United Kingdom's Strategy for Countering Terrorism*. Norwich: The Stationary Office.
- HM Government.** (2023) *CONTEST: The United Kingdom's Strategy for Countering Terrorism 2023*. Norwich: The Stationary Office.
- Hodge, E., & Hallgrímsdóttir, H. K.** (2021). Networks of hate: the alt-right, "troll culture", and the cultural geography of social movement spaces online. In *British Columbia's Borders in Globalization* (pp. 102-119). Routledge.
- Hoffman, B., Ware, J., & Shapiro, E.** (2020). Assessing the threat of incel violence. *Studies in Conflict & Terrorism*, 43(7), 565-587.
- Hope Not Hate.** (2021). *The Proud Boys Implode*, Retrieved 26 May 2021, from <https://www.hopenothate.org.uk/2018/12/02/proud-boys-implode/>.
- Hopkins, N.** (2017). Facebook moderators: a quick guide to their job and its challenges. *The Guardian*. Retrieved 21 December 2019, from <https://www.theguardian.com/news/2017/may/21/facebook-moderators-quick-guide-job-challenges>.
- Horgan, J.** (2009). *Walking Away from Terrorism: Accounts of Disengagement from Radical and Extremist Movements*. pp. 1-186. DOI: 10.4324/9780203874738.
- Hundeide, K.** (2003). Becoming a committed insider. *Culture & Psychology*, 9(2), 107-127.

- Independent Reviewer of Terrorism Legislation. (2022). Hazardous Convergence: Extreme Right Wing Terrorism [PDF document]. ISC. Retrieved 20 August 2023, from [https://isc.independent.gov.uk/wp-content/uploads/2022/07/E02710035-HCP-Extreme-Right-Wing-Terrorism\\_Accessible.pdf](https://isc.independent.gov.uk/wp-content/uploads/2022/07/E02710035-HCP-Extreme-Right-Wing-Terrorism_Accessible.pdf).
- Ingram, H.** (2014). "Three Traits of the Islamic State's Information Warfare." RUSI Journal 159 (6): 4–11.
- Ingram, H.** (2015). "The Strategic Logic of Islamic State Information Operations." Australian Journal of International Affairs 69 (6): 729–52.
- Ingram, H. J.** (2017) "The Strategic Logic of the "Linkage-Based" Approach to Combating Militant Islamist Propaganda: Conceptual and Empirical Foundations", The International Centre for Counter-Terrorism – The Hague 8, no. 6. DOI: <http://dx.doi.org/10.19165/2017.1.06>.
- ISD.** (2024). Memes: The Extreme Right-Wing. Retrieved from <https://www.isdglobal.org/explainers/memes-the-extreme-right-wing/>.
- Jewitt, C., Bezemer, J., & O'Halloran, K.** (2016). Introducing multimodality. Routledge.
- Joffe, H.** (2008). The power of visual material: Persuasion, emotion and identification. Diogenes, 55(1), 84-93.
- Johnsen, M. H.** (2014). A case study of Anders B. Breivik's intergroup conceptualisation. Contemporary Voices: St Andrews Journal of International Relations, 5(2).
- Johnson, B., & Feldman, M.** (2023). Siege: "Sheer Political Terror". Studies in Conflict & Terrorism, 1-21.
- Keatinge, T., Keen, F., & Izenman, K.** (2019). Fundraising for right-wing extremist movements: How they raise funds and how to counter it. The RUSI Journal, 164(2), 10-23.
- Counter Extremism Project** (2024) Blood & Honour (B&H). Counter Extremism Project. Retrieved 21 November 2024, from <https://www.counterextremism.com/supremacy/blood-honour-bh>.
- Kelly, M.** (2019). New Zealand ISPs are blocking sites that do not remove Christchurch shooting video. Retrieved 8 May 2021, from <https://www.theverge.com/2019/3/18/18271511/new-zealand-isp-blocking-sites-christchurch-shooting-video-removal-4chan-8chan-liveleak>.
- Kenes, B.** (2021). The Proud Boys: Chauvinist poster child of far-right extremism.

