

April, 2026

Hiding in Plain Sight: the Conceptualisation and Operationalisation of Online Philosemitism.

Gabriella Cohen

Master's Thesis supervised by Sylvain Parasie

Second member of the Jury: Nadia Marzouki

Master in Public Policy
Digital, New Technology

Contents

Main Contribution	3
Introduction	4
Chapter One: Establishing the Study	9
The Current State of Knowledge	9
Introduction	9
Conceptualising Philosemitism	9
Online Antisemitism	14
Cognitive Perspectives	16
Automated Detection	18
Methodology	20
Data Collection: Interviews	20
Data Collection: Social Media	21
Chapter Two: The Mechanism of Online Philosemitism	30
Category One: The Essentialising Mechanism	30
Four Sub-Types of Essentialised Praise	30
Cultural and Intellectual Praise	30
Unique Possession of Economic Skill and Power	32
A Resilient Community	34
Model Minority Identification	36
Category Two: Instrumental Formations	39
Anti-Muslim Sentiment and the Far-Right	40
The Indian Context	41
Religious Instrumentalising	44
Guilt-Engendered Philosemitism	46
Chapter Three: From Discourse to Detection	50
Category Three: The Authenticity Threshold	50
Defining the Threshold	50
The ‘Authenticity Test’: Four Markers	51
Threshold Instability	52
‘Anti-Antisemitism’ Minimal Threshold	53
Category Four: A Detection Gap	54
The Structural Detection Gap	55
The Challenges for Automation	56
How to Combat these Challenges	58
A Consideration of Platform Architecture	60
Creating an Annotation Framework	61
Limitations and Areas for Future Research	63
Conclusion	65
Policy Recommendations	67
Bibliography	69
Appendices	79

Main Contribution

This master's thesis explores how philosemitism, commonly understood as excessive admiration for Jewish people, operates in online environments and why it remains largely invisible to both researchers and automated detection systems. While antisemitism has been extensively studied, philosemitism has received far less systematic attention. This study addresses that gap by developing conceptualising findings and operationalising them in an annotation guideline to identify and analyse online philosemitism across platforms.

Three reasons for you to read this study are as follows. First, the thesis presents timely insights into online discourse surrounding Jewish people. The results complement a literature gap, connecting offline studies to broader social media environments, demonstrating how philosemitic narratives also operate online. Although some research has examined philosemitic narratives instrumentalised by far-right groups, a more holistic conceptualisation of philosemitism has not yet been developed. This paper does not attempt an exhaustive treatment of the movement of offline philosemitism to online contexts. This would be impossible given the scale of the phenomenon and the vastness of online communication. Instead, it is suggestive of a conceptual framework to be built upon and provides an annotation guideline to operationalise its findings for future empirical research.

Second, the study advances a critical rethinking of how positive discourse about minority groups functions. Drawing from twelve expert interviews and cross-platform social media data, the findings demonstrate that philosemitism frequently reproduces essentialised stereotypes of Jewish identity. These include portrayals of Jews as uniquely intelligent, resilient, moral or powerful. While these representations appear favourable, they rely on the same reductive logic as antisemitism and construct Jewish people as fundamentally distinct. Even well-intentioned support can reinforce this process of 'othering' that is revealed in this conceptualisation of online philosemitism.

Third, the thesis makes a practical contribution by identifying a structural detection gap in current automated systems. This blind spot exists because philosemitic discourse is typically framed in positive or neutral language, which causes it to fall below the threshold of most hate speech detection tools. These tools are primarily calibrated to detect explicitly negative or hostile content and therefore struggle to recognise more positive framings of stereotyping.

The thesis concludes that philosemitism and antisemitism operate within the same broader system of meaning-making. Both construct the 'Imaginary Jew', shaped predominantly by external projections. As such, understanding one requires engaging with the other. On this basis, future educational and monitoring recommendations are proposed to ensure both forms of discourse will be addressed within future definitional, pedagogical and detection frameworks.

Introduction

Q: Which is preferable – the antisemite or the philosemite?

A: The antisemite. At least he isn't lying.¹

Introducing the Issue of Online Philosemitism

The Internet has never known a time without antisemitism. It is long outlived by this ancient form of hatred that has spanned over two thousand years of history (Becker & Allington, 2021, p.1). After the Second World War and the horrors of the Shoah, antisemitism mutated in appearance towards more implicit features which built on traditional antisemitic content that avoided censorship from social stigma and legal restrictions. In the 21st century, this has propagated in online spaces (Becker & Troschke, 2022), operating below any threshold of detection on a daily basis from institutional decision-making bodies to the virtual world (Stenzler-Koblentz, 2026).

Online communication of antisemitism relies heavily on evolving codes and linguistic practices through which users signal meaning and group affiliations (Bossetta, 2020, p.227; Klempner & Stenzler-Koblentz, 2024; Becker, 2024). Mainstream social media users have unknowingly worked in tandem with far-right fringe groups by adopting more coded forms of antisemitism (Finkelstein et al., 2018). These extremist groups use word play, such as irony, humour and double meaning, to formulate in-group identities and normalise racist ideologies in more moderated spaces (Becker et al., 2026). The blurring boundaries between explicit and ambiguous forms of antisemitic expression has made detection for researchers and moderators challenging (Bossetta, 2022, p.230). This thesis explores one element of this coded language: philosemitism.

Philosemitism in its most elementary form is any 'action or sentiment "directed in favour of Jews."' (Lassner & Trubowitz, 2008, p.7). It signifies an inclination to admire Jews, Jewish culture, belief system, ideas and more broadly, Jewishness in a manner 'more than necessary' (Schilling, 2020, p.6). Despite deriving from the Greek *philo*, meaning 'to love', philosemitism in academia is dominated by scepticism about whether non-strategic love for Jews is possible. Crucially, philosemitism does not necessarily erase the history of antisemitism but appears, in some formulations, a disguised form of it (Zubrzycki, 2022, p.195; Sutcliffe & Karp, 2011, p.1). It serves the non-Jews' own agenda (Cohen, 2025; Fischer, 2011; Zubrzycki, 2022) and appears differently depending on its context.

