EQUAL FAITH EUROPE

EQUALITY. DIGNITY. JUSTICE.

The Enforcement Gap

ONLINE ISLAMOPHOBIA AND THE FAILURE OF DIGITAL GOVERNANCE IN EUROPE (2020–2025)

Goran Kolouch

Independent Researcher & Founder, Initiative EqualFaith Europe

An EqualFaith Europe Research Report

Full Draft

EQUALITY · DIGNITY · JUSTICE

Table of Contents

Executive Briefing

The central finding of this report is not that Europe lacks laws against anti-Muslim discrimination. It is that Europe's existing legal architecture — anti-discrimination law, hate-speech law, and now the Digital Services Act — is not reaching the people it was built to protect. The problem is not an absence of rules. It is a persistent, structural enforcement gap between what the law promises and what platforms, regulators, and institutions actually deliver.

This report traces that enforcement gap across three domains — the scale of online anti-Muslim hate, the behaviour of the two platforms most central to its spread, and the European Union's own regulatory response — and argues that all three point to the same underlying failure: systems designed to catch harm are not being resourced, audited, or held accountable in ways that make them work in practice.

Key findings

- **Discrimination against Muslims in Europe is rising, not stabilising.** FRA's EU-wide survey found 47% of Muslims reported racial discrimination in the preceding five years — up from 39% in 2016. National documentation is often more granular and more alarming: Austria's Dokustelle recorded a year-on-year increase of over 100% in victim-reported anti-Muslim cases between 2022 and 2023, two-thirds of them online.
- **Online hostility spikes are large, fast, and independently corroborated.** Three methodologically distinct trackers — keyword-based platform monitoring, manual cross-platform review, and automated AI toxicity classification — independently recorded a sharp escalation in anti-Muslim online content following the October 2023 escalation of the Israel-Gaza conflict, with increases of over 400% on individual platforms within 48 hours.
- **The clearest documented pathway from online falsehood to offline violence in this report's evidence base is the July 2024 Southport case**, in which a false claim about a Muslim attacker, amplified to hundreds of millions of views, preceded a direct attack on a mosque and six days of national rioting — with Meta's own Oversight Board later ruling that specific content should have been removed and was not.
- **Platform self-reported enforcement data confirms the gap independently.** X's own statutory DSA disclosures show a sub-1% appeal rate against content decisions; Meta's transparency reporting does not isolate "illegal hate speech" as a distinct category at all, making it structurally impossible to verify enforcement performance against religious discrimination specifically using the platforms' own published data.
- **The Digital Services Act represents a genuine shift in regulatory design — from moderating individual posts to assessing systemic risk** — but enforcement to date has addressed transparency and procedural failures rather than the substantive question of algorithmic amplification of hate content. The Commission's own findings that both X and Meta have obstructed researcher data access are, this report argues, both a violation in their own right and the reason the amplification question remains empirically unresolved.
- **The enforcement architecture built to close this gap — trusted flaggers (Article 22), researcher access (Article 40(12)), systemic risk assessment (Article 34) — itself shows structural weaknesses** that limit its practical value to religious-minority civil society specifically, independent of whether individual platforms comply.

Principal legal conclusion

The DSA's shift from content-level to systemic-risk governance is legally significant and, on the evidence reviewed, under-realised. Systemic risk assessment under Article 34 explicitly includes non-discrimination under Article 21 of the EU Charter of Fundamental Rights — yet no dedicated Commission guidance exists for religious-minority risk comparable to that already issued for electoral integrity or minor protection. This is not a gap in the law's scope; it is a gap in its operationalisation.

Policy implications

Closing the enforcement gap requires action at three levels: platform accountability (enforcing existing transparency and risk-assessment obligations with the same rigour applied to advertising and minor-protection violations), regulatory design (extending guidance and disaggregated reporting requirements to religious-minority risk specifically), and civil society capacity (ensuring organisations like EqualFaith Europe can access the data and flagging mechanisms the DSA already envisions for them).

Recommendations

Full recommendations, prioritised and linked to supporting evidence, appear in Chapter 13. In summary: (1) reform Article 22 trusted-flagger designation to require funding and affiliation disclosure; (2) expand Article 40(12) civil-society research access on the same legal basis already accepted for electoral-integrity research; (3) issue dedicated Commission guidance on religious-minority systemic risk; (4) mandate disaggregated hate-speech reporting by protected characteristic in platform transparency disclosures; (5) establish structured post-2025 monitoring of Meta's relaxed content-moderation policy specifically for its effect on anti-Muslim content.

Methodology

This report synthesises three categories of evidence, selected and weighted according to the following approach.

Source selection. Primary sources — European Commission enforcement decisions, FRA and ODIHR survey and hate-crime data, Meta Oversight Board rulings, and platforms' own statutory DSA transparency disclosures — are treated as the strongest evidentiary tier because they carry formal institutional or legal standing and, in the case of enforcement decisions, have been tested through an adversarial process (the right of reply exercised by both X and Meta). Peer-reviewed academic research is treated as the second tier, selected for direct relevance to online anti-Muslim hate, algorithmic amplification, or platform governance, and reviewed for methodology, sample size, and stated limitations rather than simply cited for its conclusions. Civil-society and NGO research (CCDH, ISD, the Online Hate Prevention Institute, national documentation centres) is treated as a third tier: methodologically transparent and, where cross-corroborated by other tiers, treated as reliable, but flagged

where its advocacy orientation may shape framing (see Chapter 11, Evidence Quality Assessment).

Case study selection. Three case studies were selected for their evidentiary distinctiveness rather than to build a representative sample of all anti-Muslim incidents in the period. The 2024 Southport case was selected because it offers the clearest documented causal chain — from a specific false claim, through measurable platform amplification, to formal institutional findings of moderation failure, to documented offline violence — of any incident identified in the research process. The Charlie Hebdo case was selected specifically because it does *not* fit that pattern: it involves institutional press speech rather than coordinated hostile amplification, and is included to test and qualify this report's central thesis rather than simply to illustrate it. The 2024 European Parliament and national election cycles were selected because they document anti-Muslim narrative content being produced by, rather than merely tolerated by, formal political actors — a structurally distinct mechanism from platform under-moderation.

Evidence limitations. This report's account of algorithmic amplification is, by necessity, correlational rather than causal, because platform-internal recommender-system data was not accessible during the research process — a limitation this report treats not as an incidental gap but as itself a central finding, given that the Commission's own enforcement decisions confirm both platforms studied have restricted the researcher access that would resolve this limitation. Cross-country hate-crime comparison is constrained by substantial, ODIHR-acknowledged variance in national reporting completeness. Where a claim could not be independently verified during this report's research process, it has been removed rather than retained on the basis of a single secondary citation; two such instances are documented in the bibliography's verification notes.

Chapter 1: Introduction — The Central Argument

Europe does not lack law to protect Muslims from discrimination. The EU Charter of Fundamental Rights prohibits discrimination on grounds of religion or belief. The Racial Equality and Employment Equality Directives create binding obligations across member states. The Digital Services Act — the most significant platform-governance instrument enacted anywhere to date — requires the largest online platforms to assess and mitigate systemic risks to fundamental rights, non-discrimination explicitly included. On paper, the architecture is comprehensive.

And yet the evidence assembled in this report shows discrimination against Muslims rising, not falling; online hostility spiking sharply and repeatedly; and a documented, causally traceable case in which platform failure preceded an attack on a place of worship and six days of national unrest.

The question this report asks is not whether the law exists. It is why the law is not reaching the people it was written for. The answer, developed across the chapters that follow and argued directly in Chapter 6, is an **enforcement gap**: a persistent, structural distance between the legal protections that exist in principle and their practical operation — in platform content-moderation systems, in regulatory guidance, and in the civil-society infrastructure needed to make enforcement mechanisms usable in practice.

This is a different diagnosis than either of the two explanations that dominate public debate. It is not, primarily, a story about insufficient law — Europe has more relevant legal instruments now than at any prior point, and the DSA in particular represents a genuine expansion of platform obligation. Nor is it, primarily, a story about individual bad actors — while specific incidents (a false claim, a hateful post, a discriminatory landlord) are the visible surface of the problem, this report's evidence shows those incidents recurring in a *pattern* traceable to systemic features of platform design, regulatory sequencing, and institutional capacity, not to isolated failures.

Every chapter in this report is organised around testing and substantiating this argument. Chapters 3 through 5 establish the scale and mechanisms of the problem — what is actually happening, on which platforms, with what documented consequences. Chapter 6 names and analyses the enforcement gap directly as this report's intellectual centrepiece. Chapter 7 examines the Digital Services Act specifically as the instrument best positioned to close that gap, and the specific ways its implementation to date has fallen short. The remaining chapters test this argument against the wider academic and institutional evidence base, cross-country variation, and legitimate counterarguments, before concluding with recommendations directly tied to the gaps identified.

Chapter 2: Conceptual and Legal Definitions

Definitional precision matters in this field for a reason that is not merely academic: as Chapter 8 details, imprecise definitions measurably affect automated content-moderation accuracy, and definitional disagreement affects which legal protections apply to which conduct. This chapter sets out the working definitions used throughout, and flags the points of genuine, unresolved disagreement in the literature and in EU policy.

2.1 Islamophobia

No single legally binding definition of Islamophobia exists in EU or international law, and this absence is itself analytically significant — Chapter 6 argues it is one structural component of the enforcement gap. Three formulations dominate the literature:

- **The Runnymede Trust (1997, updated 2017)** originated the term's contemporary usage in British policy discourse, characterising Islamophobia as a form of anti-Muslim racism.
- **The All-Party Parliamentary Group (APPG) on British Muslims (2018)**, following a two-year inquiry, proposed a working definition holding that Islamophobia is grounded in racism and functions as a type of racism directed at expressions — real or perceived — of Muslim identity. This definition has been adopted by numerous UK local authorities and institutions, but has also drawn sustained criticism, including from Muslim organisations such as FOSIS, on the grounds that framing Islamophobia primarily as racism risks obscuring its distinctly religious dimension and may complicate its relationship to protections under instruments like the UK Equality Act 2010, which defines race by colour, nationality, and ethnic or national origin rather than religion.
- **The UN Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance** has described Islamophobia as baseless hostility and fear toward Islam and, by extension, toward Muslims generally, extending to the discriminatory treatment, prejudice, and social and political exclusion that results from that hostility.

This report treats Islamophobia as an umbrella term encompassing both racialised hostility toward Muslims and hostility grounded specifically in religious identity or practice, consistent with the UN Special Rapporteur formulation, while noting the APPG/Runnymede racialisation debate where it affects legal classification — a distinction with practical consequences under EU law, since the Racial Equality Directive and religion-specific protections are not coextensive.

2.2 Anti-Muslim Hatred / Anti-Muslim Racism

FRA and the European Commission's anti-Muslim hatred database (2012–2024) use "anti-Muslim hatred" and "anti-Muslim racism" largely interchangeably. FRA's own convening on the topic has identified the continued absence of a single agreed EU definition as a direct obstacle to comprehensive equality data collection — the first of several instances, catalogued fully in Chapter 6, where definitional ambiguity translates directly into measurement and enforcement weakness.

2.3 Religious Discrimination

Distinct from Islamophobia as a social and political phenomenon, religious discrimination is the narrower legal category of unequal treatment on grounds of religion or belief, protected in employment contexts under EU Directive 2000/78/EC. Not all Islamophobic conduct meets the legal threshold for actionable religious discrimination, and not all religious discrimination against Muslims is motivated by the racialised hostility captured in the APPG/Runnymede definitions. The categories overlap substantially but are not coextensive — a distinction that matters directly for which legal remedy, if any, is available to a given victim.

2.4 Online Hate Speech and Digital Incitement

For EU purposes, this report follows the definition used in the EU Code of Conduct on Countering Illegal Hate Speech Online (2016, now integrated into the DSA's co-regulatory framework) and Council Framework Decision 2008/913/JHA: illegal hate speech is the public incitement to violence or hatred directed against a group or member of a group defined by race, colour, religion, descent, or national or ethnic origin. Digital incitement is treated as a narrower subcategory — online content that calls for or provokes violent or discriminatory action — distinct from hostile but non-incitement expression, which may still cause serious harm without meeting the criminal-law incitement threshold. This distinction is legally consequential: it marks the boundary of what the DSA's illegal-content provisions reach, as against content that is harmful but lawful, addressed instead (if at all) through the DSA's broader systemic-risk framework discussed in Chapter 7.

2.5 Algorithmic Amplification

Algorithmic amplification refers to the effect by which recommender systems and engagement-optimised ranking increase the reach, visibility, or virality of content — including hateful or discriminatory content — beyond what its initial organic distribution would achieve. This is not a peripheral technical detail; it is, as Chapter 7 argues, the central object of DSA systemic-risk regulation, because it is the mechanism by which an individual hateful post becomes a foreseeable, structural risk rather than an isolated incident.