**Kimmel, M.** (2017). *Angry white men: American masculinity at the end of an era*. Hachette UK.

**Kingdon, A.** (2024). The Methodological and Ethical Implications of Using Imagery as the Primary Method of Researching Online Extremism. In *The World White Web: Uncovering the Hidden Meanings of Online Far-Right Propaganda* (pp. 49-67). Cham: Springer Nature Switzerland.

**Koetsier, J.** (2020) Report: Facebook makes 300,000 content moderation mistakes every day. *Forbes*, Retrieved 3 June 2020, from. URL

<https://www.forbes.com/sites/johnkoetsier/2020/06/09/300000-facebook-content-moderation-mistakes-daily-report-says/>.

**Kozyrkov, C.** (2018). What great data analysts do—and why every organization needs them. *Harvard Business Review*.

**Kress, G., & Van Leeuwen, T.** (2020). *Reading images: The grammar of visual design*. Routledge.

**Kriner, M., & Lewis, J.** (2021). Pride & prejudice: The violent evolution of the proud boys. *CTC sentinel*, 14(6), 26-38.

**Kutner, S.** (2020). *Swiping Right: The Allure of Hyper Masculinity and Cryptofascism for Men Who Join the Proud Boys*. International Centre for Counter-Terrorism. Retrieved 22 January 2021, from <https://www.istor.org/stable/pdf/resrep25259.pdf>.

**Kydd, A. H.** (2021). Decline, radicalization and the attack on the US Capitol. *Violence: An International Journal*, 2(1), 3-23.

**Lacan, J.** (1968). The Mirror-Phase as Formative of the Function of the I. *New Left Review*, 51, 71-77.

**LaFree, G.** (2018). Is Antifa a terrorist group?. *Society*, 55, 248-252. *Law and Technology*. Abingdon, Oxon: Routledge, pp. 35–53.

**Lewis, J. A.** (2005, January). The Internet and terrorism. In *Proceedings of the ASIL Annual Meeting* (Vol. 99, pp. 112-115). Cambridge University Press.

**Link, D., Hellingrath, B., & Ling, J.** (2016). A Human-is-the-Loop Approach for Semi-Automated Content Moderation. In *ISCRAM*.

**Llansó, E. J.** (2020). No amount of “AI” in content moderation will solve filtering’s prior-restraint problem. *Big Data & Society*, 7(1), 2053951720920686.

**Long, D.** (1990) *The Anatomy of Terrorism* (New York: The Free Press).

**Lorenzo-Dus, N., & Nouri, L.** (2021). The discourse of the US alt-right online—a case study of the Traditionalist Worker Party blog. *Critical Discourse Studies*, 18(4), 410-428.

**Lowe, D.** (2017). Prevent strategies: The problems associated in defining extremism: The case of the United Kingdom. *Studies in Conflict & Terrorism*, 40(11), 917-933.

**Macdonald, S., Rees, C., & Remove, J. S.** (2022). Impede, disrupt, redirect: Understanding & combating pro-Islamic State use of file-sharing platforms. Washington, D.C.: RESOLVE Network. Retrieved 20 Demeber 2022, from <https://doi.org/10.37805/ogrr2022.1>

**Mackinac Center for Public Policy.** (2024). The Overton Window. Mackinac Center. <https://www.mackinac.org/OvertonWindow>

**Malthaner, S.** (2017) “Radicalization: The Evolution of an Analytical Paradigm,” *European Journal of Sociology*. Cambridge University Press, 58(3), pp. 369–401. doi: 10.1017/S0003975617000182.

**Mattheis, A.** (2022) Atomwaffen Division and its Affiliates on Telegram: Variations, Practices, and Interconnections. Washington, D.C.: RESOLVE Network. Retrieved 27 January 2023, from <https://doi.org/10.37805/remve2022.1>.

**McBain, S.** (2020). “The rise of the Proud Boys.” *New Statesman*. October 9, 2020. Vol. 149, no. 5541, p. 26.

**McCauley, C., and Moskalenko, S.** (2011) *Friction: How Radicalization Happens to Them and US*. Philadelphia:University of Pennsylvania Press.

**McKelvey, F., Tiessen, M., & Simcoe, L.** (2015). A consensual hallucination no more? The Internet as simulation machine. *European Journal of Cultural Studies*, 18(4-5), 577-594.