This is what makes philosemitism so difficult to define. It can theoretically be positive - a genuine admiration for Jewish people, culture or tradition surpasses ideological instrumentalisation (Karp, 2011; Levenson, 2004; Speinwall, 2011). Under certain

¹ Karp, J. and Sutcliffe, A. (2011). *Philosemitism in History*. Cambridge: Cambridge University Press. p.1

circumstances, the same expression can become offensive depending on context and positionality. People have their own sensibilities, shaped by their own trajectories as members of insider or outsider groups and these shape how they interpret information (Brewer, 1999). The same words, expressed by a different person or context and read by a different audience can produce entirely different meanings (Goffman, 1974). This is the fine line that lies at the heart of this thesis. Online, where context is stripped away, knowing precisely where the line is becomes evidently more complicated.

What is clear is that philosemitism crosses borders, morphing differently depending on its geographical, historical, cultural and socio-political context. Traced back to Greco-Roman antiquity, it has cultivated itself even beyond the Jewish diaspora, such as South East Asia (Schilling, 2020; Kowner, Ainslie & Podoler, 2023). Philosemitism involves the construction of Jewish people as a distinctive group characterised by particular moral, intellectual, cultural, racial and religious qualities. While such representations may appear complimentary, they frequently rely on essentialised stereotypes that position Jews as fundamentally different from other groups (Lassner & Trubowitz, 2008, p.9). There are multiple distinct types of Jewish cultural appropriation, some unreflective and others more conscious. Where appropriation is unreflective, non-Jews are not considering the impact on Jewish individuals when engaging with any means of a practice or symbols (Zubrzycki, 2022, p.157). This tension lies at the heart of the phenomenon as beneath this linguistic surface of admiration lies a more problematic dynamic that reinforces essentialised assumptions about collective Jewish identity.

Philosemitism is an overlooked notion in the broader landscape of discourse about Jewish people. This is concerning as antisemitism increases following significant political events, such as national elections or crises such as Covid-19 (Finkelstein et al., 2018, p.1). At the same time, the world is becoming more divided in online communication reaching new heights of political polarisation reinforced by platform structures which generate increasingly untrustworthy information ecosystems (Humprecht et al., 2023; Pennycook et al., 2021, 59; Munger., 2020). Discourse around attitudes towards Jewish people has become further convoluted by October 7th, 2023, when Hamas entered Israel and carried out coordinated attacks. This has sparked heated scholarly debates around the terms ‘antisemitism’ and ‘anti-Zionism’ that breach the scope of this thesis but remain important to note contextually. Concomitantly, the rise in right wing populism globally has increased philosemitic discussions of Jewish people from both the left and right sides of the political spectrum (Rose, 2020; Zubrzycki, 2018). It is therefore important to examine how such discourses manifest in mainstream online spaces and whether existing tools of automated detection are equipped to capture their nuances.

A significant portion of non-Jewish people only encounter Judaism through social media (Woo et al., 2025). For many, this is their primary point of contact due to limited exposure to Jewish communities in their daily lives. Without pre-existing frameworks for evaluating religious discourse about minority groups, these encounters shape their understanding in ways that are rarely interrogated. The more people know about Judaism, the warmer they feel

towards Jewish people (Lipka., 2019). In an era of widespread mis- and disinformation, amid the global rise of far-right and sensationalist politics, the risk is that positive linguistic structures can conceal problematic representations of Jewish identity.

The Illusory Truth Effect depicts how repeated exposure to content, regardless of its accuracy, increases its perceived credibility which is acute for young people in the early stages of forming their understanding of Judaism (Udry & Barber, 2024). Almost 90% of young Jewish Europeans identify social media as the ‘context in which antisemitism is most problematic today’ (FRA, 2019). The impact of commentary about Judaism on Jewish identity is well-documented (Becker et al., 2024). However, what is missing is a full comprehension of the benign discussions of the Jewish community - whether philosemitic in sentiment or not. The process of ‘othering’ is a recurring problem in Jewish history. Therefore, engaging in a holistic analysis of online discourse of the Jewish community will display if the same essentialised tropes are reproduced. Understanding how people talk about the Jewish community online, including through excessive admiration, is a precondition for being able to label, monitor and in the future, potentially contest it.

Philosemitism is a new angle for thinking about its assiduous partner, antisemitism (Lassner & Trubowitz, 2008, p.8; Schilling, 2020, p.298). It reveals how seemingly positive representations of Jewish people can still rely on instrumentalising and political motivated narratives. The two phenomena are not antithetical as philosemitism can coexist and reinforce antisemitism making it a critical framework for understanding the complexities of contemporary discourse (Schilling, 2020, p.298; Lassner & Trubowitz, 2008, p.7). Unlike overt antisemitism, where hostile language is a clearer indicator of meaning, philosemitic content online remains largely invisible within both academic research and the automated systems designed to detect online hate speech (Chandra et al., 2021). With limited evidence for philosemitic life online (Rose, 2020), establishing whether philosemitism has translated from its historical existence to the Internet is important for understanding the full spectrum of how Jewish people are discussed in digital spaces.

These considerations give rise to the central question of this thesis: how can online philosemitism be conceptualised and operationalised in order to detect and analyse its presence in contemporary online communication?

The following hypotheses will be tested:

H1. Philosemitism online emerges across a typologically distinct range of discursive formations that include stereotypical praise, theological instrumentalisation and political approbation, all of which do not map neatly onto existing frameworks developed for offline contexts.

H2. The majority of philosemitic discourse identified in online platforms reproduces essentialised stereotypes of Jewish identity, regardless of whether the expressed intent is positive or supportive.

H3. The positive linguistic features of philosemitic discourse results in a systemic detection gap in automated content detection systems which are calibrated to identify explicitly negative or hateful language.

Methodology and Scope

To address these questions, this thesis adopts a mixed-methods approach. The first component consists of a theoretical literature review examining existing scholarship on philosemitism, the monitoring of antisemitism in online environments and the challenges of automated detection tools. This situates the study within broader academic debates about existing conceptualisations of the phenomenon and will inform the subsequent identification and analysis of online philosemitism.

The second component involves qualitative interviews used to create a codebook through the framework Constructivist Grounded Theory (CGT). This primary data will be combined with sampled data from a cross-section of social media profiles suggested by interviewees, alongside a small sample of the Blue Square Alliance's (BSA) dataset and a thread from the Facebook Forum 'where non-Jews can openly ask Jewish people questions'. This forum was selected due to one interviewee being its moderator and gave permission for the data collection. Through an iterative process of coding and category development through the lens of Charmaz (2014), the coded system aims to capture the nuances of philosemitism and its perception by different stakeholders whilst simultaneously being grounded in real online discourse.