2.6 Coordinated Harassment and Platform Governance

Coordinated harassment describes organised campaigns — by networked individuals, inauthentic accounts, or hybrid human-automated activity — that target individuals or communities with hostile content at scale, often exploiting hashtag or trending mechanisms. Platform governance, as used throughout this report, refers to the combination of platform terms of service, content-moderation systems, transparency reporting, and — since 2024 — binding external regulatory obligations under the DSA that together determine how hateful content is identified, ranked, and, in principle, removed or down-ranked. The distance between this formal architecture and its practical operation is precisely the enforcement gap this report is structured to demonstrate.

Chapter 3: The Scale and Trajectory of Online Anti-Muslim Hate (2020–2025)

3.1 The trend beneath the spikes

Public attention to online Islamophobia tends to follow discrete trigger events — a terror attack, a war, a riot. This chapter's evidence shows why that framing, while intuitive, understates the problem: the trigger events are real and consequential, but they sit on top of a baseline that was already rising before any single event, and the mechanisms that produce the spikes are structural rather than incidental.

FRA's "Being Muslim in the EU" survey — fielded 2021–2022 across 13 member states, with 9,604 respondents — found 47% of Muslim respondents had experienced racial discrimination in the preceding five years, up from 39% in a comparable 2016 survey. This baseline increase, which predates the acute events analysed later in this chapter, is the first evidence against a purely event-driven account of the problem: something was already deteriorating before 7 October 2023 or the Southport attack gave that deterioration a visible flashpoint.

3.2 The October 2023 inflection point: convergent evidence from three independent methodologies

The escalation of the Israel-Gaza conflict from 7 October 2023 produced the most heavily documented and quantified spike in this report's period of study. What makes this spike analytically significant — beyond its magnitude — is that it was independently measured by three methodologically distinct approaches, each with different susceptibility to bias or artifact, all converging on the same conclusion.

Keyword-based platform monitoring. The Institute for Strategic Dialogue (ISD) recorded a 422% increase in discriminatory posts targeting Muslims on X between 7–8 October 2023 compared with the preceding two days, and a 250% increase across the following week relative to the prior week. ISD separately documented a 43-fold increase in anti-Muslim comments on YouTube in the same period.

Manual cross-platform review. The Online Hate Prevention Institute, monitoring ten platforms (Facebook, Instagram, TikTok, X, YouTube, Telegram, LinkedIn, Gab, Reddit, and BitChute) using a matched 160-hour sample between 27 October 2023 and 8 February 2024, found anti-Muslim hate disproportionately concentrated on X relative to the other nine platforms sampled.

Automated AI toxicity classification. The EU-funded European Observatory of Online Hate (EOOH), using an explainable-AI classifier applied to close to eight million messages, found online toxicity increased more than 30% between January and September 2023, with Islam-related content the second-most toxic topic category measured (average score 0.19 on EOOH's 0–1 scale) — behind only Judaism-related content (0.28), and ahead of Ukraine- and refugee-related content.

Why the convergence matters. Each of these three methods has a different failure mode. Keyword tracking can be skewed by which terms researchers choose to monitor. Manual review is limited by sample size and reviewer judgment. Automated classification depends on classifier training data and can misclassify context (a limitation examined directly in Chapter 8.3). A finding that survives all three approaches simultaneously is far less likely to be a methodological artifact than a finding produced by any single tracker alone. This convergence is, in effect, the strongest available evidence that the October 2023 spike was a real, substantial phenomenon rather than a measurement effect — and it establishes the evidentiary baseline against which this report's platform-specific findings (Chapter 4) and its central causal case study (Chapter 5) should be read.

3.3 Geographic and structural distribution

FRA's national-level data shows substantial cross-country variation, with Austria (71%), Germany (68%), and Finland (63%) recording the highest twelve-month discrimination rates in the 2021–2022 survey wave. FRA's data also documents a structural, non-discursive dimension to this discrimination: two in five Muslim respondents (41%) were overqualified for their current employment. This finding matters for how this report frames the problem as a whole — online hostility does not exist in a separate sphere from material disadvantage; the evidence suggests the two reinforce each other, a dynamic developed further in Chapter 6.

National-level civil-society documentation, where it exists, tends to be more granular than EU aggregate data and correspondingly more alarming. Austria's Documentation Centre (Dokustelle) recorded 1,522 racist attacks against Muslims and perceived Muslims in 2023 — of which 66.7% were classified as online incidents — up from 1,324 in 2022 (1,080 online), a year-on-year increase exceeding 100% in victim-reported cases specifically. This is the most granular online/offline-disaggregated national dataset identified in this report's research process, and it is presented in full in the cross-country comparison in Chapter 9, alongside the significant reporting gaps documented for other countries in the same period.

3.4 From online spike to offline consequence: the general pattern

The clearest documented instance of online anti-Muslim content translating into offline violence in this report's evidence base is the wave of rioting that followed the 29 July 2024 Southport stabbing attack, examined in full as a case study in Chapter 5. It is introduced here because it illustrates a distinct dynamic from the diffuse October 2023 spike: rather than a broad increase in hostile content, a single false claim achieved rapid, high-volume distribution and directly preceded an attack on a specific mosque and six days of national rioting. Reading Sections 3.2 and 3.4 together establishes the two mechanisms this report treats as analytically distinct but mutually reinforcing: *diffuse escalation* (a general rise in hostile content following a geopolitical trigger) and *acute amplification* (a single false claim achieving outsized reach through platform mechanics). Chapter 6 argues both mechanisms trace back to the same underlying enforcement gap, differently expressed.

3.5 Why comprehensive measurement remains impossible

This chapter's findings should be read with an important caveat that recurs throughout this report: no comprehensive, EU-wide, religion-disaggregated measurement infrastructure for anti-Muslim online hate exists. FRA's own convened workshop on this question identified underreporting, denial of anti-Muslim racism, and the stigmatisation of Muslim advocacy organisations as ongoing obstacles to comprehensive measurement. Civil-society trackers (ISD, the Online Hate Prevention Institute, national documentation centres such as Austria's Dokustelle) fill part of this gap, but use differing methodologies, sampling windows, and platform access levels, which complicates direct comparison across trackers or over time. As Chapter 7 details, this measurement gap is not incidental — it is directly connected to the restricted researcher data access that both X and Meta have been formally found, by the European Commission, to have obstructed. The measurement problem and the enforcement problem are, in this respect, the same problem viewed from two angles.

Chapter summary. The evidence in this chapter establishes three things a policymaker should take as established fact rather than contested claim: discrimination against Muslims in Europe was already rising before 2023; the October 2023 spike in online hostility is corroborated by

three independent methodologies and should not be dismissed as a single tracker's artifact; and the absence of comprehensive, disaggregated measurement infrastructure is not a neutral data gap but a structural feature of the enforcement gap this report is centrally concerned with.

Chapter 4: Platform Governance in Practice — X and Meta

Chapter 3 established the scale of the problem. This chapter asks a narrower and more legally consequential question: what have the two platforms most implicated in that evidence base — X and Meta — actually done about it, and what does their documented conduct reveal about the gap between platform governance as designed and platform governance as practised?

4.1 X: regulatory history and the limits of transparency-focused enforcement

X was designated a Very Large Online Platform (VLOP) under the DSA on 25 April 2023, triggering the Act's most stringent systemic-risk obligations. The European Commission opened formal proceedings against X on 18 December 2023 addressing the dissemination of illegal content and the platform's handling of disinformation — the strand of investigation most directly relevant to hate-speech amplification, and, notably, the strand that remained open and unresolved throughout this report's period of study.

A narrower, parallel investigation into X's advertising transparency, "dark patterns," and researcher data access produced preliminary findings of non-compliance on 12 July 2024: a deceptive "Blue checkmark" verification system found to compromise account authenticity signals; the absence of a properly searchable advertising repository; and restricted, high-fee researcher data access in contravention of Article 40(12) obligations. In January 2025 the Commission escalated its scrutiny of X's recommender systems specifically, and this narrower investigation culminated on 5 December 2025 in the DSA's first-ever non-compliance fine: €120 million.

Analytical significance. The sequencing here is the single most important structural finding in this chapter. The completed, sanctioned violations concern transparency and procedural failures — not a substantive finding that X's recommender system amplifies anti-Muslim hate. The investigation that would produce such a finding remains open. This is not necessarily evidence of regulatory failure — transparency violations are often more straightforward to establish evidentially than amplification effects — but it means that, as a matter of adjudicated legal fact, no DSA enforcement action to date directly addresses the amplification mechanisms this report's evidence (Chapters 3 and 5) documents. Chapter 6 develops this sequencing gap as one of the enforcement gap's core components.

Independent evidence corroborates the underlying concern the open investigation is meant to resolve. The European Commission's own preliminary findings noted that X's crowd-sourced fact-checking system (Community Notes) does not reliably ensure systematic detection of illegal or inappropriate content, and that visibility-limiting mechanisms do not prevent harmful posts from reaching substantial audiences even where partial restriction is applied. This finding is echoed in X's own DSA-mandated risk assessment, cited in the European Board for Digital Services' first Article 35(2) systemic-risk report, which acknowledges that hateful conduct on the platform creates risks to freedom of expression through self-censorship among targeted users — a significant admission because it comes from the platform itself in a legally mandated disclosure, even though its framing centres the chilling effect on targets of abuse rather than conceding the platform's own systems actively amplify that abuse.

4.2 Meta: parallel proceedings and a self-acknowledged policy reversal

The Commission's DSA scrutiny of Meta followed a comparable trajectory: an October 2023 request for information on risk assessment and mitigation measures; formal proceedings opened 30 April 2024; and, on 24 October 2025, preliminary findings that Meta had breached DSA transparency obligations — specifically, that Facebook and Instagram made it unduly difficult for users to flag illegal content and challenge moderation decisions, and deployed "dark patterns" in reporting flows for serious content categories including terrorist material.

The single most consequential development for this report's subject matter, however, is not a Commission finding at all — it is Meta's own January 2025 decision to end third-party fact-checking, replace it with a user-generated "Community Notes" model, and rewrite its "Hateful Conduct" policy in a direction CEO Mark Zuckerberg himself described as accepting a deliberate "tradeoff": more harmful content remaining visible on the platform. Meta's own Oversight Board — an entity separate from Meta, funded through an independent trust, whose standard decisions are binding on the company — formally criticised this change as having been made "hastily," without evidence of prior human rights due diligence, and urged Meta to assess the policy's human rights impact on affected groups.

Analytical significance. This is the clearest instance in this report's evidence base of a platform *voluntarily* widening the enforcement gap rather than merely failing to close it through inadequate resourcing. The distinction matters for how policymakers should read platform self-regulation: it is not simply that content-moderation systems have gaps due to scale or technical limitation (a defensible, if inadequate, explanation) — Meta's own framing acknowledges this specific policy change as a deliberate choice to tolerate more harmful content. Chapter 6 treats this as direct evidence that platform incentives, absent binding regulatory pressure, do not reliably align with the DSA's stated risk-mitigation objectives.

4.3 The Oversight Board's Southport findings as a case of documented, binding failure

The clearest instance of Meta's content moderation directly failing to prevent amplification of anti-Muslim content in this report's period is the Oversight Board's own review of Meta's handling of the 2024 UK riots, examined in full in Chapter 5. Reviewing three posts shared on Facebook during the unrest, the Board overturned Meta's original decisions to leave the content up, finding each created a risk of "likely and imminent harm" that should have led to removal, and explicitly linked the posts to "a period of contagious anger and growing violence, fuelled by misinformation and disinformation on social media," in which "anti-Muslim and anti-immigrant sentiment spilled onto the streets." The Board further found Meta's own crisis-response mechanisms were triggered only after the most acute period of violence — including the attack on Southport's mosque — had already occurred.

This finding is significant beyond its immediate factual content because it is not a civil-society allegation but a binding institutional ruling. It converts what might otherwise be treated as a contested claim about platform failure into adjudicated fact, and it is the single strongest piece of evidence in this report that the enforcement gap identified in Chapter 6 has produced measurable, attributable real-world harm rather than remaining a theoretical or statistical concern.

4.4 Structural transparency limitations affecting both platforms

Both platforms' own DSA-mandated transparency disclosures reveal a shared structural weakness directly relevant to this chapter's central question. X's own reporting shows 238,000 illegal-content reports received in the period analysed, with 115,000 (approximately 48%) actioned, a median resolution time of 2.7 hours, but only 667 appeals filed — an appeal rate of roughly 0.58% — of which 190 were overturned. X frames this low appeal rate as evidence of decision accuracy. This report treats that framing as unproven: a sub-1% appeal rate is equally consistent with

limited user awareness of, or trust in, the appeals mechanism, and X's self-reported data cannot distinguish between the two explanations.