**Meta.** (2022). Privacy Within Meta's Integrity Systems. Facebook. Retrieved 8 July 2023, from <https://about.fb.com/wp-content/uploads/2022/07/Privacy-Within-Metas-Integrity-Systems.pdf>

**Michael, G., & Minkenberg, M.** (2007). A continuum for responding to the extreme right: A comparison between the United States and Germany. *Studies in Conflict & Terrorism*, 30(12), 1109-1123.

**Miller, C.** (2018). McInnes, Molyneux, and 4chan: Investigating pathways to the alt-right. Retrieved 25 May 2021, from <https://www.splcenter.org/20180419/mcinnes-molyneux-and-4chan-investigating-pathways-alt-right>

**Milman, N., & Doerr, N.** (2021). Visual analysis. *Oxford handbook of political participation research*, 1-15.

- Milner, R. M.** (2014). Hacking the social: Internet memes, identity antagonism and the logic of lulz. *The Fiberculture Journal* 22.
- Milton, D.** (2016). "Communication Breakdown: Unraveling the Islamic State's Media Efforts." *Combatting Terrorism Center at West Point*. Accessed May 1, 2022.  
<https://ctc.usma.edu/posts/communication-breakdown-unraveling-the-islamic-states-media-efforts>
- Mina, A. X.** (2018). *Memes to Movements: How the World's Most Viral Media is Changing Social Protest and Power*. Boston: Beacon Press.
- Morada, N. M.** (2023). Hate Speech and Incitement in Myanmar before and after the February 2021 Coup. *Global Responsibility to Protect*, 15(2-3), 107-134.  
<https://doi.org/10.1163/1875984X-20230003>
- Motyl, M. and Ellington, G.** (2024). "The Unbearably High Cost of Cutting Trust & Safety Corners." *Tech Policy Press*. Retrieved 2 October 2024, from  
<https://www.techpolicy.press/the-unbearably-high-cost-of-cutting-trust-safety-corners/#>.
- Mudde, C.** (2004). The populist zeitgeist. *Government and opposition*, 39(4), 541-563.
- Mudde, C.** (2007) *Populist Radical Right Parties in Europe*. Cambridge: Cambridge.
- Mudde, C.** (2010). The populist radical right: A pathological normalcy. *West European Politics*, 33(6), 1167-1186.
- Mudde, C.** (2015). Populist radical right parties in Europe today. *Transformations of populism in Europe and the Americas: History and recent trends*, 295-307.
- Mudde, C.** (2019). *The far right today*. John Wiley & Sons.
- Müller, M. G., & Özcan, E.** (2007). The political iconography of Muhammad cartoons: Understanding cultural conflict and political action. *PS: Political Science & Politics*, 40(2), 287-291.
- Nagle, A.** (2017). *Kill all normies: Online culture wars from 4chan and Tumblr to Trump and the alt-right*. John Hunt Publishing.
- Neumann R P.** (2013) "The Trouble with Radicalization," *International Affairs* 89(4), pp. 875.
- Newhouse, A.** (2021). The threat is the network: The multi-node structure of neo-fascist accelerationism. *CTC Sentinel*, 14(5), 17-25.
- Newton, C.** (2019). Facebook moderators are in revolt over 'inhumane' working conditions. *The Verge*. Retrieved 8 January 2020, from

<https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona>.

**Norris, P.** (2002). The bridging and bonding role of online communities. *Harvard International Journal of Press/Politics*, 7(3), 3-13.

**Nouri, L., & Morgan, M.** (2023). Introducing StreetSnap: An Initiative to Transform the Capture, Recording & Intelligence Gathered about Hate Graffiti across the Bridgend County Borough.

**Nouri, L., Lorenzo-Dus, N., & Watkin, A. L.** (2019). Following the Whack-a-Mole: Britain First's Visual Strategy from Facebook to Gab.

**Nouri, L., Lorenzo-Dus, N., & Watkin, A. L.** (2020). Impacts of radical right groups' movements across social media platforms—A case study of changes to Britain first's visual strategy in its removal from Facebook to Gab. *Studies in Conflict & Terrorism*, 1-27.

**O'Halloran, K. L.** (2015). Multimodal digital humanities. *International handbook of semiotics*, 389-415.

**O'Halloran, K. L., Tan, S., Wignell, P., Bateman, J. A., Pham, D. S., Grossman, M., & Moere, A. V.** (2019). Interpreting text and image relations in violent extremist discourse: A mixed methods approach for big data analytics. *Terrorism and Political Violence*, 31(3), 454-474.