The social media data was partially collected to capture the language real people use online when discussing Jewish people and Judaism. Data collection began from 2022 to present day, deliberately allowing for philosemitic discourse prior to October 7th 2023. The analysis is cross-platform, covering Twitter/X, Instagram, Bluesky, Facebook, YouTube, Reddit and LinkedIn. This is an exploratory rather than substantive paper on the conceptualisation of online philosemitism so geographical scope has not been narrowed. The phenomenon operates across borders and any early restriction would risk foreclosing findings before the conceptual groundwork has been established.

The operationalising output of this thesis is an annotation guideline, produced from the conceptual and technical findings across both chapters. An annotation guideline is a structured instrument that enables future researchers to label instances of philosemitism in a reproducible way. The long-term goal is the production of a gold standard dataset for the use in Large Learning Model (LLM) benchmarking experiments and training automated monitoring tools (Pustejovsky & Stubbs, 2013; Waseem & Hovy, 2016; Becker et al., 2026).

This matters because philosemitism's defining instability requires interpretive flexibility for future labelling efforts. The annotation guideline produced here does not attempt to resolve that ambiguity but to make it legible and provide future coders with inclusion/exclusion criteria that records, rather than erases, interpretive disagreement. By incorporating

philosemitic data into training and testing datasets, future detection systems will be better equipped to recognise the full spectrum of discourse surrounding Jewish identity online and assess whether philosemitism functions as a structural signal for hate.

The structure of the thesis reflects these stages of analysis. The first chapter presents the theoretical literature review. The findings are presented across two chapters. Chapter Two examines how philosemitic narratives are framed online through the essentialising and instrumentalising of Jewish people. Chapter Three addresses the operationalising of these findings, the difficulty of authenticating online philosemitism and the technical challenges facing automated monitoring, concluding with the annotation guideline as a resource for future quantitative research on the phenomenon.

Overall, this study contributes to a deeper understanding of philosemitism. It is a phenomenon that remains both conceptually ambiguous and empirically underexplored, particularly in its online formations. Through a multi-layered approach, it highlights the complex ways in which admiration, stereotyping and ideology intersect within online discourse. In doing so, it also raises broader questions about how societies recognise and respond to subtle forms of prejudice that may appear, at first glance, to be expressions of support.

Chapter One

Establishing the Study

The Current State of Knowledge

Introduction

This literature review covers four areas. First, it traces the competing definitions and historical arc of philosemitism from antiquity to present, including its non-Western formations. Second, it situates philosemitism within the broader literature of antisemitism. Third, it examines the current state of automated detection tools and their structural limitations. Finally, cognitive insights are used to explain how biases and perceptual thresholds shape online discourse interpretation. These establish the conceptual and technical gaps this thesis addresses.

Conceptualising Philosemitism

a. A Contested Term

To assess how philosemitism appears online, understanding its unstable etymology is essential as its historical resistance to ossification plays a part of this thesis's puzzle. The term has been contested from the outset. Coined in Germany during the 1880s, philosemitism emerged as an antonym to antisemitism and entered the Oxford English Dictionary in 1914 (Sutcliffe & Karp, 2011, p.1; Rose, 2020, p.4). Cohen asserts its late arrival into formal vocabulary is itself instructive as public political figures who admired Jews in the early twentieth century 'did not need the word to engage in apologetics' referring to the polemic derivation rather than a genuine taxonomy of the term (Cohen, 2025, p.4).

It was almost immediately weaponised when Hitler's right-hand man, Joseph Goebbels, published eight issues of *Der Philosemit* from 1928 in the official newspaper *Der Angriff* as a satirical instrument of antisemitic propaganda (Cohen, 2025, p.5). The term was not widespread and most Europeans did not know the term by 1949 (Roth, 1962, p.10). Only among surviving German Jews did it provide a negative meaning and a posed 'friendly disposition detrimental to Jews (Cohen, 2025, p.8).

b. Competing Terms

The literature on philosemitism is marked by definitional fragmentation, with scholars differing on whether it should be understood as a discrete attitude, a broader process, or rejected altogether as a category. Philosemitic literature is dominated by four texts: Lassner and Trubowitz (2008), Edelstein (1982), Rubenstein and Rubenstein (1999) and Levenson (2004) (Schilling, 2018, p.298). The simplest interpretation is by Lassner and Trubowitz who declare it as any 'action or sentiment "directed in favour of Jews"' (2008, p.7).

Cohen confutes this, advancing a more expansive interpretation. He posits philosemitism goes beyond harbouring admiration for Jews or the creation of a good ‘Imaginary Jew’ and claims it is an ambiguous process involving a ‘spectrum of iterations’. These span from opposition to anti-Jewish prejudice, through Holocaust memory, to compensatory *Judenfreundlichkeit* [friendliness towards Jews] and philo-Zionism (2025, p.20). This difference lies in Cohen’s contextual focus on German and French post-war philosemitism, that elicits the hope of absolution in return for the befriending of Jews and Israel, shifting the focus from static definition to dynamic function (ibid., p.19).

To complicate the phenomenon further, some authors reject the term philosemitism itself, proposing ‘allosemitism’ as a more pertinent replacement. Argued by Zygmunt Bauman and popularised by David Wertheim, ‘allosemitism’ is the legitimising of non-Jewish thought through the “Jew” (Bauman, 1998; Wertheim, 2017a). This reification of Jewish alterity is more objective as it highlights perceived singularity of the Jewish people whether positive or negative (Cohen, 2025, p.19). It is, however, a broad conceptual lens to understand how Jews are used to validate non-Jewish ‘ideas, decisions and exploits’ (Wertheim, 2017a, p.2).

‘Anti-antisemitism’ is an alternative narrower descriptor of reactive, pro-Jewish sentiment though, as Rose (2020), acutely posits, is too limited by only positioning philosemitism in relation to the hatred it opposes. Like Cohen’s iterative spectrum suggests, philosemitism can be both reactive and proactive to a contextually broad range of catalysts explaining the vastness of its formation that emerges around the world. Therefore, anti-antisemitism sits as a sub-category of this overall phenomenon and is insufficient in capturing its dynamics in its entirety.