Meta's transparency reporting reveals a more fundamental structural gap. Independent analysis by the International Network Against Cyber Hate (INACH) found that Meta's statutory disclosures do not provide a distinct category for "illegal hate speech" at all — hate-speech notices are subsumed within a broad "other illegal content" catch-all that excludes intellectual property, defamation, and privacy violations but does not isolate hate speech, religious discrimination, or any other specific harm type by protected characteristic. This means the raw notice and removal figures Meta discloses (113,638 notices to Facebook, 16,052 removals; 61,888 notices to Instagram, 12,418 removals, in the period analysed) cannot be further disaggregated to determine what share, if any, concerned anti-Muslim content specifically — using Meta's own published data.

Analytical significance. This is not a minor reporting inconvenience. It is a concrete, documented instance of the measurement gap identified in Chapter 3.5, now demonstrated at the level of the platforms' actual statutory compliance practice rather than as a general observation about data availability. A regulatory framework cannot verify compliance with a non-discrimination risk-assessment obligation if the platform's own reporting structure makes that specific category of harm invisible by design. Chapter 6 treats this disaggregation gap as one of the enforcement gap's clearest structural manifestations, and Chapter 13's recommendations address it directly.

4.5 Chapter conclusion: why this matters for regulators

Three findings from this chapter should inform how policymakers assess platform compliance going forward. First, completed DSA enforcement to date has addressed transparency and procedural violations rather than the substantive amplification question — a sequencing gap regulators should treat as an open, urgent priority rather than a settled matter. Second, at least one platform (Meta) has demonstrated that self-regulation, absent binding pressure, can move in the direction of *increased* rather than decreased risk tolerance — a finding that should weigh against reliance on voluntary compliance mechanisms alone. Third, both platforms' own transparency disclosures contain structural gaps — appeal-rate ambiguity for X, complete category-level invisibility for hate speech at Meta — that mean current reporting requirements are insufficient to verify DSA compliance with non-discrimination obligations specifically. This is a concrete, addressable regulatory design problem, not merely a platform behaviour problem, and Chapter 13 proposes a specific remedy.

Chapter 5: Case Studies in Digital-to-Physical Harm

Chapters 3 and 4 established, respectively, the scale of online anti-Muslim hate and the documented conduct of the two platforms most implicated in its spread. This chapter examines three case studies selected, as explained in the Methodology, not to illustrate a single uniform pattern but to test this report's central thesis against genuinely different mechanisms of harm: acute algorithmic amplification of a specific falsehood (Southport); institutional press speech generating recurring controversy through a structurally different pathway (Charlie Hebdo); and formal political normalisation of anti-Muslim narrative content (the 2024 election cycles). Reading the three together shows that the enforcement gap identified in Chapter 6 operates across more than one mechanism — a finding with direct implications for how comprehensive any regulatory remedy needs to be.

5.1 Southport: the clearest documented causal chain

The trigger. On 29 July 2024, three young girls — Alice da Silva Aguiar, Bebe King, and Elsie Dot Stancombe — were murdered, and ten others injured, in a knife attack at a children's dance class in Southport, England. The attacker was 17-year-old Axel Rudakubana, who was neither Muslim nor an asylum seeker.

The falsehood and its spread. Within hours, a false claim — that the attacker was a Muslim asylum seeker who had recently arrived by boat — began circulating on social media, attached at points to a fabricated name, "Ali al-Shakati," later debunked by police. Research by the Institute for Strategic Dialogue documented that far-right online networks began organising protests the day after the attack, by which point the false narrative was already in wide circulation, with promotional material on X explicitly linking a planned protest to anti-migrant and anti-Muslim framing of the attack. Pakistani authorities later arrested a freelance web developer, Farhan Asif, charging him with cyber terrorism for allegedly originating the false suspect-identity claim.

Platform amplification, measured. On X, individual posts promoting the disinformation received tens of thousands of engagements; one video alone reached approximately 230,000 views before correction. Far-right activist Tommy Robinson, with roughly 900,000 X followers, encouraged his audience to join the unrest; Amnesty International's subsequent analysis found his posts received over 580 million views in the two weeks following the attack — a reach Amnesty characterised as extraordinary for a figure previously banned from most mainstream platforms for hate-speech violations. This reach figure is analytically significant beyond its scale: as discussed in Chapter 7's treatment of algorithmic amplification, a single account's follower count cannot organically explain 580 million views in two weeks, implying substantial algorithmic redistribution beyond the poster's direct audience — the clearest quantitative anchor for the amplification concern this report otherwise has to argue more abstractly. X's own ownership engaged directly with riot-related content during the unrest, including a reply describing "civil war" as "inevitable."

On Meta's platforms, as detailed in Chapter 4.3, the Oversight Board found at least three Facebook posts should have been removed under Meta's own policy and were not, and that Meta's crisis-escalation protocols activated only after the most acute violence had already occurred.

Consequences. Rioting occurred across multiple English and Northern Irish towns and cities between 30 July and 5 August 2024. Documented consequences include an attack on Southport Islamic Society Mosque — windows smashed, a police van set alight outside the building — attacks on other mosques and accommodation housing asylum seekers, over 361 injured police officers, and 1,840 arrests, of which 1,103 resulted in charges. A subsequent public inquiry into the Southport attack has been the subject of calls from Muslim community leaders to extend its scope to the riots and the underlying disinformation dynamics.

Why this case anchors the report. Southport is unusual in offering a complete, independently verifiable causal chain: a specific, falsifiable claim; measurable, quantified amplification on a named platform; a binding institutional finding (the Oversight Board ruling) that content should have been removed and was not; and documented, attributable offline violence against a named place of worship. Most accounts of online-to-offline harm in this field have to infer the causal link; Southport allows this report to demonstrate it directly. It is, accordingly, the case study Chapter 6 draws on most heavily in constructing the enforcement-gap argument.

5.2 Charlie Hebdo: a structurally different case, deliberately included

The Charlie Hebdo case is included specifically because it does not fit the Southport pattern, and testing this report's thesis against a genuinely different case is a methodological safeguard against overclaiming.

On 7 January 2015, two Islamist gunmen attacked the Paris offices of the satirical weekly *Charlie Hebdo*, killing 12 people in direct retaliation for the magazine's repeated publication of caricatures of the Prophet Muhammad. The attack produced the #JeSuisCharlie movement and a decade-long, still-unresolved debate over the boundary between satire, blasphemy, and Islamophobia.

Unlike Southport, this is not a case of coordinated hostile amplification of a falsehood. It is recurring online controversy generated by an institutional, legally protected press actor engaging in repeated, deliberate provocation — including the magazine's August 2017 cover linking the Barcelona terror attack to Islam as a religion, condemned by critics (including French commentators) as risking the fanning of Islamophobia, and a 2017 cartoon targeting scholar Tariq Ramadan that generated fresh, specific death threats against staff. The magazine's January 2025 tenth-anniversary edition, soliciting a fresh caricature contest, drew substantial and reportedly growing French public support for unconditional satirical freedom regarding religious subject matter — a sentiment increasingly in tension with the concerns raised by French Muslim communities and international commentators, including Turkey's president, who characterised rising European Islamophobia as having reached "intolerable" levels at the UN in September 2023.

Analytical significance. This case demonstrates that not all online anti-Muslim controversy is amenable to the same regulatory remedy this report proposes elsewhere. Content-moderation reform, researcher access, and systemic-risk assessment — the remedies developed in Chapters 6 and 7 — are designed to address coordinated amplification of illegal or borderline content, not the harder question of where legitimate satirical and press freedom ends and religiously targeted provocation begins. This report does not attempt to resolve that boundary question; it flags the case as a genuine limit on the scope of the enforcement-gap framework, directly relevant to the free-expression counterarguments addressed in Chapter 12.

5.3 The 2024 election cycles: normalisation through formal political channels

The June 2024 European Parliament elections, alongside concurrent national campaigns in Germany, France, and elsewhere, document a third and distinct mechanism: anti-Muslim narrative content produced *by*, rather than merely tolerated by, actors operating within formal democratic political structures.

Research published by the European Digital Media Observatory (EDMO) documented Germany's Alternative für Deutschland (AfD) making sustained use of AI-generated visual content throughout its 2024 and early-2025 campaigning, consistently contrasting an "ideal Germany" (blond, blue-eyed imagery) against AI-generated depictions of migrants portrayed as dark-skinned and threatening — imagery EDMO's analysis notes drew implicit visual reference to the December 2024 Magdeburg Christmas market attack. EDMO documented a parallel pattern in French far-right figure Éric Zemmour's (Reconquête) 2024 campaign content, and noted this technique was used by multiple French far-right parties in the same period, not as an isolated instance.

Beyond individual campaign content, the "Patriots for Europe" political group — formed after the June 2024 elections and led by Jordan Bardella of France's Rassemblement National, then the European Parliament's third-largest group — has been documented explicitly invoking "Great Replacement" conspiracy framing at the institutional political level, alleging a "cultural and religious invasion incompatible with European values" driven by Muslim immigration.

Peer-reviewed research on the 2024 election cycle, published in *Media and Communication*, found migration and electoral-integrity framing were the predominant topics of election-period disinformation, that this disinformation's life cycle extended well beyond election day, and that the problem was more intense in Southern and Eastern Europe specifically.

Analytical significance. This case demonstrates that the same "Great Replacement" narrative CCDH found platforms failed to act on in 89% of flagged instances (Chapter 10) is simultaneously being mainstreamed through a formal, elected pan-European political grouping. Platform content moderation and political normalisation are not independent phenomena operating in separate spheres; they are parallel, mutually reinforcing channels for the same narrative content — directly relevant to the DSA's Article 34(1)(c) civic-discourse and electoral-integrity risk category, discussed alongside the non-discrimination risk category in Chapter 7.

5.4 Chapter conclusion: three mechanisms, one underlying gap

These three case studies demonstrate that "online Islamophobia" is not a single phenomenon with a single remedy. Southport shows acute algorithmic amplification of a specific falsehood, with platform moderation failure as the proximate, addressable cause. Charlie Hebdo shows a harder case at the boundary of legitimate expression, where content-moderation remedies are poorly suited and a different — legal, cultural, and long-term — response is required. The 2024 elections show formal political normalisation operating in parallel with, and reinforcing, platform-level under-moderation. A regulatory framework addressed only to platform content moderation, without also addressing political accountability and the definitional clarity discussed in Chapter 2, will close only part of the gap these three cases collectively expose. This is the argument Chapter 6 develops in full.

Chapter 6: The Enforcement Gap

6.1 Naming the problem precisely

The preceding three chapters have shown, in sequence: that discrimination and online hostility toward Muslims in Europe are measurably rising (Chapter 3); that the two platforms most implicated in that rise have, at best, addressed procedural symptoms rather than the substantive amplification question, and at worst have voluntarily loosened their own safeguards (Chapter 4); and that the mechanisms of harm are more varied — acute amplification, institutional press provocation, formal political normalisation — than any single remedy can address (Chapter 5).

This chapter names the pattern connecting these findings: **the enforcement gap**. It is the persistent, structural distance between the legal and institutional protections that exist against anti-Muslim discrimination in principle, and their practical operation in the systems — platform, regulatory, and civil-society — that are supposed to give those protections effect. This report's central claim, stated in the Introduction, is that this gap, not the absence of law, is the principal challenge Europe now faces in this domain.

This is a stronger and more precise claim than "platforms don't enforce their rules" or "the law is too weak," and it matters to distinguish it from both. Europe's legal protections against religious discrimination are not weak in comparison to most jurisdictions globally; the EU Charter, the Racial and Employment Equality Directives, and now the DSA together constitute one of the more comprehensive frameworks in existence. The problem this report documents is not a drafting failure. It is an implementation failure, occurring at several identifiable points, each with a distinct cause and — critically for policymakers — a distinct and addressable remedy.

6.2 The four components of the enforcement gap

Component one: definitional fragmentation

Chapter 2 established that no single agreed definition of Islamophobia exists in EU or international law, and that this fragmentation has direct downstream consequences: FRA's own convened experts identify the absence of an agreed definition as an obstacle to comprehensive equality data collection, and empirical research reviewed in Chapter 8 demonstrates that imprecise definitions measurably reduce the accuracy of automated hate-speech detection systems — producing both false negatives (genuine Islamophobic content that goes undetected) and false positives (legitimate speech wrongly flagged). A regulatory or platform-moderation system cannot enforce a category it cannot precisely define, and this report's evidence shows that imprecision is not merely a theoretical problem but one with measured, practical enforcement consequences.