**Ohlheiser, A., & Dewey, C.** (2016). Hillary Clinton's Alt-Right speech, annotated. *The Washington Post*. Retrieved 8 October 2019, from <https://www.washingtonpost.com/news/the-fix/wp/2016/08/25/hillary-clintons-alt-right-speech-annotated/>.

**Pantucci, R.** (2011). What have we learned about lone wolves from Anders Behring Breivik?. *Perspectives on Terrorism*, 5(5/6), 27-42.

**Piazza, J. A.** (2017). The determinants of domestic right-wing terrorism in the USA: Economic grievance, societal change and political resentment. *Conflict management and peace science*, 34(1), 52-80.

**Piazza, J. A.** (2022). Fake news: The effects of social media disinformation on domestic terrorism. *Dynamics of Asymmetric Conflict*, 15(1), 55-77.

**Pirro, A. L.** (2023). Far right: The significance of an umbrella concept. *Nations and Nationalism*, 29(1), 101-112.

**Pitcavage, M.** (2019). Surveying the landscape of the American far right. *The George Washington University*.

**Powell, L.** (2016) "Counter-Productive Counter-Terrorism. How is the dysfunctional discourse of Prevent failing to restrain radicalisation?," *Perspectives on Terrorism* 11(4), pp. 73-77.

**Powers, B. M.** (2022). *DOMESTIC EXTREMISTS AND LEADERLESS RESISTANCE* (Doctoral dissertation, Monterey, CA; Naval Postgraduate School).

**Price, R.** (2017). Google's app store has banned Gab, a social network popular with the far-right, for 'hate speech'. <http://uk.businessinsider.com/google-app-store-gab-ban-hate-speech-2017-8>. (2017).

**Prisk, D.** (2017). The hyperreality of the Alt Right: how meme magic works to create a space for far right politics.

**Public Safety Canada.** (2021). Currently Listed Entities. Retrieved 8 March 2021, from - <https://www.publicsafety.gc.ca/cnt/ntnl-scrt/cntr-trrrsm/lstd-ntts/crrnt-lstd-ntts-en.aspx>.

**Quinn, B.** (2024). Senior Tories criticise No 10 plans to broaden extremism definition. *The Guardian*. Retrieved 1 April 2024, from <https://www.theguardian.com/politics/2024/mar/05/senior-tories-criticise-no-10-plans-to-broaden-extremism-definition>.

**Rapoport, D.** (2014). Terrorism. In Hawkesworth, M., & Kogan, M. (Eds.), *Encyclopedia of Government and Politics* (Chapter 69). Taylor & Francis. Retrieved 14 June 2023, from <https://www.taylorfrancis.com/books/edit/10.4324/9780203426159/encyclopedia-government-politics-mary-hawkesworth-maurice-kogan>.

**Rapoport, D. C.** (2004). The four waves of modern terrorism. *Attacking Terrorism: Elements of a Grand Strategy* (Washington, DC: Georgetown University Press 2004) p, 54, 3-11.

**Rasoulkolamaki, S., & Kaur, S.** (2022). The Representational Strategies of Lionization and Victimization in ISIS's Online Magazine, *Dabiq. Terrorism and Political Violence*, 1-20.

**Rasoulkolamaki, S., Kaur, S., Zhdanova, A., & Mat Isa, N. A. N.** (2023). In-group and out-group identity construction in extremist discourse: a critical multimodal approach. *Behavioral Sciences of Terrorism and Political Aggression*, 1-19.

**Ravndal, J. A., & Bjørge, T.** (2018). Investigating terrorism from the extreme right: A review of past and present research. *Perspectives on Terrorism*, 12(6), 5-22.

**Reed, A., & Henschke, A.** (2021). Who Should Regulate Extremist Content Online?. In *Counter-Terrorism, Ethics and Technology: Emerging Challenges at the Frontiers of Counter-Terrorism* (pp. 175-198). Cham: Springer International Publishing.