Despite these divergences, the most cited definition is Karp and Sutcliffe’s ‘*sine qua non* of philosemitism’ that Jews are perceived as ‘resolutely distinct people’ with ‘distinctly admirable characteristics’ (2011, p.13). They discuss the generalities of philosemitism through a list of (sometimes overlapping) criteria useful to establish the phenomenon. They find philosemitism follows a recurring ‘similar cluster of classifications’ that include: ‘economic, utilitarian, millenarian, humanistic, romantic, intellectual, liberal, Christian, and Zionist’ themes (ibid.) The criteria are useful for foreshadowing how Jewish identity may be constructed and operationalised in online environments.

c. Philosemitism’s Historical Arc

Philosemitic attitudes predate modernity considerably. Karp and Sutcliffe trace its roots to Greek and Roman writers, such as Hecataeus of Abdera and Marcus Varro, who regarded Jews as ‘philosophical people’, while similarly Theophrastus writes they are ‘philosophers by birth’ (2011, p.13; Feldman, 1993, pp.201-3). Jewish antiquity connoted cardinal virtues, including courage and temperance but especially wisdom, whilst Moses was analogised with lawmaker figures like Lycurgus of Sparta (Karp and Sutcliffe, p.15). There was a significant religious dimension where non-Jews recognised and sometimes appropriated the power of the Jewish God (Cohen, 1989, pp.15-6). Greco-Jewish literature included stories of gentiles who saw the ‘power of the god of the Jews’ and viewed him as superior (ibid.).

Jewish-Roman authors, Josephus and Augustine, wrote about non-Jews who were labelled 'Judaizers' or 'Godfearers', adopting Jewish rituals, customs and beliefs without converting (Berthlot, 2020, pp.500-1). The Greco-Roman era established an intellectual culture viewing Jews as bookish people with a 'particular identity' and religious superiority.

The Enlightenment and European Romantic period produced a strand of intellectual philosemitism most prominently in Friedrich Nietzsche, who saw Jews in 'superlative terms'. Jews were labelled as 'uniquely resilient, intellectually energetic and historically significant people' (Sutcliffe & Karp, 2011, p.20). Nietzsche expressed public disapproval for antisemitic behaviours of his time; his focus remained on Jews as an intellectual asset. The burst of literature during this time, including Gotthold Ephraim Lessing's *The Jews* (1754) and *Nathan the Wise* (1779), established philosemitism as an aesthetic and cultural marker in post-war liberal discourse. Yet, the period equally confirmed philosemitism as a nexus to antisemitism. The majority praised Jews ambivalently, admiring abstracted Jewish intellect while remaining hostile to actual Jewish communities (Sutcliffe & Karp, 2011, p.21; Cohen 2025, p.4), continuing into the twentieth century.

d. After the Shoah

After the Shoah, literature focuses on guilt, redemption and the lachrymose dimension of philosemitism that pushed a 'love for Jews' that suppressed Jewish voices through instrumentalising Holocaust memory (Cohen, 2025, p.12). For the first time in European history, overt animus against Jews became off-limits in public discourse in Western Europe. This suppression did not eliminate anti-Jewish feeling but redirected it as for non-Jews with 'a lingering Jewish problem', philosemitism offered a new form of expression (Cohen, 2025, p.29).

Post-war philosemitic studies concentrate on West Germany where philosemitism is framed as a guilt mechanism, or 'state philosemitism' (Ozyurek, 2023; Selent & Kortman, 2025; Dekel, 2022; Kansteiner, 2011; Taberner 2005; Von Bieberstein, 2016; Levenson, 2004; Fischer, 2018). There were other motivations behind philosemitic attitudes at play, such as legitimising post-war democracy by expressing safety for Jewish people and Israeli support (Von Bieberstein, 2016; Dekel, 2022, p.573; Bouteldja, 2017, p.8; Ozyurek, 2023). *Judenfreundlichkeit* [friendliness towards Jews] increased and antisemitism seemed to vanish entirely. However, it was really a transition of negative to positive prejudice formation while antisemitism remained immanent (Cohen, 2025, p.9).

The Netherlands mirrors this German trajectory, shifting from pre-war ambivalence to intense support sparked by Anne Frank Diary's global success (Levy, 2018, p.1). Both nations adopted the German term *Erinnerungsabwehrantisemitismus* [antisemitism to fend off [guilty] memory], using a 'special relationship' with Israel to redeem from historical passivity during the Nazi regime (Gans, 2017, p.153). While the Netherlands was late to support Israel in 1948, the 1956 Suez Crisis cemented a bond between leaders Drees and Ben-Gurion, positioning Israel as the West's frontier against "aggressor" Arab states (Gans,

2017; Rose, 2020). By 1967, Dutch sympathy for Israel (67%) far outstripped that of the US, UK or France (ibid.). Both countries developed a selective philosemitism rooted in post-war guilt that distanced themselves from historical antisemitism and reduced Jewish identity to a functional tool for national redemption.

Post-war guilt also shaped how Jewish identity appeared in cultural life. Wieviorka conceptualises the recurring theme of the 'Imaginary Jew' in their idealisation. Post-1960, Jews entered mainstream media through film and cultural life, yet remained once again a projection of non-Jewish ideas (Wieviorka, 2018, p.43). In Germany and Austria after the war, there evolved a 'coded presence' where Jews no longer occupied the physical spaces but were projected onto the television screens of people's homes (Taberner, 2005; Tanzer, 2024). Although Jewish identity is 'too fractured and multifarious' to be homogenised, these cultural appearances of Jews after the war re-imagined Jewishness as a symbolic construct shaped by non-Jewish memory.

e. Beyond the West

A small subset of literature demonstrates that philosemitism does not belong only to the West nor requires Jewish presence to thrive. Across the globe, Jewish identity has been adopted as a symbolic framework by communities far removed from the traditional diaspora, including in East and Southeast Asia (Schilling, 2020, p.297; Kowner, Ainslie & Podoler, 2023; Sutcliffe & Karp, 2011). In these environments, 'dualist attitudes' emerged, fusing antisemitic notions with a profound philosemitic attraction (Schilling, 2020, p.297).

In countries like Japan, China and South Korea, philosemitism was almost entirely a product of the 'Imaginary Jew' (Schilling, 2020; Wieviorka, 2018). Schilling highlights how an 'information vacuum' stems from academic theorising and pure fantasy due to a lack of lived interaction, such as the commercial success of the Talmud in Japan (Schilling, 2020, pp.298-307). Books published under the name of Conservative Rabbi Marvin Tokayer became bestsellers, yet Japanese editions were often produced without his oversight, containing cartoons of Christian symbols and giving credence to 'stubborn stereotypes' (ibid.).