Component two: measurement invisibility

Chapters 3 and 4 documented, from two different angles, that anti-Muslim online hate is structurally difficult to measure. At the survey and hate-crime level, ODIHR's own 2024 reporting shows significant, self-acknowledged national gaps — the Netherlands recorded zero anti-Muslim hate-crime submissions for 2024, a result ODIHR itself flags as indicating underreporting rather than an absence of incidents. At the platform level, Meta's own DSA transparency disclosures do not isolate "illegal hate speech" as a distinct reporting category at all, making it structurally impossible to verify, from Meta's own published data, what share of its enforcement actions address anti-Muslim content specifically. A harm that cannot be measured cannot be demonstrably mitigated, and cannot be the basis for regulatory enforcement that withstands legal scrutiny — which is precisely why the researcher-access question addressed in Component three is not a peripheral technical issue but central to closing this gap.

Component three: restricted verification access

This report's own evidentiary limitations, disclosed in the Methodology, are not incidental — they are direct evidence of this component. Both X and Meta have been formally found by the European Commission, in adjudicated enforcement decisions, to have obstructed the researcher data access that Article 40(12) of the DSA envisions for exactly this purpose. This creates what this report terms an *evidentiary bootstrapping problem*: the platforms most in need of independent scrutiny for algorithmic amplification of hate content are the same platforms whose own restricted data access prevents that scrutiny from being conducted with full rigour. EqualFaith Europe's own legal memorandum on Article 40(12) — understood to be among the first analyses specifically addressing this provision as a pathway for civil-society research into systemic anti-Muslim discrimination — argues that the legal reasoning already accepted by the Commission for election-integrity data access extends, without a meaningful legal distinction, to religious-discrimination research. The Commission's own enforcement findings against both platforms for researcher-access obstruction lend that argument independent, contemporaneous support: the same structural failure the Commission has now formally sanctioned in the transparency context is, on this analysis, equally present and equally actionable in the religious-discrimination context.

Component four: capture-vulnerable enforcement mechanisms

The DSA's Article 22 trusted-flagger mechanism illustrates a fourth, distinct failure mode: not restricted access, but a mechanism vulnerable to being formally available while practically ineffective for its intended beneficiaries. Article 22 does not require disclosure of a flagger's funding sources or political affiliations as a condition of designation. In principle, this mechanism could substantially accelerate removal of anti-Muslim illegal content if religious-minority civil-society organisations were systematically designated and resourced as trusted flaggers. In practice, without disclosure requirements, the mechanism is equally available to entities whose institutional interests diverge from the communities Article 22 is meant to protect — a capture risk this report treats as structurally comparable to, though legally distinct from, the researcher-access restriction in Component three: in both cases, a formally available legal mechanism can fail to deliver its intended protective effect for reasons that have nothing to do with the underlying substantive law.

6.3 Why systemic governance failure matters more than individual content decisions

A recurring temptation in this field is to treat each documented failure — a post left up, an appeal denied, a flagger overlooked — as an isolated error correctable through better individual moderation decisions. This report's evidence argues against that framing directly. The Meta

Oversight Board's Southport findings (Chapter 4.3) did not identify three isolated moderation mistakes; they identified a crisis-response protocol that activated too late as a matter of institutional design. The CCDH finding that platforms fail to act on 89% of clearly identifiable, manually reported anti-Muslim hate content (Chapter 10) is not a story about individual reviewer error at that scale; it is a story about systemic under-resourcing or systemic policy choice. Meta's January 2025 policy reversal (Chapter 4.2) was not a moderation failure at all — it was a deliberate, acknowledged, system-wide recalibration of risk tolerance.

This distinction is precisely what the DSA's Article 34/35 systemic-risk framework, discussed fully in Chapter 7, is designed to capture, and it is why this report treats the DSA — imperfectly implemented as it currently is — as the correct legal instrument for closing the enforcement gap, rather than treating incremental improvements to individual content moderation as sufficient on their own. A regulatory response aimed only at individual bad decisions will always be one step behind a problem whose root causes, on this report's evidence, are structural: definitional, measurement-related, access-related, and mechanism-design-related.

6.4 The cycle this gap sustains

Synthesising the evidence from Chapters 3 through 5, this report identifies a recurring cycle, not merely a static gap: online hostility (Chapter 3) is inadequately measured (Component two) and inadequately moderated (Chapter 4), which allows specific falsehoods to achieve outsized reach (Chapter 5.1); that reach produces real-world consequences — violence, but also, as documented in Chapter 9's cross-country evidence, discriminatory outcomes in employment and housing and policy responses such as face-covering bans that Chapter 12 argues are frequently justified by reference to a climate of suspicion those same online dynamics helped generate; those consequences do not generate corrective regulatory pressure proportionate to their scale, because the restricted access and measurement gaps documented in Components two and three make the causal chain harder to demonstrate with the rigour regulators require; and the cycle repeats at the next trigger event. Breaking this cycle at any single point would help; this report's recommendations in Chapter 13 are designed to intervene at all four components simultaneously, on the argument that a partial fix at only one point leaves the underlying cycle largely intact.

Chapter conclusion. The enforcement gap is not a rhetorical framing device. It is a specific, four-part, evidenced diagnosis: definitional fragmentation that undermines detection accuracy; measurement invisibility that prevents harm from being demonstrated; restricted verification access that prevents independent scrutiny of the platforms most implicated; and capture-vulnerable protective mechanisms that can exist on paper without functioning in practice. Each component has a distinct, addressable remedy, developed in Chapter 13. The remainder of this report — the DSA analysis in Chapter 7, the wider evidence base in Chapter 8, the cross-country comparison in Chapter 9 — should be read as further substantiation of this diagnosis, not as separate, disconnected findings.

Chapter 7: The Digital Services Act — From Content Moderation to Systemic Risk Governance

7.1 A genuine shift in regulatory design

The Digital Services Act represents a legally significant departure from prior content-governance approaches, and this report's analysis depends on that distinction being understood precisely. Earlier instruments — including the 2016 EU Code of Conduct on Countering Illegal Hate Speech Online, discussed in Chapter 2.4 — operated primarily at the level of individual content decisions: a piece of content is reported, assessed against a policy, and removed or retained. The DSA's Article 34/35 framework for Very Large Online Platforms operates at a categorically different level: it requires platforms to assess and mitigate *systemic risk* — the foreseeable, structural probability that a platform's design, algorithms, and scale will produce society-wide harm, independent of any single content decision.

This distinction is not merely technical. Recital 80 of the DSA clarifies that "systemic risk" under Article 34(1)(a) extends to a platform's *dissemination, amplification, or propagation* of illegal content, not merely its hosting of that content. This is the direct legal hook for the algorithmic amplification concern documented throughout this report — most concretely in the Southport reach figures analysed in Chapter 5.1 — and it means that a pattern of repeated individual moderation failures, of the kind documented in Chapter 4, can itself constitute evidence of a systemic risk-assessment failure, rather than remaining a series of unconnected incidents. Article 34(1)(b) extends this obligation explicitly to negative effects on Charter fundamental rights, including the right to non-discrimination under Article 21 — placing anti-Muslim discrimination squarely, if not yet fully operationally, within the DSA's intended scope.

7.2 Systemic risk assessment in practice: an operational gap, not a scope gap

The European Board for Digital Services' first Article 35(2) report on systemic risks and mitigations, published in late 2025, confirms that non-discrimination risk under Charter Article 21 has been explicitly raised in VLOP risk-assessment reporting, with providers and civil-society organisations flagging risks from the large-scale dissemination and amplification of content reinforcing discriminatory views. This confirms the legal scope point made above: religious-minority discrimination is not excluded from DSA risk assessment as a matter of law.

The gap this report identifies is operational, not jurisdictional. No dedicated Commission guideline currently addresses religious-minority hate-content amplification specifically, in contrast to the detailed, purpose-built guidance already issued for electoral risk (the March 2024 Guidelines on the Mitigation of Systemic Risks for Electoral Processes) and, more recently, minor protection. This asymmetry is consistent with the pattern identified in Chapter 3.5 and Component two of the enforcement gap (Chapter 6.2): religious-minority harm is measured and guided less comprehensively than other DSA-recognised risk categories, not because the law excludes it, but because the operational infrastructure to address it specifically has not yet been built.

7.3 Article 22: designed protection, undelivered protection

Article 22 requires platforms to give priority handling to notices from Digital Services Coordinator-designated "trusted flaggers." As Chapter 6.2 (Component four) established, the absence of funding and affiliation disclosure requirements in the designation criteria creates a capture risk that is not hypothetical: EqualFaith Europe's own policy analysis (briefing ref. EFE/DSA/2026-01), using Germany and Ireland as illustrative examples, has identified this independence gap directly and proposed amendments requiring disclosure as a condition of designation.

The legal analysis here is precise and important to state correctly: this is not a claim that Article 22 has been misused in any documented instance identified in this report's research. It is a claim about the mechanism's *design* — that its current criteria do not require the safeguard that would prevent misuse, and that this design gap has direct, foreseeable consequences for whether religious-minority civil society can both attain trusted-flagger status and trust the designations already granted to others.

7.4 Article 40(12): the researcher-access provision and its self-reinforcing restriction

Article 40(12) establishes a pathway for vetted researchers — and, on the reading this report and EqualFaith's own legal memorandum advance, appropriately qualified civil-society organisations — to obtain platform data for systemic-risk research, non-discrimination risk included. As Chapter 4.4 documented, both X and Meta have been found, in the Commission's own preliminary and final enforcement decisions, to have restricted exactly this kind of access: X's practices were cited among the three grounds for the December 2025 €120 million fine, and Meta's October 2025 preliminary findings identified comparable obstruction, following Meta's August 2024 discontinuation of its CrowdTangle research tool without an adequate DSA-compliant replacement.

This report treats this finding as legally and analytically central rather than as one item in a longer list, for the reason developed in Chapter 6.2 (Component three): it is *self-reinforcing*. Restricted researcher access prevents exactly the kind of rigorous, platform-internal-data-based amplification research that would allow regulators, courts, or civil society to demonstrate a systemic-risk violation with the evidentiary rigour DSA enforcement requires — which in turn makes it harder to compel the platforms to grant the access that would resolve the underlying uncertainty. The Commission's own adjudicated finding that this obstruction has occurred, at both platforms studied, converts what might otherwise be treated as a civil-society grievance into a documented regulatory fact, and it is the strongest available legal support for the Article 40(12) civil-society-access expansion this report recommends in Chapter 13.

7.5 Enforcement sequencing: what has been sanctioned, and what has not

Synthesising Chapter 4's platform-specific findings at the level of DSA enforcement design: completed Commission enforcement actions against both X and Meta to date address procedural and transparency failures — dark patterns, advertising repositories, researcher access, minor-protection risk-assessment methodology — rather than reaching a substantive finding on whether either platform's systems systemically amplify illegal or harmful anti-Muslim content specifically. This is not, in itself, evidence of inadequate regulatory intent; transparency violations are generally more straightforward to establish evidentially than amplification effects, and sequencing enforcement from more to less tractable

violations is a defensible prosecutorial strategy.

It does mean, however, that as of this report's period of study, no DSA enforcement action yet directly and substantively addresses the amplification mechanisms documented in Chapters 3 through 5. The transparency-focused fines remain significant for a different reason: they establish, as adjudicated fact rather than civil-society allegation, that both platforms have obstructed precisely the kind of independent researcher access that would be needed to establish amplification effects conclusively. This is the evidentiary bootstrapping problem named in Chapter 6.2 restated at the level of formal enforcement outcomes.

7.6 Chapter conclusion: why this matters for regulators and Digital Services Coordinators

Three conclusions follow directly from this chapter's analysis and should inform Digital Services Coordinators' and the Commission's next enforcement priorities. First, the DSA's systemic-risk framework is legally adequate in scope to address religious-minority discrimination; the gap is in operational guidance and disaggregated reporting requirements, not in the underlying legal text — meaning the remedy is achievable through Commission guidance and implementing measures rather than legislative amendment. Second, the researcher-access obstruction the Commission has already sanctioned at both platforms studied should be treated as directly probative of the broader amplification question this report cannot itself resolve without that same access — an argument for prioritising Article 40(12) civil-society access expansion specifically, not merely as one item among several reform proposals. Third, the sequencing pattern in current enforcement — transparency violations addressed first, amplification questions left open — is defensible as prosecutorial strategy but should not be mistaken for the underlying problem being resolved; regulators and civil society should treat the current fines as a first step in a longer enforcement trajectory, not as evidence the core concern has been addressed.

Chapter 8: The Academic and Institutional Evidence Base

This chapter synthesises the peer-reviewed and official-institutional evidence underpinning this report's argument, organised to show how each source contributes to the enforcement-gap diagnosis rather than as a standalone literature catalogue. Figures already presented in Chapters 3 through 5 are referenced rather than restated in full.