- Renard, T. and Rekawek, K.** (2024). The Base, , and the Basis for Listing Far-Right Terror Groups. ICCT research paper. URL <https://icct.nl/publication/base-and-basis-listing-far-right-terror-groups>.
- Roberts, S. T.** (2017). Content moderation. In: Schintler, L., McNeely, C. (eds) Encyclopedia of Big Data. Springer, Cham. [https://doi.org/10.1007/978-3-319-32001-4\\_44-1](https://doi.org/10.1007/978-3-319-32001-4_44-1).
- RUSI.** (2024). Far Right. Retrieved 22 February 2024, from <https://rusi.org/explore-our-research/projects/extremism-and-terrorism-programme#far-right>.
- Sabir, R.** (2016) Preventing And Countering Violent Extremism Through Civil, Political, And Human Rights. [online] Ohchr.org. Retrieved 8 January 2021, from Available at: [https://www.ohchr.org/Documents/Issues/RuleOfLaw/PCVE/Rizwaan\\_Sabir.pdf](https://www.ohchr.org/Documents/Issues/RuleOfLaw/PCVE/Rizwaan_Sabir.pdf).
- Sageman, M.** (2008) Leaderless Jihad: Terror Networks in the Twenty First Century. Philadelphia: University of Pennsylvania Press.
- Sander, B.** (2019). Freedom of expression in the age of online platforms: The promise and pitfalls of a human rights-based approach to content moderation. *Fordham Int'l LJ*, 43, 939.
- Sayogie, F., Farkhan, M., Julian, H. P., AL Hakim, H. S. F., & Wiralaksana, M. G.** (2023). Patriarchal Ideology, Andrew Tate, and Rumble's Podcasts. *3L: Southeast Asian Journal of English Language Studies*, 29(2).
- Schain, M. A.** (2006). The extreme-right and immigration policy-making: Measuring direct and indirect effects. *West European Politics*, 29(2), 270-289.
- Schmid, A.** (2013). The Definition of Terrorism. In *The Routledge Handbook of Terrorism Research*, ed. Alex Schmid, 38–98. Abingdon-on-Thames: Routledge.
- Schmid, A. P.** (2020). Discussion 1-Revisiting the wicked problem of defining terrorism. *Terrorism: Its Past, Present & Future Study A Special Issue to Commemorate CSTPV* at 25, 3.
- Schuurman, B.** (2019). Topics in terrorism research: reviewing trends and gaps, 2007-2016. *Critical Studies on Terrorism*, 12(3), 463-480.
- Scrivens, R., Davies, G., & Frank, R.** (2020). Measuring the evolution of radical right-wing posting behaviors online. *Deviant Behavior*, 41(2), 216-232.
- Scrivens, R., Osuna, A. I., Chermak, S. M., Whitney, M. A., & Frank, R.** (2023). Examining online indicators of extremism in violent right-wing extremist forums. *Studies in Conflict & Terrorism*, 46(11), 2149-2173.

**Shenkman, C., Thakur, D., & Llansó, E.** (2021). Do you see what I see? Capabilities and limits of automated multimedia content analysis. arXiv preprint arXiv:2201.11105.

**Shifman, L.** (2014). *Memes in Digital Culture*. Cambridge, MA: MIT Press.

**Sleeping Giants** [@slpng\_giants]. (2024). Twitter feed. Retrieved 18 February 2024, from [https://twitter.com/slpng\\_giants](https://twitter.com/slpng_giants).

**Southern Poverty Law Centre.** (2024). BLOOD & HONOUR. Retrieved 28 October 2024, from <https://www.splcenter.org/fighting-hate/extremist-files/group/blood-honour>.

**Speckhard, A., Ellenberg, M., Morton, J., & Ash, A.** (2021). Involuntary celibates' experiences of and grievance over sexual exclusion and the potential threat of violence among those active in an online incel forum. *Journal of Strategic Security*, 14(2), 89-121.

**Spence, R., Bifulco, A., Bradbury, P., Martellozzo, E., & DeMarco, J.** (2023). The psychological impacts of content moderation on content moderators: A qualitative study. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 17(4).

**SPLC.** (2021). "Proud Boys", Southern Poverty Law Center, 2019, <https://www.splcenter.org/fighting-hate/extremist-files/group/proud-boys>.

**Steiger, M., Bharucha, T. J., Venkatagiri, S., Riedl, M. J., & Lease, M.** (2021, May). The psychological well-being of content moderators: the emotional labor of commercial moderation and avenues for improving support. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1-14).