There are patterns of collective identification with the Jewish community extending across history and geography. Non-Jewish communities adopted Jewish identity as a symbolic resource to position themselves against hegemonic powers or dominant ideological frameworks they wanted to resist (Sutcliffe & Karp, 2011; Schilling, 2020; Zubrzycki, 2022; Wieviorka, 2018). Afrikaner nationalists strongly identified with Jews to oppose themselves to the British, a liberal Spanish movement (*filosefardismo*) emerged to counter-position Catholic conservatism (Sutcliffe & Karp, 2011, p.21). In post-Communist Poland, non-Jewish activists 'resurrected' a Jewish past to defend liberal pluralism against nationalism (Zubrzycki, 2022).

Wieviorka's definition of philosemitism as 'a love of non-Jews for a Judaism that is somewhat imaginary' applies with particular force to these global contexts (2018, p.43). Wedgwood's concept of 'judeity', a term to describe the 'fractured and multifarious' way a Jew is a Jew, is

a useful corrective to this homogenized ‘Imaginary Jew’ that encodes stereotypes without engagement with actual diversity (Wedgewood, 2022). This international dimension is important for online research due to the borderless nature of social media platforms.

f. Placing Christianity into the Historical Context

There is a complex relationship where, in its Christian inflection, philosemitism derives from supersessionism and structural dependencies of Christianity with Judaism. At a foundational level, Christianity philosemitism is both dependent on and supersessionist toward Judaism. Karp and Sutcliffe’s distinction between ‘pure’ and ‘applied’ philosemitism (2011, p.13) applies to the dynamic between these two religious groups. Different Christian denominations across history accumulated praise for Jewish texts and traditions but have treated poorly the community in practice. Hebrew texts were foundational for shaping the Christian political imagination (Sutcliffe & Karp, 2011, p.21). Paul’s allegorical reinterpretation of Jewish law established a long-standing pattern in which admiration for Judaism remained inseparable from the expectation of Jewish conversion (ibid. pp.16-21).

In the sixteenth and seventeenth centuries, this was furthered by Christian Hebraists’ scholarly engagement with Jewish texts which were valued for linguistic utility rather for Jewish perspectives themselves (Levy, 2017, p.3). Works such as Knorr von Rosenroth’s *Kabbala Denudata* (1677-8 & 1684) translated Kabbalistic texts from Hebrew into Latin and German. It spread the idea across Europe that the Zohar (a chief Kabbalistic text) was unified through Christian aegis, conveying how intellectual admiration coexisted with epistemic appropriation (Kilcher, 2017).

The entanglement between admiration and identification with Judaism blurred by the late 18th century. Christian philosemitism led to political radicalism as questions of citizenship came to the forefront. The French Revolution generated secular and religious symbolism around the ‘regeneration of Jews’ and campaigns for Jewish political rights drew on both traditions simultaneously. Philosemitic discourse shifted from concerns of salvation to those of political inclusion, with Jewish emancipation symbolising broader ideals of secular progress (Sutcliffe & Karp, 2011, p.18). This shift did not eliminate instrumentalisation. Jewish identity became instead a vehicle through which the modern state articulated its own values.

At the same time, the limits of Christian philosemitism remained evident. Figures such as British politician Lord George Gordon convey how identification with Judaism was tolerated only at distance. His conversion in 1787 with his ‘marginal eccentrics’ and subsequent marginalisation depicts that while admiration for Jewish traditions was fashionable, full participation in Jewish life was socially and politically constrained (Levenson, 2004, p.144). This tension between symbolic appreciation and lived exclusion persisted into the nineteenth and twentieth centuries, including within Protestant missionary movements, which aimed to preserve Jewish historical significance and bring its theological suppression simultaneously through conversion (Levy, 2018, p.2)

Contemporary expressions of philosemitism continue to reflect this historical pattern. Modern literature characterises contemporary Christian Zionism and Evangelical support for Israel as explicit forms of philosemitism, driven by eschatological commitments and a literal reading of biblical prophecy (Inbari et al., 2020). These movements often frame Jewish existence as a necessary precondition for the Second Coming, positioning Jewish people as, what McDermott terms, a ‘secular parable’ within a Christian teleological narrative (2016, p.62)

While some modern movements, such as Judeocentric Catholicism, attempts to shift towards a philosemitism based on post-Holocaust atonement and the thriving of Jewish people, they remain inherently tied to a theological paradox regarding the necessity of Jesus for salvation (Polyakov, 2020). These tensions are visible in movements like Messianic Judaism, which many mainstream Jewish denominations reject as ‘spiritual terrorism’ because it violates established religious boundaries (Shapiro, 2011, p.7; Smith, 2007). Across these contexts, Christian philosemitism oscillates between genuine admiration and structural instrumentalisation.

g. ‘New Philosemitism’

Hannah Rose’s ‘new wave of philosemitism’ is one of few papers acknowledging the phenomenon’s existence online (2020, p.6). ‘New Philosemitism’ refers to the European far-rights use of pro-Israel positions as a tool to deflect accusations of racism and distance themselves from Nazi ties, similar to ‘state philosemitism’ (Rose, 2020, p.ii; Dekel, 2022; Gans, 2014). Multiple scholars turn to Geert Wilders to demonstrate this new era as he explicitly uses framing such as Israel is the ‘West’s first line of defence’ against Islam (Rose, 2020, p.8; Wertheim, 2017b, p.280). Other far-right parties across Europe portray similar patterns, such as the AfD whose pro-Jewish statements often target Muslims or the left opposition (Selent & Kortman 2025).

It must be acknowledged directly that the scholarly literature on online philosemitism is sparse. Rose (2020) notes ‘limited evidence for philosemitic life online’, and there is no existing corpus or dataset designed to capture it computationally. This is the gap that this thesis sets out to investigate and is its main contribution.

Online Antisemitism

a. Antisemitism as Philosemitic Nexus

The literature does not pinpoint the exact relationship between philo- and antisemitism. However, it is clear there is a connection between the two phenomena that reveals new dynamics about how non-Jews construct their views of Jewry that makes discussing the two concepts together important (Levy, 2018, p.2). There is an absence of philosemitic literature compared to the numerous published volumes of ‘old’ and ‘new’ antisemitism papers (Cohen, 2025, p.21).