8.1 Peer-reviewed research: three convergent strands

Computational and prevalence research corroborates the scale findings in Chapter 3 through methodologies independent of the civil-society trackers already discussed. Kasahara, Schroeder, Yazidi and Lind's Norwegian study, applying a neural-network classifier to over one million Twitter and Facebook comments (2010–2021), found hateful content increasing significantly over the study period using automated rather than manual or civil-society-flagged methods — an independent check on trend direction, albeit with a dataset ending before the October 2023 inflection point that dominates this report's more recent evidence. Rosen and Walther (2025), examining X specifically in the period following 7 October 2023, identified a distinct amplification mechanism operating through *social approval* dynamics: users receiving more positive engagement on hateful posts go on to produce more extreme content in subsequent posts — a mechanism operating through engagement-signal design rather than recommender-system ranking directly, and therefore analytically distinct from, but compounding, the algorithmic amplification concern addressed in Chapter 7. Ünlü and colleagues (2026), publishing in *New Media & Society*, applied mixed-methods classification to Finnish Twitter data (2018–2023) and identified 32 distinct anti-Muslim thematic clusters, providing granularity on *what* the rising volume of content documented in Chapter 3 actually contains.

Behavioural and attitudinal research extends this report's concern beyond content prevalence to downstream consequence. Lajevardi, Oskooii, and Walker (2022), published in the *Journal of Public Policy*, found a strong and consistent association between reliance on social media as a primary news source and support for anti-Muslim policies, using three original U.S.-focused surveys and the Nationscape dataset. While outside this report's primary European geographic focus, this study is methodologically important because it demonstrates a downstream behavioural outcome — measurable political-preference formation — providing a direct evidentiary link between the online content documented in this report and the real-world discriminatory outcomes (employment, housing, policy) discussed in Chapter 9, complementing the more visible but rarer violence pathway documented in the Southport case study.

Detection-methodology research provides the direct evidentiary basis for Component one of the enforcement gap (Chapter 6.2): a 2023 IEEE/ACM conference paper found that imprecise definitions of Islamophobia in automated detection systems produce both false negatives and false positives, and that a refined definition reduced misclassification while maintaining detection accuracy — empirical confirmation that the definitional fragmentation discussed in Chapter 2 has measured, not merely theoretical, consequences for platform enforcement quality.

A distinct and more recent strand examines bias in the large language models increasingly integrated into content-moderation and recommendation pipelines. One frequently cited study found GPT-3 associated the word "Muslim" with "terrorist" in 23% of tested completions, markedly higher than comparable associations for other religious identity terms — a forward-looking concern this report addresses directly in its concluding risk assessment (Chapter 13.3), given the increasing integration of such systems into the platform infrastructure examined in Chapter 4.

8.2 Official and institutional reporting: convergent, cross-institutional findings

The FRA, ODIHR, and Council of Europe reporting already introduced in Chapters 3 and 9 converges across institutions in a way that strengthens confidence in this report's core findings independent of any single body's methodology.

FRA's "Being Muslim in the EU" survey and its 2025 Fundamental Rights Report — the latter describing anti-Muslim and antisemitic hate crime as having "exploded" across Europe, in language attributed to Commission Vice-President Margaritis Schinas — are corroborated by ODIHR's independently collected hate-crime data (Chapter 9) and by the Council of Europe's ECRI 2024 Annual Report, which found antisemitic incidents in the final quarter of 2023 substantially exceeded typical annual totals in several countries and that anti-Muslim hate incidents, including online hate speech specifically, proliferated over the same period. The European Parliament's own research service has synthesised this convergence in its 2024 briefing, referencing the 29 November 2023 Joint Statement of EU Coordinators, Special Representatives, Envoys and Ambassadors on Combating Anti-Muslim and Antisemitic Hatred and Discrimination.

This cross-institutional convergence — an EU agency, an OSCE body, a Council of Europe monitoring mechanism, and the European Parliament's own analysts independently describing the same escalation in the same period — is presented here as a distinct evidentiary category from the platform-specific and academic evidence discussed elsewhere in this report: it establishes that this report's core empirical claim (rising anti-Muslim hostility, online and offline, particularly from late 2023) is not contested or marginal within the EU's own institutional assessment of the problem.

8.3 Civil-society monitoring: value and limits

Civil-society sources — the Institute for Strategic Dialogue, the Center for Countering Digital Hate, the Online Hate Prevention Institute, and national documentation centres such as Austria's Dokustelle — provide the platform- and country-specific granularity that official institutional reporting generally does not. Their evidentiary value and limitations are assessed systematically in Chapter 11; the key point for this chapter is that, wherever this report's civil-society findings have been checked against an independent methodology (Chapter 3.2's three-way convergence; the Meta Oversight Board's independent confirmation of Southport-related moderation failure), they have held up. This pattern supports treating civil-society monitoring as a reliable evidentiary tier for this report's purposes, while still applying the source-specific caveats documented in Chapter 11.

For completeness, this report's research process also identified the Organisation of Islamic Cooperation's Sixteenth Report on Islamophobia and Bayrakli and Hafez's annual European Islamophobia Report as indexed sources within FRA's own anti-Muslim hatred database; both are weighted, per Chapter 11, as stakeholder rather than neutral-monitor sources, appropriate for corroborating qualitative trend direction (for instance, the OIC report's finding of a stronger upward trend in the UK, France, and Germany than in Italy or Spain, referenced in Chapter 9) rather than as primary quantitative evidence.

Chapter conclusion. The academic and institutional evidence reviewed in this chapter does two things for this report's central argument. First, it corroborates the scale and mechanism findings of Chapters 3 through 5 through methodologies entirely independent of this report's own platform- and case-study-level analysis, substantially reducing the risk that this report's conclusions reflect a selection effect in the sources chosen. Second, several strands — the detection-methodology research in particular — provide direct empirical support for specific components of the enforcement-gap diagnosis in Chapter 6, moving that diagnosis from analytical inference to evidenced claim.

Chapter 9: Cross-Country Comparison

9.1 Purpose and interpretive caveat

This chapter compares the ten focus countries specified in this report's brief plus the United Kingdom, not to produce a strict quantitative ranking — the data limitations discussed below make that inappropriate — but to test whether the enforcement-gap pattern identified in Chapter 6 holds across different national legal and reporting environments, or whether it is an artifact specific to the countries most heavily featured in this report's case studies (the UK, Sweden, Denmark).

9.2 Summary comparison

Country	FRA discrimination rate (5-yr, 2021–22 survey)	ODIHR 2024 hate incidents (all bias categories)	Notable incident in report period	Policy response
Austria	71% (highest in FRA survey)	Not separately isolated in ODIHR's 2024 total; Dokustelle recorded 1,522 anti-Muslim attacks in 2023 (66.7% online), a 100%+ year-on-year increase in victim-reported cases (Chapter 3.3)	Dokustelle's online/offline breakdown is this report's most granular national dataset	—
Germany	68%	305 total incidents across all bias categories; ~10% multiple/overlapping bias motivations	—	—
Finland	63%	Not isolated in sources reviewed	Subject of Ünlü et al. (2026) (Chapter 8.1), finding rising anti-Muslim content 2018–2023 across 32 thematic clusters	Quran burning illegal (unlike Sweden/Denmark/Norway)
France	Not among the three highest FRA rates; high overqualification/labour-market effects reported nationally	332 incidents (all categories); ODIHR noted incidents including harassment of women and girls linked to visible Muslim dress	OIC report identified France among the countries with the strongest observed upward trend, Feb 2023–Apr 2024	Face-covering ban (subject of separate EqualFaith research)
Belgium	Included in FRA survey (rate not isolated)	Only 18 incidents reported for 2024 — a figure ODIHR itself flagged as strikingly low, alongside a mid-2024 national police recording methodology revision	—	Face-covering ban (subject of separate EqualFaith research)
Netherlands	Included in FRA survey (rate not isolated)	**Zero** anti-Muslim submissions received for 2024 — explicitly flagged by ODIHR as a probable data gap rather than an absence of incidents	November 2024 Amsterdam football-related riots (Ajax–Maccabi Tel Aviv) illustrate a comparable online-to-offline escalation pattern for a different bias category (antisemitism)	New Article 44bis of the Dutch Criminal Code (July 2025) increases sentencing for discriminatory crimes
Sweden	Not among FRA's three highest	Not isolated in sources reviewed	2023 Quran-burning controversy (Chapter 10.3); Salwan Momika killing (Jan. 2025)	Religious-text desecration ban considered but not passed; raised terror threat level; NATO accession complications
Denmark	—	Not isolated in sources reviewed	2023 Quran-burning protests outside Copenhagen embassies and a mosque during Friday prayers	Passed legislation criminalising "inappropriate treatment" of religious texts — a direct policy reversal prompted by the diplomatic fallout
Italy	Included in FRA survey (rate not isolated)	Not isolated in sources reviewed	OIC report identified Italy (with Spain) as showing a comparatively weaker upward trend than the UK, France, and Germany	—
Spain	Included in FRA survey (rate not isolated)	2023 data grouped anti-Muslim and anti-Christian hate crimes in a combined category	OIC report identified Spain (with Italy) as showing a weaker relative trend	Face-covering ban (subject of separate EqualFaith research); Strategic Framework against racism and xenophobia (2023–2027)
United Kingdom (non-EU, included per report scope)	Not FRA-surveyed	858 incidents (all categories) — but over 80% antisemitic, reported by a single organisation (CST); not a reliable anti-Muslim-specific indicator without further disaggregation	2024 Southport riots (Chapter 5.1), this report's central case study	Public inquiry into the Southport attack, with calls to extend its scope to the riots and disinformation dynamics

9.3 What the comparison demonstrates

Three findings from this table directly substantiate the enforcement-gap argument developed in Chapter 6, beyond simply describing national variation.

The measurement gap is itself cross-nationally variable, which is diagnostic in its own right. The Netherlands' zero anti-Muslim submissions and Belgium's 18-incident aggregate sit alongside Austria's much more granular, disaggregated Dokustelle dataset — not because anti-Muslim incidents are genuinely absent in the Netherlands or Belgium, but because national reporting infrastructure varies enormously in maturity. This is Component two of the enforcement gap (Chapter 6.2) demonstrated cross-nationally rather than asserted in the abstract: the same EU legal framework produces radically different visibility of the same underlying problem depending on national implementation capacity, independent of the underlying rate of discrimination.

Legal responses to the same underlying phenomenon (religious-text desecration) diverged sharply between adjacent, similarly situated states — illegal in Finland, newly restricted in Denmark following acute diplomatic fallout, still permitted with active debate in Sweden. This demonstrates that "online Islamophobia" policy responses are not converging toward a single EU standard even where the underlying platform-governance framework (the DSA) is fully harmonised — a finding with direct relevance to Chapter 7's argument that operational guidance, not just legal text, determines practical outcomes.

The UK case requires a specific caution this report treats as methodologically important. The UK's large ODIHR aggregate figure (858 incidents) is driven overwhelmingly by antisemitic reporting from a single well-resourced civil-society organisation, not by anti-Muslim-specific data. Reading this raw aggregate as an anti-Muslim indicator would produce a misleading cross-country comparison — precisely the kind of measurement artifact Component two of the enforcement gap warns against, and a caution this report applies to its own data rather than only to others'.

Chapter conclusion. The enforcement gap identified in Chapter 6 is not a phenomenon specific to any single country's legal or platform environment; it recurs, in different specific forms, across national contexts with otherwise quite different legal traditions, media environments, and Muslim community sizes. This cross-national consistency is itself evidence that the gap's primary drivers are EU-level and platform-level — definitional fragmentation in the underlying legal concept, platform-level measurement and access restrictions — rather than idiosyncratic to any one member state's implementation failures, reinforcing this report's focus on DSA-level rather than purely national remedies in Chapter 13.

Chapter 10: Quantitative Evidence and Platform Transparency Data

This chapter consolidates the report's core quantitative findings into a single reference point and adds the platforms' own DSA-mandated transparency-report data not fully developed elsewhere. Figures already presented and analysed in Chapters 3, 4, and 5 are tabulated here for reference rather than re-argued.

10.1 Civil-society enforcement-failure data: the CCDH baseline

Using the platforms' own user-facing reporting tools, the Center for Countering Digital Hate flagged 530 posts containing anti-Muslim hate speech, collectively viewed at least 25 million times, across five platforms. Reported failure-to-act rates: YouTube 100%, Twitter 97%, Facebook 94%, Instagram 86%, TikTok 64% — an aggregate cross-platform failure rate of 89%. CCDH additionally documented 23 Facebook groups "mainly or wholly dedicated to anti-Muslim hatred," combined following exceeding 320,000, none removed despite reporting; and found platforms failed to act on 89% of posts promoting the "Great Replacement" conspiracy theory — the same narrative documented as a formal political talking point in Chapter 5.3.

As established in Chapter 11, this 89% figure should be read as a floor on failure rates for clearly identifiable content — reflecting CCDH's non-random, manually curated sampling methodology — rather than a representative prevalence estimate. Its evidentiary value lies in corroboration: it is independently consistent with the Meta Oversight Board's binding Southport findings (Chapter 4.3) and the Commission's own preliminary DSA findings (Chapter 4.1, 4.2), three entirely independent sources converging on the same underlying conclusion — that platform stated policy and platform practice diverge substantially.