**Stephens, W., Sieckelinck, S., and Boutellier, H.** (2019) Preventing Violent Extremism: A Review of the Literature, *Studies in Conflict & Terrorism*, DOI: [10.1080/1057610X.2018.1543144](https://doi.org/10.1080/1057610X.2018.1543144).

**Suttmoeller, M. J., Chermak, S. M., & Freilich, J. D.** (2018). Is more violent better? The impact of group participation in violence on group longevity for far-right extremist groups. *Studies in Conflict & Terrorism*, 41(5), 365-387.

**Tajfel, H.** (1981). *Human groups and social categories: Studies in social psychology*.

**Tajfel, H., & Turner, J. C.** (2004). The social identity theory of intergroup behavior. In *Political psychology* (pp. 276-293). Psychology Press.

**Tan, S., O'Halloran, K. L., Wignell, P., Chai, K., & Lange, R.** (2018). A multimodal mixed methods approach for examining recontextualisation patterns of violent extremist images in online media. *Discourse, Context & Media*, 21, 18-35.

**Taylor, B.** (2020). Alt-right. In *Encyclopedia of critical whiteness studies in education* (pp. 15-22). Brill.

**Taylor, C.** (2004). *Modern Social Imaginaries*. Duke UP.

**Tech Against Terrorism.** (2022). *Transparency Report Year One: Terrorist Content Analytics Platform*. Retrieved 8 March 2021, from [https://www.techagainstterrorism.org/wp-content/uploads/2022/03/Tech-Against-Terrorism-TCAP-Report-March-2022\\_v6.pdf](https://www.techagainstterrorism.org/wp-content/uploads/2022/03/Tech-Against-Terrorism-TCAP-Report-March-2022_v6.pdf) pp. 20.

**Tech Against Terrorism.** (2023). *TCAP Insights: Patterns of Online Terrorist Exploitation*. Retrieved from <https://techagainstterrorism.org/news/patterns-of-online-terrorist-exploitation>

**Tech Against Terrorism.** (2023a). *Transparency Report Year Two: Terrorist Content Analytics Platform*. Retrieved 14 February 2024, from <https://tcap-website.s3.eu-west-2.amazonaws.com/about/Tech+Against+Terrorism+-+TCAP+Transparency+Report+2021-2022.pdf>.

**Tech Against Terrorism.** (2023b). *Tech Against Terrorism – The Online Regulation Series 3.0 – The Handbook*. Retrieved 14 February 2023, from <https://techagainstterrorism.org/hubfs/Online%20Regulation%20Series%203-0-%20Tech%20Against%20Terrorism.pdf>.

**Tony Blair Institute** (2019). *Narratives of hate: The spectrum of far-right worldviews in the UK*, Aug 2019. Retrieved 14 February 2021, from <https://www.institute.global/insights/geopolitics-and-security/narratives-hate-spectrum-far-right-worldviews-uk>.

**Topinka, R. J.** (2018). Politically incorrect participatory media: Racist nationalism on r/ImGoingToHellForThis. *New media & society*, 20(5), 2050-2069.

**Turner, J.C., and Reynolds, K.** (2012). Self-categorization theory. 10.4135/9781446249222.n46.

**UK Government.** (2024). *New definition of extremism 2024*. GOV.UK. Retrieved 1 April 2024, from <https://www.gov.uk/government/publications/new-definition-of-extremism-2024/new-definition-of-extremism-2024#:~:text=2.,The%20definition,footnote%205%5D%20of%20others%3B%20or>.

**United Nations General Assembly.** (1948). *Universal Declaration of Human Rights*. Retrieved 19 October 2023, from <https://www.un.org/en/universal-declaration-human-rights/>.

**United Nations General Assembly.** (1966). International Covenant on Civil and Political Rights (Treaty Series No. 999). Retrieved 14 October 2023, from <https://treaties.un.org/doc/Publication/UNTS/Volume%20999/volume-999-I-14668-English.pdf>.

**United States Attorney's Office.** (2020). Las Vegas Resident Who Discussed Setting Fire To A Synagogue With A White Supremacist Extremist Group Sentenced For Possession Of Bomb-Making Components, Retrieved 18 September 2024, from. URL <https://www.justice.gov/usao-nv/pr/las-vegas-resident-who-discussed-setting-fire-synagogue-white-supremacist-extremist-group>.