As a ‘flexible concept’, philosemitism undertakes the ‘othering’ process familiar to antisemitism (Gans, 2017, p.156; Cohen, 2025). This means idealisation can ‘lead to an

overly negative interpretation' (ibid.). Both involve essentialising Jews meaning Jewish people are defined through assumptions imposed by other communities continuously (Rose, 2020, p.5). In some cases, philosemites can replicate the 'othering' process by splitting Jews into "good" and "bad" categories' to display they do not group the Jewish community together (Cohen, 2025, p.17). Though, this just repeats the same distancing process that leads to distrust of a constructed outgroup and confirms that the phenomena are 'two sides of the same coin' (Rose, 2020, p.5). It demonstrates how understanding the key patterns of antisemitism is useful for this study as it makes identifying positive appearing tropes easier.

b. Classical Antisemitism

An awareness of classical antisemitic tropes and conspiracy theories is useful for flagging their appearance under the guise of philosemitism. There are four main categories of antisemitism described in the literature: Christian anti-Jewish hatred, racial antisemitism from the 1800s, post-war antisemitism and Israel-related antisemitism (Becker & Allington, 2021, p.1; Guhl, 2022, p.61). From these categories, two main concepts dominate. The first links Jews to power - one of the most popular discourses found online (Becker et al., 2024, p.6; Guhl, 2022; p.60). The second presents the Jew as the Other. This 'Imaginary Jew' is well established in both antisemitic and philosemitic academia, displaying its relevance in the construction of Jewish identity.

These correlate with key conspiracy theories monitored by researchers. The most common of all is the Global Dominant Conspiracy Theory that claims Jewish people advance their own interests over others (Chapelan, 2024, p.178). Other conspiracy theories include the belief Jews control the media, Warmongering theories, 'White Genocide', Holocaust-related and QAnon theories (ibid.; Levin et al., 2022).

With the rise of social media, classical antisemitism has been repacked by the far-right and far-left with the impact of dehumanising the Jewish figure (Weimann & Masri, 2022, p.169). The same classical tropes are echoed online of 'hidden Jewish power' and dual loyalty (Weinberg et al., 2025; Jikeli et al., 2024). Its shift from offline to online contexts has been closely monitored, yet this same manifestation with philosemitism has been overlooked.

c. Coded and Implicit Language

Antisemitic expression is even more diverse online than offline making it challenging to study (Becker et al., 2024, p.11; Becker & Troschke, 2022). This has implications for philosemitism, as it follows many of the same processes in its formation. Most antisemitic online antisemitic content is increasingly coded, sharing its own cultural system that is easy to obscure and near-impossible to detect as it requires prior historical and cultural knowledge (Weinberg et al., 2025; Becker & Allington, 2021). Becker's *Lexicon* attempts to systematise this, gathering 175 categories of explicit and implicit antisemitic hate speech (Becker et al., 2024, p.10). However, philosemitism does not appear in this dictionary of definitions and examples, suggesting that philosemitism lacks the same systematic categorisation as antisemitism despite potentially operating through similarly implicit structures online.

Researchers are finding it increasingly difficult to detect implicit antisemitism as it appears in less overt formations and is focused on ‘cultural, political and conspiratorial discourse’ which is easier to share freely online (BSA, 2025a). Fringe groups have utilised platform dynamics to normalise this coded antisemitism and spread bottom-up narratives creating an algorithmic radicalisation process using methods such as ‘dog whistles’ (Becker et al., 2024, p.8; Mendelsohn et al., 2023). With the impact of Artificial Intelligence (AI) and the broader attention economy, these processes are further amplified, enabling the rapid circulation of coded content at scale. Understanding how non-Jews cognitively process such content is a necessary step before examining what automated systems can and cannot do.

Cognitive Perspectives

a. Perceiving the Jew

There is no substantial literature on how non-Jews learn about Jews. In research, focus leans towards understanding whether non-Jews hold antisemitic views because of the immediate threat that negative views pose to the Jewish community (Hübscher & Von Mering, 2022, p.12; Weimann & Marsi, 2022). Some studies demonstrate how online commentary about Judaism impacts Jewish identity and wellbeing which are insightful for understanding how online and offline realities intersect in identity formation (Becker & Allington, 2021). With the rising spread of antisemitic disinformation online, the understanding of Jewishness by non-Jews is inevitably evolving (Antisemitism Policy Trust, 2024, p.6).

However, what is missing is a full comprehension about the benign discussions of the Jewish community, whether philosemitic or not. The processing of ‘othering’ is a problem in Jewish history. Philosemitism could be as damaging in the long-term as overt antisemitism because it appears positively to the naked eye (Norman Lear Center, 2024). Therefore, it does not demand immediate pushback on essentialised ideas described in the literature above. This is especially the case with the increasing use of coded language used by fringe groups to derogate out-groups as it makes identifying instrumentalising narratives more challenging (Farrand, 2023).

Social media architecture makes the subjectivity, cultural and social dependency of hate speech formation more complicated (Badamchi, 2021; Boromisza-Habashi, 2012; Cowan & Hodge, 1996). Implicit hate speech has been shown to radicalise and normalise ‘in group/out group’ attitudes until they are no longer viewed as offensive (Bilewicz & Soral, 2020). Philosemitism is not deemed hate speech due to its positive linguistic framing. However, it undertakes the same processes and mean-making formations as antisemitism and therefore fulfils the criteria above.

A consideration of cognitive insights provides a useful framework for understanding where a threshold for philosemitism can be drawn, given the decontextualised environment of social media platforms. The two following cognitive frameworks have been selected to illuminate this study:

b. Illusory Truth Effect

The illusory truth effect is a psychological phenomenon in which repeated statements are judged as more likely to be true than new ones, regardless of their actual veracity (Jalbert et al., 2020, p.603). It reduces the ability to understand information accurately because it prompts individuals to rely on processing fluency, the ease of processing information, as a heuristic for truth (Wang et al., 2016, p.2). The truth effect can override a person's existing knowledge, causing them to accept false information without performing the effortful analysis required to identify it as false (Jahn et al., 2023, p.3). Repeated exposure to misinformation in contexts like social media can therefore allow falsehoods to be seamlessly integrated into an individual's knowledge base (ibid.).