10.2 Reference tables: scale and discrimination data

(Consolidated from Chapters 3 and 5; see those chapters for full analytical discussion.)

Metric	Platform	Figure	Source
Increase in discriminatory posts targeting Muslims, 7–8 Oct 2023 vs. prior 2 days	X	+422%	ISD
Increase in anti-Muslim posts, 7–13 Oct 2023 vs. prior week	X	+250%	ISD
Increase in anti-Muslim comments following 7 Oct attack	YouTube	43-fold	ISD
Overall online toxicity increase, Jan–Sep 2023	Cross-platform	+30%	EOOH
Views received by a single previously-banned figure's posts, two weeks post-Southport	X	580 million+	Amnesty International

Metric	Figure	Source
Muslims reporting racial discrimination in preceding 5 years, EU average	47% (up from 39% in 2016)	FRA
Muslims overqualified for current employment	41%	FRA
Austria: anti-Muslim attacks, 2023, share online	1,522 incidents; 66.7% online	Dokustelle

10.3 The Quran-burning controversies: a comparative causal pattern

Beginning January 2023, Danish-Swedish activist Rasmus Paludan burned a Quran outside Turkey's Stockholm embassy; the controversy escalated when Iraqi refugee Salwan Momika burned Quran pages, combined with placing bacon on the text, outside Stockholm's central mosque on 28 June 2023, deliberately timed to coincide with Eid al-Adha. Further burnings occurred outside Copenhagen embassies, including one held directly opposite a mosque during Friday prayers. Swedish authorities identified an active online amplification dimension distinct from the physical protests: intelligence services intercepted online chatter in which groups researchers characterised as terrorist organisations discussed violent responses targeting Sweden, Denmark, and the Netherlands specifically.

The diplomatic consequences were severe and rapid: Iraq expelled Sweden's ambassador and severed diplomatic ties after protesters stormed and set fire to Sweden's Baghdad embassy on 19 July 2023; Sweden's Prime Minister described the situation as the country's most serious security crisis since the Second World War; the controversy directly complicated Sweden's NATO accession given Turkey's objections; and both Sweden and Denmark moved toward legislating against religious-text desecration, with Denmark's bill passing.

Analytical significance. This case illustrates the reverse causal direction from Southport (Chapter 5.1): rather than online disinformation preceding offline violence, a real-world provocative act generated its own large-scale online amplification and cross-border security consequences. Read alongside Southport, it demonstrates that the online-offline harm pathway this report documents operates in both directions — online content driving offline consequences, and offline acts driving online amplification with its own severe consequences — reinforcing Chapter 6's argument that the enforcement gap is a systemic feature of the online-offline information ecosystem as a whole, not a one-directional pipeline addressable through content moderation alone.

10.4 Platform self-reported transparency data

(Full analysis in Chapter 4.4; figures reproduced here for reference.) X's own DSA disclosures: 238,000 illegal-content reports received, 115,000

(~48%) actioned, median 2.7-hour resolution, 667 appeals (~0.58% rate), 190 overturned (~0.17%). Meta's disclosures, per INACH's independent analysis: Facebook received 113,638 notices (16,052 removals); Instagram received 61,888 notices (12,418 removals); hate speech is not reported as a distinct legal category. CCDH's analysis of Meta's own data found proactive enforcement had accounted for over 97% of actions in the categories affected by the January 2025 policy reversal, correctly actioning approximately 277 million pieces of content in those categories beforehand.

A pre-DSA baseline from the European Commission's 2021–2022 Code of Conduct monitoring shows deteriorating platform performance in the period immediately preceding this report's core window: Twitter's 24-hour hate-speech review rate fell from 82% to 51%, and its removal rate fell from 49.8% to 45.4% — occurring in the period directly following the platform's 2022 ownership change, providing useful longitudinal context for the "freedom of speech, not reach" philosophy discussed in Chapter 4.1.

Chapter conclusion. The quantitative evidence consolidated in this chapter should be read as corroboration and reference material for the analytical arguments developed in Chapters 3 through 7, not as a separate body of findings. Its cumulative weight — independent civil-society testing, platform self-disclosure, and pre-DSA regulatory monitoring all pointing toward the same conclusion of substantial, persistent enforcement shortfall — is what allows this report to treat the enforcement gap identified in Chapter 6 as an evidenced diagnosis rather than an interpretive framing.

Chapter 11: Evidence Quality Assessment

This chapter assesses, source by source, the reliability, methodology, and potential bias of the evidence underpinning this report, consistent with the source-tiering approach set out in the Methodology. It is presented as a discrete chapter, rather than embedded as caveats throughout, so that policymakers and reviewers can assess this report's evidentiary foundation independently of its analytical arguments.

Source	Reliability	Methodology	Potential bias	Institutional status
FRA "Being Muslim in the EU" (2024)	High	Large-sample (n=9,604), stratified 13-country survey	Low; established EU-agency survey methodology, standard self-report limitations apply	Official EU agency
ODIHR Hate Crime Data	Moderate–High	Aggregates official state statistics and civil-society submissions	State-reporting variance (several states submit incomplete data, self-flagged by ODIHR) is the primary limitation	Official OSCE body
Institute for Strategic Dialogue (ISD)	Moderate–High	Mixed qualitative/quantitative platform monitoring	Anti-extremism institutional mission; methodologically transparent, not independently audited in sources reviewed	Established think tank, widely cited in government and academic contexts
Center for Countering Digital Hate (CCDH)	Moderate	Manual identification, platform-tool submission (non-random sample)	Advocacy orientation; produces a floor estimate on identifiable content, not a representative prevalence figure	Registered non-profit; core findings not substantively rebutted in sources reviewed despite industry pushback on methodology
Meta Oversight Board decisions	High for adjudicated facts	Formal case review, published reasoning	Funded via a Meta-established independent trust — a structural relationship worth noting even though decisions are binding and have repeatedly ruled against Meta	Quasi-judicial body
European Commission enforcement findings	High	Formal investigation with adversarial due process (right of reply exercised by both platforms)	Low direct bias risk; enforcement priorities may reflect resourcing and sequencing choices rather than the full universe of violations	Official EU regulatory body; legally binding once finalised
Peer-reviewed academic sources (Chapter 8)	High	Varies by study; documented, replicable methodologies	Standard academic limitations disclosed within each study	Published in peer-reviewed journals
National documentation centres (e.g., Dokustelle Austria)	Moderate–High for national context; not cross-nationally comparable	Documentation of reported/registered cases	Support-organisation orientation typical of this source type; likely undercounts non-reported incidents	National civil-society body; one of few sources offering online/offline-disaggregated data
Platform DSA transparency reports (X, Meta)	Moderate	Statutory self-disclosure under DSA Articles 15/24/42	Direct conflict-of-interest risk: platforms report their own performance; categorisation choices (e.g., Meta's absent hate-speech category) may reflect motivated design	Legally mandated but self-reported; a floor on platform-favourable framing, not an independently audited record
European Observatory of Online Hate (EOOH)	Moderate–High	Automated AI toxicity classification across a near-comprehensive sample	EU-funded; standard automated-classifier limitations apply	EU-funded research project; methodology publicly documented, cited by the European Parliament's own research service
OIC "Sixteenth Report on Islamophobia"	Lower for cross-referencing outside its own framework	Not detailed in sources reviewed	Intergovernmental body representing Muslim-majority states; likely advocacy orientation given institutional mandate	Included for completeness; weighted as a stakeholder rather than neutral-monitor source
Journalistic sources (Reuters, AP, BBC, Guardian, Al Jazeera, CNN, etc.)	High for factual/event reporting	Standard journalistic sourcing	Outlet-specific editorial variance; used primarily for verifiable factual sequences rather than analytical claims	Industry-standard verification practices apply

Interpretive principle applied throughout this report. Where a single-tier source's finding stands alone, it is presented with the caveats in this table attached explicitly. Where findings from two or more independent tiers converge on the same conclusion — as in the three-way corroboration of the October 2023 spike (Chapter 3.2), or the independent confirmation of Southport-related platform failure by both civil-society tracking and the binding Meta Oversight Board ruling (Chapter 4.3) — this report treats the convergence itself as adding evidentiary weight beyond what any single source would carry alone, and states so explicitly rather than presenting convergent findings with the same hedging as single-source claims.

Chapter 12: Counterarguments and Alternative Perspectives

Consistent with this report's commitment to rigour, this chapter addresses the strongest available counterarguments to its central thesis, rather than treating the enforcement-gap diagnosis as immune from legitimate challenge.

12.1 The definitional challenge

As Chapter 2 established, the APPG/Runnymede racialisation framing of Islamophobia has been explicitly rejected by some Muslim organisations, including FOSIS, on the grounds that it under-emphasises the specifically religious dimension of the hostility in question. This is not a peripheral dispute: because different definitions produce different classifier outputs (Chapter 8.1) and different eligibility criteria for legal protection (Chapter 2.3), some portion of the quantitative variance across the studies cited in this report may reflect definitional choice rather than underlying prevalence difference. This report has attempted to manage this risk by favouring convergent, cross-methodology findings (Chapter 3.2) over single-source claims, but it does not claim to have resolved the underlying definitional dispute, which remains a live and legitimate area of scholarly and political disagreement this report's recommendations (Chapter 13) do not presume to settle.

12.2 The free-expression challenge

The Quran-burning case study (Chapter 10.3) and the Charlie Hebdo case (Chapter 5.2) together illustrate a genuine, unresolved tension this report's evidence does not resolve. Sweden's and Denmark's historically strong free-expression protections permitted conduct that generated severe diplomatic and security consequences, prompting policy reversals some free-expression advocates have characterised as concerning capitulation to external pressure rather than legitimate domestic hate-speech regulation. The DSA's own hate-speech framework (Chapter 2.4) is explicitly bounded by the *illegal* hate-speech standard rather than a broader harm standard, reflecting a deliberate EU-level judgment that not all offensive or provocative religious-identity-related content should be subject to platform removal obligations.

This tension extends to DSA enforcement itself. Industry and some civil-liberties organisations have argued that DSA enforcement — including the X fine analysed in Chapter 4.1 — risks becoming a vector for politically contested speech regulation, a concern voiced explicitly by U.S. officials in response to the December 2025 fine, notwithstanding that the specific fine addressed transparency rather than content standards directly. This report's recommendations in Chapter 13 are framed to respect this boundary: they target transparency, researcher access, and procedural safeguards rather than proposing an expanded substantive definition of prohibited speech, precisely because the evidence in this chapter shows the latter to be genuinely contested terrain this report is not positioned to resolve.

12.3 The evidentiary-limitation challenge

This report's own Methodology discloses that its account of algorithmic amplification is correlational rather than causal, given restricted access to platform-internal recommender-system data. A critical reader could reasonably argue this limitation undermines the strength of the amplification claims made in Chapters 5 and 7. This report's response, developed fully in Chapter 6.2 (Component three), is that this limitation is itself part of the evidenced finding, not merely a caveat weakening it: the Commission's own adjudicated findings that both platforms studied have obstructed researcher access mean the evidentiary gap is not a research-design shortfall this report could have closed with more effort, but a documented regulatory violation independently established by the EU's own enforcement authority. This reframes the limitation from a weakness in this report's method to further substantiation of its central argument — though this report acknowledges that a reader who does not accept this reframing would reasonably discount the amplification claims in Chapters 5 and 7 to correlational rather than causal status, which is the status this report itself assigns them throughout.

12.4 The cross-category comparison challenge

Chapter 9.2's inclusion of comparative hate-crime figures for antisemitic and anti-Christian incidents raises a legitimate methodological caution this report addresses directly rather than eliding: differential reporting infrastructure, willingness to report, and national legal recognition vary substantially by bias category, meaning raw incident counts across categories (Muslim, Jewish, Christian) are not straightforwardly comparable measures of relative prevalence or severity. This report includes these comparative figures for evidentiary completeness and to situate its specific focus within the broader landscape of religiously motivated hate crime reporting in Europe — not to argue for equivalence or non-equivalence between categories, a claim the evidence reviewed does not support and this report does not make.

12.5 The CCDH methodology challenge

As Chapter 11 notes, CCDH's "Failure to Protect" methodology — manual identification and reporting-tool submission of a curated, non-random sample — has drawn industry criticism disputing whether its 89% aggregate failure figure generalises to platforms' overall content-moderation performance across the much larger universe of anti-Muslim content never manually curated and submitted by CCDH researchers. This report treats the CCDH figures, throughout, as evidence of failure on clearly identifiable content specifically — a floor rather than a representative average — consistent with the caveat applied consistently since its first introduction in Chapter 10.1, rather than as a general prevalence or overall-performance statistic that would require a different, randomised sampling methodology to establish with confidence.