**Valentini, D., Lorusso, A. M. and Stephan, A.** (2020): "Onlife Extremism: Dynamic Integration of Digital and Physical Spaces in Radicalization," *Frontiers in Psychology*, 11(March), p. 12.

**Van Der Vegt, I., Gill, P., Macdonald, S., & Kleinberg, B.** (2019). Shedding light on terrorist and extremist content removal.

**Van Prooijen, J. W., & Kuijper, S. M.** (2020). A comparison of extreme religious and political ideologies: Similar worldviews but different grievances. *Personality and Individual Differences*, 159, 109888.

**Vittori, J.** (2009). All struggles must end: The longevity of terrorist groups. *Contemporary Security Policy*, 30(3), 444-466.

**Ware, J.** (2019). Fighting back: The Atomwaffen Division, countering violent extremism, and the evolving crackdown on far-right terrorism in America. *Journal for Deradicalization*, (25), 74-116.

**Watkin, A. L.** (2023). Developing a responsive regulatory approach to online terrorist content on tech platforms. *Studies in Conflict & Terrorism*, 1-22.

**Weimann, G.** (2004). *Cyberterrorism: How real is the threat?* (Vol. 31). United States Institute of Peace.

**Weimann, G.** (2006). *Terror on the Internet: The new arena, the new challenges*. US Institute of Peace Press.

**Weimann, G., & Vellante, A.** (2021). The Dead Drops of Online Terrorism. *Perspectives on Terrorism*, 15(4), 39-53.

**Whittaker, J. J.** (2022). *Online radicalisation: the use of the internet by Islamic State terrorists in the US (2012-2018)*(Doctoral dissertation, Leiden University).

- Widdicombe, L.** (2013). The Bad-Boy Brand. Retrieved 25 May 2021, from <https://www.newyorker.com/magazine/2013/04/08/the-bad-boy-brand>.
- Wignell, P., Tan, S., O'Halloran, K. L., & Lange, R.** (2017). A Mixed Methods Empirical Examination of Changes in Emphasis and Style in the Extremist Magazines *Dabiq* and *Rumiyah*. *Perspectives on Terrorism*, 11(2), 2–20.
- Windisch, S., & Simi, P.** (2022). More than a Joke: White Supremacist Humor as a Daily Form of Resistance. *Deviant Behavior*, 1-17.
- Winter, C., Neumann, P., Meleagrou-Hitchens, A., Ranstorp, M., Vidino, L., & Fürst, J.** (2020). Online extremism: Research trends in internet activism, radicalization, and counter-strategies. *International Journal of Conflict and Violence*, 14(2), 1-20. doi: 10.4119/ijcv-3809
- Witt, T.** (2020). 'If i cannot have it, i will do everything i can to destroy it.'the canonization of Elliot Rodger:'Incel' masculinities, secular sainthood, and justifications of ideological violence. *Social Identities*, 26(5), 675-689.
- X Company.** (2024). Rules Enforcement. Retrieved from <https://transparency.x.com/en/reports/rules-enforcement.html#2021-jul-dec>.
- Yiannopoulos, M. and Bokhari, A.** (2016) An Establishment Conservative's Guide To The Alt-Right. [online] Breitbart. Retrieved 9 May 2021, from <https://www.breitbart.com/tech/2016/03/29/an-establishment-conservatives-guide-to-the-alt-right/>.  
youtube.com/watch?v=L2B6\_s3UvZA.
- Yudkowsky, E.** (2008). Artificial intelligence as a positive and negative factor in global risk. *Global catastrophic risks*, 1(303), 184.
- Zannettou, S., Bradlyn, B., De Cristofaro, E., Kwak, H., Sirivianos, M., Stringini, G., & Blackburn, J.** (2018). What is gab: A bastion of free speech or an alt-right echo chamber. In *Companion Proceedings of the The Web Conference 2018*(pp. 1007-1014).
- Zannettou, S., Caulfield, T., Blackburn, J., De Cristofaro, E., Sirivianos, M., Stringhini, G., & Zittrain, J. L.** (2019). Three eras of digital governance. Available at SSRN 3458435.