Increasingly more young people believe the Holocaust did not occur due to consuming conspiracy theories and Holocaust denial posts on social media (Becker et al., 2024, p.7). There are pre-exposure warnings, or debunking, that can reduce the illusory truth effect by eliciting more critical processing at the time of consumption. When warned that information may be false, individuals disrupt the tacit assumption that communication is truthful which leads to increased scepticism (Jalbert et al., 2020, p.607). Without these warnings, the size of the truth effect is doubled (ibid.). However, the timings of these warning signals are important as if only given at the time of testing, or in this case reading an online post, then the warning has little impact (ibid.). For philosemitism, the positive framing of content acts as its own counter-warning because the information does not read as threatening or false, neither the reader nor the platform is prompted to issue a corrective signal, meaning the illusory truth effect can in theory operate entirely unchecked.

c. Signal Detection Theory

Green and Swets' (1996) behavioural framework of Signal Detection Theory (SDT) offers another complementary framework for understanding why philosemitic content evades scrutiny at the level of individual judgement. It categorises the possible outcomes of a binary decision, such as determining whether a piece of content is harmful, into four results: hits (correctly identifying a harmful signal), correct rejections (accurately identifying genial content), misses (failing to detect harm that is present), and false alarms (flagging content as harmful when it is not) (Heeger, 1997, p.1; Batailler et al., 2023). The theory distinguishes between the observer's sensitivity to the signal and criterion (threshold) (ibid.).

For the study of philosemitism, SDT is particularly instructive because it demonstrates that detection failure does not mean ignorance from the individual. Where antisemitic content triggers a relatively low criterion, prompting users to flag or reject it, philosemitic content raises that threshold considerably. Its positive framing reduces the perceived cost of a miss. So, if the content appears warm toward Jews, the social and cognitive cost of failing to scrutinise it seems negligible. Therefore interviewees will likely differ drastically on their interpretation of philosemitism depending on their positionality.

Another useful element of SDT is its consideration of 'background noise' that makes the phenomenon hard to detect (Heeger, 1997). The overlap between genuine support and

philosemitism means even a sensitive observer may struggle to separate signal from background noise. This is because philosemitism's harm lies in its framing and cumulative effect. SDT therefore frames philosemitism as likely to be missed by individual readers but also that any surface-level cue will systematically underdetect it.

Together, these frameworks reveal how repeated exposure to essentialised content normalises it regardless of intent, while the ambiguity of philosemitic framing makes it difficult for individuals (and systems) to register it. These cognitive limitations experienced by human readers are mirrored in automated detection systems. The following section examines the current state of automated antisemitism detection and the specific obstacles that arise when the content in question does not conform to overtly negative linguistic patterns.

Automated Detection

a. Gold Standard Datasets

The application of AI to social media content carries well-documented limitations and advantages. While automated tools offer analytical scale that manual methods cannot match, they are vulnerable to the biases embedded in the data used to train them (Antisemitism Policy Trust, 2024, p.4). This makes the quality of training data a foundational concern. Mutable, contested concepts like antisemitism are difficult to quantify (Becker et al., 2024, p.58) and so the construction of gold standard datasets has become a central preoccupation in computational social science (Jikeli et al., 2022, p.193). A gold standard dataset curates labelled data, validated by annotation experts, providing a reliable baseline for classification when evaluating the accuracy and performance of models or quantitative methods (ibid.).

The process of building such datasets is extensive. It involves encompassing the definition of a classification schema, establishing annotation guidelines, data collection and pre-processing, expert training and the construction of a final corpus. Each piece of labelled data is annotated with meaningful tags, such as categories or descriptions, for a model to better learn the inputs' relationship and provide desired output. The field has expanded with increasingly more publicly available labelled datasets, each capturing a different area of online discourse (Gomez et al., 2020; Davidson et al., 2017; Waseem, 2016).

There are still methodological tensions the literature acknowledges. Chief among these is the subjectivity of annotation: human annotators frequently label identical content differently, reflecting the culturally and socially contingent nature of hate speech identification (Jikeli et al., 2022, p.194). Debates around definitional frameworks expose this clearly. The IHRA definition of antisemitism is the most widely adopted across annotated datasets but yet it is too general to be operationalisable as the 'linguistic dimension' of implicit language is ignored (Becker & Allington, 2021, p.2; Jikeli et al., 2022, p.195).

Antisemitic intent is not a requirement to classify a message as antisemitic, which has significant implications for how annotation guidelines are designed and applied (Jikeli et al., 2022, p.199). Without labelled data, it was not possible to distinguish content according to

antisemitic speech, limiting the analytical value of keyword and lexical network approaches (Allington, 2022, p.34). Labelled data enables analysis of label frequencies, including the distribution of ‘stereotype categories’ and the construction of ‘co-occurrence diagrams’ that reveal relationships between labels used across different online communities (Becker et al., 2024, pp.58–59). Despite the growth of antisemitism datasets, no equivalent gold standard exists for philosemitism. This study's annotation guideline directly addresses that gap.

b. Key Identified Challenges of Current Models

A comprehensive evaluation of models designed to detect implicit or coded hate speech surpasses the scope of this thesis. However, identifying the key challenges facing current approaches clarifies the gaps that an annotation guideline for philosemitism must account for.

Existing models perform better when trained on corpora that capture greater linguistic and contextual nuance, yet such training data remains scarce (Bjola & Manor, 2020; Becker et al., p.53). Identified challenges include content-structural diversity, the ‘elusive nature’ of antisemitism as a phenomenon as well as practical constraints such as comment length variation, world knowledge gaps and the context-dependency of harmful content (Becker et al., 2024, p.51). Antisemitism is not always contained within a single text. It frequently emerges from a system of social practices and relationships, meaning that ideological intent is constructed across interactions rather than within them (Becker et al., 2024, p.48). BERT-based and conversational models have shown limited effectiveness in identifying antisemitic text, particularly where comment-to-post relationships must be inferred (Chiu et al., 2022, p.56; Prabhu & Seethalakshmi, 2025).² Reassessing the boundary between contextualised and non-contextualised antisemitism in annotation may therefore better leverage the capabilities of large language models than previous approaches have achieved (ibid.).

These challenges apply with equal, if not greater, force to philosemitism, where the absence of overtly negative language makes surface-level detection unreliable and the need for theoretically grounded annotation all the more pressing.

² BERT-based models are a type of natural language processing model (NLP) built to understand text by reading it bidirectionally.

Methodology

To understand how online philosemitism can be conceptualised and operationalised for automated detection, a mixed-method approach must be undertaken. A quantitative approach is not suitable as no pre-existing gold standard dataset exists to test elements of causation or statistical significance around the phenomenon. Multiple qualitative methods were therefore adopted to effectively explore the research question and accompanying hypotheses. The aim was to produce an annotation guideline, a principled instrument for future labelled datasets, grounded in both expert knowledge and real online discourse.