Chapter conclusion. None of the counterarguments addressed in this chapter overturn this report's central thesis, but each meaningfully qualifies its scope. The enforcement-gap diagnosis in Chapter 6 should be read as strongest where this report's evidence shows convergence across independent sources and methodologies (the October 2023 spike, the Southport causal chain, the researcher-access obstruction finding), and as more provisional where it rests on single-source or definitionally contested evidence (the precise magnitude of cross-country prevalence differences, the boundary between legitimate satire and religiously targeted provocation). Chapter 13's recommendations are calibrated to this distinction: the strongest, most specific recommendations target the areas of strongest evidentiary convergence.

Chapter 13: Conclusions and Recommendations

13.1 What this report has established

This report set out to test a specific thesis: that the principal challenge facing Europe in addressing anti-Muslim discrimination is no longer the absence of legal protection, but a persistent enforcement gap between that protection and its practical operation. The evidence assembled across twelve preceding chapters supports this thesis on three independent grounds.

First, the *scale* evidence (Chapter 3) shows discrimination and online hostility toward Muslims rising across a baseline period (2016–2022) that predates any single trigger event, and escalating sharply around identifiable events (October 2023, Southport) in a pattern corroborated by three independent measurement methodologies. Second, the *platform-conduct* evidence (Chapters 4 and 5) shows that where failures have been documented — the Southport case above all — they trace to systemic, institutional causes (delayed crisis protocols, deliberately loosened content policy, restricted researcher verification) rather than isolated moderation errors, and that completed regulatory enforcement to date has addressed procedural symptoms rather than this systemic root cause. Third, the *structural* evidence (Chapters 6 through 10) identifies four specific, addressable components of the enforcement gap — definitional fragmentation, measurement invisibility, restricted verification access, and capture-vulnerable protective mechanisms — each independently evidenced and each with a distinct legal or regulatory remedy.

13.2 The strongest findings, and their evidentiary basis

Three findings in this report rest on the strongest and most convergent evidence available, and should be treated by policymakers as established rather than contested. The October 2023 spike in anti-Muslim online content (Chapter 3.2) is corroborated by three independent methodologies. The Southport case (Chapter 5.1) offers the clearest documented causal chain from false claim to platform amplification to binding institutional finding of moderation failure to offline violence identified anywhere in this report's research process. And the researcher-access obstruction finding (Chapters 4.4, 6.2, 7.4) is now adjudicated regulatory fact, independently established by the European Commission against both platforms studied, not a civil-society allegation this report has had to take on trust.

13.3 Remaining research gaps and forward-looking risk

The most significant gap this report identifies is the absence of platform-internal recommender-system data in the public evidence base — a limitation this report has argued is itself diagnostic of the enforcement gap rather than merely a constraint on this report's own analysis. A second gap is the absence of comprehensive, religion-disaggregated EU equality data infrastructure comparable to what exists for other protected categories. A third is the underdeveloped academic literature specifically evaluating recommender systems, as opposed to organic engagement dynamics or classifier accuracy, for anti-Muslim content amplification.

Two forward-looking risks warrant structured monitoring beyond this report's period of study. Meta's January 2025 moderation reversal (Chapter 4.2) is an explicit, self-acknowledged increase in tolerance for harmful content whose measured effect on anti-Muslim content specifically falls outside this report's evidence window. And the emerging large-language-model bias literature (Chapter 8.1) suggests that as generative AI systems become further integrated into moderation and recommendation pipelines, model-level biases could propagate into platform-level outcomes in ways the platform-specific literature reviewed here has not yet captured.

13.4 Recommendations

The following recommendations are organised by priority, drawing on the strength-of-evidence assessment in Section 13.2 and the four-component enforcement-gap diagnosis in Chapter 6. Each is linked explicitly to its supporting evidence rather than presented as a general policy aspiration.

Priority One — Immediate, evidence-backed, achievable through existing legal mechanisms

1. Expand Article 40(12) civil-society research access. This is this report's highest-priority recommendation because it is both the remedy with the strongest evidentiary support (Chapters 4.4, 6.2, 7.4) and a precondition for verifying whether the other recommendations below are working in practice. The Commission's own adjudicated findings that both X and Meta have obstructed researcher access provide direct legal support for extending the reasoning already accepted for election-integrity research to religious-discrimination research, as argued in EqualFaith Europe's legal memorandum on this provision. **Actionable step:** the Commission should issue implementing guidance explicitly recognising qualified civil-society organisations researching systemic religious-discrimination risk as eligible under Article 40(12) on the same basis as electoral-integrity researchers.

2. Mandate disaggregated hate-speech reporting by protected characteristic. Chapter 4.4 documented that Meta's transparency reporting does not isolate "illegal hate speech" as a category at all, making independent verification of religious-discrimination-specific enforcement structurally impossible using platform-published data. **Actionable step:** the Commission's November 2024 Implementing Regulation standardising transparency-report categories, still producing its first full harmonised dataset as of this report's period of study, should be amended or clarified to require disaggregation by protected characteristic, not merely by broad content-category.

Priority Two — Achievable through Commission guidance, not requiring new legislation

3. Issue dedicated Commission guidance on religious-minority systemic risk. Chapter 7.2 established that religious-minority discrimination falls within Article 34's legal scope but lacks the purpose-built operational guidance already issued for electoral risk and minor protection. **Actionable step:** the Commission should issue Guidelines on the Mitigation of Systemic Risks to Religious Minorities, modelled structurally on the existing March 2024 electoral-risk guidelines, developed in consultation with religious-minority civil society organisations including but not limited to EqualFaith Europe.

4. Reform Article 22 trusted-flagger designation criteria. Chapter 7.3 identified the absence of funding and affiliation disclosure requirements as a capture risk undermining the mechanism's practical value to the communities it is meant to protect. **Actionable step:** Digital

Services Coordinators should require funding-source and political-affiliation disclosure as a condition of trusted-flagger designation, following the specific amendments proposed in EqualFaith Europe's briefing EFE/DSA/2026-01.

Priority Three — Monitoring and longer-term evidence-building

5. Establish structured post-2025 monitoring of Meta's Hateful Conduct policy impact on anti-Muslim content specifically. Given the policy reversal documented in Chapter 4.2 and its currently uncaptured downstream effects (Chapter 13.3), Digital Services Coordinators should require Meta to report, as part of its ongoing DSA compliance obligations, specific before/after data on anti-Muslim content prevalence and enforcement rates attributable to the January 2025 policy change — a reporting requirement that would itself help close the disaggregation gap addressed in Recommendation 2.

6. Support the development of EU-wide, religion-disaggregated equality data infrastructure, building on FRA's existing "Being Muslim in the EU" survey instrument and the national-level model demonstrated by Austria's Dokustelle, to close the cross-country measurement gap documented in Chapter 9.3.

13.5 A closing observation on the nature of the gap

This report has deliberately avoided two simpler, more familiar diagnoses: that Europe's problem is insufficient law, or that it is a series of unconnected individual failures by bad actors. The evidence assembled here supports neither. What it supports is a more specific and, this report argues, more actionable diagnosis: a set of legal and institutional protections that are largely adequate in their formal scope, operating through systems — platform, regulatory, and civil-society — whose practical capacity to deliver that protection has not kept pace with either the scale of the harm or the sophistication of its digital mechanisms. That is a narrower problem than "Europe needs new anti-discrimination law," and a more tractable one. The recommendations in this chapter are addressed to closing it.

Bibliography

This bibliography consolidates all sources cited across this report, verified during the research and editorial process. Entries follow APA 7th edition format where full metadata was independently confirmed; two entries are flagged below as requiring further direct-source verification before use in external legal or academic citation, consistent with this report's evidence-quality standards (Chapter 11).

Al Jazeera. (2024, July 31). *Agitators accused of Islamophobia for linking Southport attack to Muslims*. https://www.aljazeera.com/news/2024/7/31/agitators-accused-of-islamophobia-for-linking-southport-attack-to-muslims

Al Jazeera. (2024, August 2). *Southport stabbing: What led to the spread of disinformation?* https://www.aljazeera.com/news/2024/8/2/southport-stabbing-what-led-to-the-spread-of-disinformation

Al Jazeera. (2017, August 23). *Charlie Hebdo draws ire with Barcelona attack cartoon*. https://www.aljazeera.com/amp/news/2017/8/23/charlie-hebdo-draws-ire-with-barcelona-attack-cartoon

Amnesty International. (2025, August 6). *How X's design and policies led to Southport-linked racist violence*. https://www.amnesty.org/en/latest/news/2025/08/xs-design-and-policies/

Bayrakli, E., & Hafez, F. (2024). *European Islamophobia report 2023*. SETA.

Center for Countering Digital Hate. (2022). *Failure to protect: How tech giants fail to act on reports of anti-Muslim hate*. https://counterhate.com/research/anti-muslim-hate/

Center for Countering Digital Hate. (2024, April 11). *Hate pays: How X accounts are exploiting the Israel-Gaza conflict to grow and profit*. https://counterhate.com/wp-content/uploads/2024/04/Hate-Pays_April2024_CCDH_FINAL.pdf

Center for Countering Digital Hate. (2025). *How will Meta's policy changes affect European users?* https://counterhate.com/blog/how-will-metas-policy-changes-affect-european-users/

Christian Science Monitor. (2025, January 8). *How has the Charlie Hebdo attack changed French cartooning?* https://www.csmonitor.com/World/Europe/2025/0108/charlie-hebdo-islam-satire-religion

CNN. (2023, August 2). *Sweden and Denmark consider ban on Quran-burning protests as security fears rise*. https://www.cnn.com/2023/08/02/europe/quran-burning-protest-denmark-sweden-intl/index.html

CNN Business. (2025, January 7). *Meta gets rid of fact checkers and says it will reduce 'censorship'*. https://www.cnn.com/2025/01/07/tech/meta-censorship-moderation

Columbia Journalism Review. (2025, January 7). *The unresolved legacy of the Charlie Hebdo massacre*. https://www.cjr.org/the_media_today/the-unresolved-legacy-of-the-charlie-hebdo-massacre.php

Council of Europe, European Commission against Racism and Intolerance. (2024, June 20). *Annual report on ECRI's activities covering the period from 1 January to 31 December 2023*. https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/annual-reports

Council of Europe, European Commission against Racism and Intolerance. (2024, October 25). *ECRI report on the United Kingdom (sixth monitoring cycle)*. https://rm.coe.int/sixth-report-on-the-united-kingdom/1680b20bdc

Digital Policy Alert. (2025). *European Commission investigation into X's compliance with DSA requirements to address illegal content and disinformation*. https://digitalpolicyalert.org/change/7366

Digital Policy Alert. (2025, October 24). *European Commission investigation into Meta's compliance with DSA requirements to address illegal content and disinformation*. https://digitalpolicyalert.org/change/7405

European Commission. (2024, April 30). *Commission opens formal proceedings against Facebook and Instagram under the Digital Services Act* [Press release]. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2373

European Commission. (2024, May 15). *Commission opens formal proceedings against Meta under the Digital Services Act related to the protection of minors on Facebook and Instagram* [Press release]. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2664

European Commission. (2026). *Commission preliminarily finds Meta in breach of Digital Services Act for failing to prevent minors under 13 from using Instagram and Facebook* [Press release]. https://ec.europa.eu/commission/presscorner/detail/en/ip_26_920

European Commission. (2026). *Commission preliminarily finds the addictive design of Instagram and Facebook in breach of the Digital Services Act* [Press release]. https://ec.europa.eu/commission/presscorner/detail/en/ip_26_1579

European Commission, Directorate-General for Communications Networks, Content and Technology. (n.d.). *How the Digital Services Act enhances transparency online*. https://digital-strategy.ec.europa.eu/en/policies/dsa-brings-transparency

European Digital Media Observatory. (2025, May 7). *How the European far-right is using AI-generated content to engage voters*. https://edmo.eu/publications/how-the-european-far-right-is-using-ai-generated-content-to-engage-voters/

European Network Against Racism. (2024). *Racial discrimination in Europe: ENAR shadow report 2016–2021*.

European Observatory of Online Hate. (2023). *Online hate speech in 2023* [Periodical report]. Indexed via FRA's Anti-Muslim Hatred Database. https://fra.europa.eu/en/databases/anti-muslim-hatred/node/9854

European Observatory of Online Hate. (n.d.). *FAQ: Methodology and toxicity scoring*. https://eooh.eu/faq

European Observatory of Online Hate. (2026, May). *Articles and periodical reports archive*. https://eooh.eu/articles

European Parliamentary Research Service. (2024). *Hate speech and hate crime: Time to act?* European Parliament. https://www.europarl.europa.eu/RegData/etudes/BRIE/2024/762389/EPRS_BRI(2024)762389_EN.pdf

European Union Agency for Fundamental Rights. (2024). *Being Muslim in the EU: Experiences of Muslims*. https://fra.europa.eu/en/event/2024/launching-fras-being-muslim-eu-report

European Union Agency for Fundamental Rights. (2025). *Fundamental Rights Report 2025*. https://fra.europa.eu/en/publication/2025/fundamental-rights-report-2025

European Union Agency for Fundamental Rights. (2025). *FRA updates its anti-Muslim hatred database* [News item]. https://fra.europa.eu/en/news/2025/fra-updates-its-anti-muslim-hatred-database

European Union Agency for Fundamental Rights. (n.d.). *Database 2012–2024 on anti-Muslim hatred*. https://fra.europa.eu/en/databases/anti-muslim-hatred

European Union Agency for Fundamental Rights, FRANET research network. (2022, 2023). *Documentation Centre Austria (Dokustelle): Racist attacks against Muslims and persons perceived as Muslim* [2022 and 2023 data]. https://fra.europa.eu/en/databases/anti-muslim-hatred/node/9689; https://fra.europa.eu/en/databases/anti-muslim-hatred/node/9690

Euronews. (2026, May 27). *Online hate speech: Who faces the most online toxicity in Europe?* https://www.euronews.com/my-europe/2026/05/27/online-hate-speech-who-faces-the-most-online-toxicity-in-europe

Forbes. (2025, April 23). *Meta oversight board urges company to assess 'human rights impact' of hateful conduct policy*. https://www.forbes.com/sites/conormurray/2025/04/23/meta-oversight-board-urges-company-to-assess-human-rights-impact-of-hateful-conduct-policy/

France 24. (2025, January 6). *France's Charlie Hebdo unveils special edition marking 10 years since deadly attack*. https://www.france24.com/en/live-news/20250106-charlie-hebdo-unveils-special-edition-10-years-since-attack

France 24. (2025, January 7). *France marks 10 years since the Charlie Hebdo attacks*. https://www.france24.com/en/live-news/20250107-france-to-remember-charlie-hebdo-attacks-10-years-on

Ghasiya, P., & Sasahara, K. (2022). Rapid sharing of Islamophobic hate on Facebook: The case of the Tablighi Jamaat controversy. *Social Media + Society, 8*(4). https://doi.org/10.1177/20563051221129151

Goodwin Procter. (2025, December). *EC issues first non-compliance fine under the DSA: X fined €120 million for dark patterns and transparency failures*. https://www.goodwinlaw.com/en/insights/publications/2025/12/alerts-practices-antc-ec-issues-first-non-compliance-fine-under

Harneys. (2024). *EU Commission's preliminary findings: X in breach of Digital Services Act*. https://www.harneys.com/our-blogs/regulatory/eu-commission-s-preliminary-findings-x-in-breach-of-digital-services-act/

Human Rights Campaign. (2025, January 27). *Meta's new policies: How they endanger LGBTQ+ communities and our tips for staying safe online*. https://www.hrc.org/news/metas-new-policies-how-they-endanger-lgbtq-communities-and-our-tips-for-staying-safe-online

Human Rights Watch. (2025). *World report 2025: European Union*. https://www.hrw.org/world-report/2025/country-chapters/european-union

Humboldt Institute for Internet and Society, Digital Society Blog. (2025, December 10). *The DSA's transparency reports*. https://www.hiig.de/en/analysis-of-the-dsas-transparency-reports/

IAPP. (2026). *European Commission fines X 120M euros for DSA violations*. https://iapp.org/news/a/european-commission-fines-x-120m-euros-for-dsa-violations

Information Technology and Innovation Foundation. (2025, October 20). *EU should improve transparency in the Digital Services Act*. https://itif.org/publications/2025/10/20/eu-should-improve-transparency-in-the-digital-services-act/

Institute for Strategic Dialogue. (2023). *Islamophobia and anti-Muslim hatred* [Explainer]. https://www.isdglobal.org/isd-explainer/islamophobia-and-anti-muslim-hatred/

Institute for Strategic Dialogue. (2026). *EU Digital Services Act* [Explainer]. https://www.isdglobal.org/isd-explainer/eu-digital-services-act/

Institute for Strategic Dialogue. (2026). *From rumours to riots: How online misinformation fuelled violence in the aftermath of the Southport attack*. https://www.isdglobal.org/digital-dispatch/from-rumours-to-riots-how-online-misinformation-fuelled-violence-in-the-aftermath-of-the-southport-attack/

International Network Against Cyber Hate. (2025). *Transparency reports 2024/2025: DSA analysis*. https://www.inach.net/wp-content/uploads/Transparency-Reports-2024_2025-DSA-2-1.pdf

Kasahara, Y., Schroeder, D. T., Yazidi, A., & Lind, P. G. (2026). The dynamics of hate speech: Assessing anti-Muslim hate speech in Norwegian social media. https://doi.org/10.1177/08944393261417730

Lajevardi, N., Oskooii, K. A. R., & Walker, H. (2022). Hate, amplified? Social media news consumption and support for anti-Muslim policies. *Journal of Public Policy, 42*(4), 656–683. https://doi.org/10.1017/S0143814X22000083

Meta Platforms Ireland Limited. (2024–2025). *EU Digital Services Act: Transparency reports for very large online platforms* [Facebook and Instagram]. https://transparency.meta.com/reports/regulatory-transparency-reports/

Middle East Eye. (2024, August 6). *UK: Anti-Muslim mob attacks Southport mosque after misinformation campaign*. https://www.middleeasteye.net/news/uk-far-right-attacks-southport-mosque-after-misinformation-campaign

New Arab. (2025, November 22). *AI and Islamophobia: The algorithmic gaze on Muslim communities*. https://www.newarab.com/analysis/ai-and-islamophobia-algorithmic-gaze-muslim-communities

NPR. (2025, January 7). *Meta says it will end fact-checking as Silicon Valley prepares for Trump*. https://www.npr.org/2025/01/07/nx-s1-5251151/meta-fact-checking-mark-zuckerberg-trump

Online Hate Prevention Institute. (2024). *Online anti-Muslim hate and racism against Palestinians and Arabs: October 2023 – February 2024*. https://ohpi.org.au/afteroctober/

Organisation for Security and Co-operation in Europe, Office for Democratic Institutions and Human Rights. (2025, November 16). *2024 hate crime report*. https://hatecrime.osce.org/hate-crime-report

OSCE/ODIHR. (2025). *Report data — Germany, 2024*. https://hatecrime.osce.org/reporting/germany/2024

OSCE/ODIHR. (2025). *Report data — France, 2024*. https://hatecrime.osce.org/reporting/france/2024

OSCE/ODIHR. (2025). *Report data — Belgium, 2024*. https://hatecrime.osce.org/reporting/belgium/2024

OSCE/ODIHR. (2025). *Report data — Netherlands, 2024*. https://hatecrime.osce.org/reporting/netherlands/2024

OSCE/ODIHR. (2025). *Report data — United Kingdom, 2024*. https://hatecrime.osce.org/reporting/united-kingdom/2024

OSCE/ODIHR. (2024). *Report data — Spain, 2023*. https://hatecrime.osce.org/reporting/spain/2023

OSCE/ODIHR. (n.d.). *Anti-Muslim hate crime* [Topic overview]. https://hatecrime.osce.org/anti-muslim-hate-crime

Oversight Board. (2025). *Wide-ranging decisions protect speech and address harms*. https://www.oversightboard.com/news/wide-ranging-decisions-protect-speech-and-address-harms/

Religion Media Centre. (2022, April 13). *Factsheet: Defining Islamophobia*. https://religionmediacentre.org.uk/factsheets/islamophobia-defined/

RNZ. (n.d.). *Social media giants failing to combat 'blatant and easy to find' anti-Muslim hate speech*. https://www.rnz.co.nz/news/world/466092/social-media-giants-failing-to-combat-blatant-and-easy-to-find-anti-muslim-hate-speech

Rosen, Z. P., & Walther, J. B. (2025). Social processes in the intensification of online hate: The effects of verbal replies to anti-Muslim and anti-Jewish posts following 7 October 2023. https://doi.org/10.1177/20563051251383635

RSIS (S. Rajaratnam School of International Studies). (2024). *From disinformation to violence: The UK far right and 2024 riots*. https://rsis.edu.sg/ctta-newsarticle/from-disinformation-to-violence-the-uk-far-right-and-2024-riots/

SBS English. (2024, August 6). *Anti-immigration riots are escalating in the UK. Here's how misinformation spurred violence*. https://www.sbs.com.au/language/english/en/article/anti-immigration-riots-are-escalating-in-the-uk-heres-how-misinformation-spurred-violence/klg25tml4

Society for Computers & Law. (2026). *European Commission preliminarily finds Meta in breach of DSA*. https://www.scl.org/european-commission-preliminarily-finds-meta-in-breach-of-dsa/

TechPolicy.Press. (2025, January). *Meta's content moderation changes are going to have a real world impact. It's not going to be good.* https://www.techpolicy.press/metas-content-moderation-changes-are-going-to-have-a-real-world-impact-its-not-going-to-be-good/

TechPolicy.Press. (2025, December 16). *Platforms report to EU regulators under DSA with an eye on US politics*. https://www.techpolicy.press/platforms-report-to-eu-regulators-under-dsa-with-an-eye-on-us-politics/

The Conversation. (2025, November 26). *Why UK's working definition of Islamophobia as a 'type of racism' is a historic step*. https://theconversation.com/why-uks-working-definition-of-islamophobia-as-a-type-of-racism-is-a-historic-step-107657

The Organization for World Peace. (2025, August 1). *The instrumentalization of political Islam by Europe's far right*. https://theowp.org/reports/the-instrumentalization-of-political-islam-by-europes-far-right/

Time. (2023, August 17). *Why Quran burning is making Sweden and Denmark so anxious*. https://time.com/6303348/quran-burning-sweden-denmark/

UK Parliament, Hansard. (2024, January 9). *Definition of Islamophobia*. https://hansard.parliament.uk/commons/2024-01-09/debates/24010969000020/DefinitionOfIslamophobia

UK Parliament, Home Affairs Committee. (2025, March). *Written evidence: Gendered Islamophobia — Evidence of disproportionate targeting of Muslim women and girls*. https://committees.parliament.uk/writtenevidence/138836/pdf/

Ünlü, A., Truong, S., Sawhney, N., Tammi, T., & Kotonen, T. (2026). From prejudice to marginalization: Tracing the forms of online hate speech targeting LGBTQ+ and Muslim communities. *New Media & Society*. https://doi.org/10.1177/14614448241312900

United Nations Regional Information Centre for Western Europe. (2025, March 14). *EU: One in two Muslims are victims of discrimination in daily life*. https://unric.org/en/eu-one-in-two-muslims-are-victims-of-discrimination-in-daily-life/

Vice. (2024, August 9). *There are dozens of Facebook groups intentionally spreading Islamophobia. Meta refuses to act.* https://www.vice.com/en/article/meta-facebook-islamophobia-groups-spread/

Vidgen, B., & Yasseri, T. (2020). Detecting weak and strong Islamophobic hate speech on social media. *Journal of Information Technology & Politics, 17*(1), 66–78.

Wikipedia. (n.d.). *2023 Quran burnings in Sweden*. https://en.wikipedia.org/wiki/2023_Quran_burnings_in_Sweden

Wikipedia. (n.d.). *2024 United Kingdom riots*. https://en.wikipedia.org/wiki/2024_United_Kingdom_riots

Wilson Center. (2023, August 9). *Why Sweden and Denmark react differently to Quran desecrations*. https://www.wilsoncenter.org/article/why-sweden-and-denmark-react-differently-quran-desecrations

X Corp / Twitter International Unlimited Company. (2024). *DSA Transparency Report — April 2024*. https://transparency.x.com/dsa-transparency-report-2024-april.html

X Corp / Twitter International Unlimited Company. (2025). *DSA Transparency Report — April 2025*. https://transparency.x.com/dsa-transparency-report-2025-april.html

X Corp / Twitter International Unlimited Company. (2024). *Report setting out the results of the systemic risk assessment (TIUC DSA-SRA 2024)*. https://transparency.x.com/content/dam/transparency-twitter/dsa/dsa-sra/dsa-sra-2024/TIUC-DSA-SRA-Report-2024.pdf

[Authors not individually attributed in source]. (2023). Enhancing automated hate speech detection: Addressing Islamophobia and freedom of speech in online discussions. In *Proceedings of the 2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. https://dl.acm.org/doi/10.1145/3625007.3627487

Flagged for further direct-source verification before external legal or academic citation:

- The specific Amnesty International 2022/2023 finding on disproportionate targeting of Muslim women, cited in this bibliography via a secondary source (UK parliamentary written evidence, above) rather than Amnesty's original publication directly.
- The European Observatory of Online Hate's underlying primary report documents (beyond the FRA index page and EOOH's own published summaries cited above) were not independently retrieved and cross-checked against the FRA-indexed summary figures used in this report.

End of report.