Data Collection: Interviews

Twelve interviews were conducted ranging from fifty minutes to two hours. Participants are named in Figure One's first column. To clarify any abbreviations: Campaign Against Antisemitism (CAA) is a volunteer-led charity, JSOC is a common term in the UK for a university Jewish Society and the forum moderator moderates a Facebook forum with over 20,000 members. Participants ranged in nationality, gender, age and religion that addressed the hypotheses fully and bridge together a more concrete conceptualisation of online philosemitism. As the undertaken method of Constructivist Grounded Theory (CGT) includes a consideration of positionality, the diversity of my participants was core to the analysis as the points of convergence strengthened the category creation.

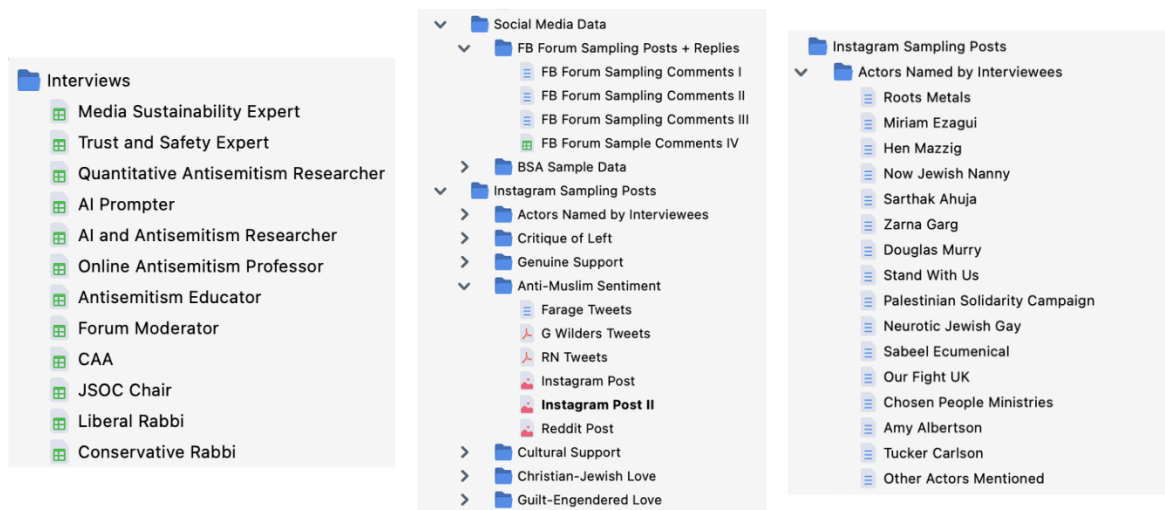


Figure One: Screenshots of All Data Collection Documents from MAXQDA.

Although beginning with semi-structured interviews, I turned to unstructured interviews as the interview sample did not share a common framework for understanding philosemitism. Interviewees approached the concept through their own distinct and even conflicting epistemologies. For example, religious participants grounded their interpretations in theological or community-based experiences of Jewish identity while AI practitioners and

moderators emphasised operational definitions, scalability and detection thresholds. This heterogeneity necessitated a more flexible approach that introduced relevant contexts and enabled the research to capture the full range of meanings attributed to philosemitism rather than constraining responses to a fixed conceptual schema.

Vignettes were still included in the interviews (available in the appendices) to see how participants interpreted various online statements about the Jewish community. These varied in their clarity of philosemitic content and were selected to reflect what the literature identified as key forms of the phenomenon. All data will be kept until the completion of the master's degree and then deleted in accordance with the thesis policy. Consent forms were signed by participants and identities have been anonymised to enable the use of direct quotations throughout.

Data Collection: Social Media

Social media data collection originated from interviewees and online actors they referenced. Collection began in January 2026 and ended in April. Once my categories became clearer through the coding process, threads could be followed and suitable examples placed under relevant folders. For example, the 'Anti-Muslim Sentiment' file visible in Figure One. The data is cross-platform, covering Twitter/X, Instagram, Bluesky, Facebook, YouTube, Reddit and LinkedIn. All content is in English to avoid misinterpretation through automated translation.

To sample data from the Facebook Forum titled 'Where Non-Jews ask Jews Questions about Judaism', responses to four 'Topsy Turvy Questions' were extracted with the permission of the Forum Moderator. The Topsy Turvy Question series allows Jewish forum members to ask non-Jews questions about their perception of Judaism, reversing the usual protocol. This series emerged directly from the interview with the Forum Moderator when a Topsy Turvy question after the interview generated such high engagement that it began a whole new series within the forum. The Forum includes a survey where members declared their purpose for joining, with 43% non-Jewish responses (06/11/2025). The results were: ally/supportive (18%), dialogue/learning (14%), curious (3%), considering conversion (4%), actively converting (2%), married to a Jewish person (2%).

This closed forum has its own set of rules that members must agree to when joining. If disrespected, members can be muted or banned. Therefore, it does not reflect public social media platforms. However, members share openly their experiences and perspectives on Judaism and must state their religious identities, making it a useful space for observational and conceptual work on religion and the Internet.

A small subset of data was later extracted from the Blue Square Alliance (BSA) dataset using Boolean queries.³ The BSA is an American-based non-profit research hub advocating for

³ Boolean Queries were created using the same methodology suggested by Anti-Defamation League, cited by Finkelstein et al., 2018, p.2. It included using expert reviews to generate a list of terminology and context-terms that informed the query.

awareness of antisemitism with the campaign “Stand up to Jewish Hate”. Only one category of queries was selected for the thesis due to the volume of data requiring too much time for processing within the project’s timeframe. This data fell under the essentialising mechanism of guilt-engendered philosemitism, which the literature described and interviewees confirmed. BSA’s data scrapping tool limited its search to twelve months from January 2025 to January 2026. It excluded key terms such as ‘Hamas’ and ‘Israel’ to avoid capturing content outside of philosemitism specifically. The following three Boolean queries were constructed from real posts:

(Holocaust OR shoah OR "never again" OR "our history" OR heartbreak)
AND ("stand with israel" OR "support jew*" OR "protect jew*")

"never stop" AND (sharing OR "speaking up" OR "spreading the light" OR remembering)

(truth OR resilience OR power OR courage OR love OR reverence) AND "of the jew*")

This brought in just under 70,000 posts. A random sample of 399 posts were extracted and cleaned using Python on Google Colab which produced 205 usable posts for analysis. The cleaning process began with 24 columns and removed irrelevant ones, retaining: page type, text, author, whether it was a post or reply, impressions, likes, shares, country, sentiment and date. The sentiment column was inaccurate but retained for transparency. The full cleaning code is available in the appendices.

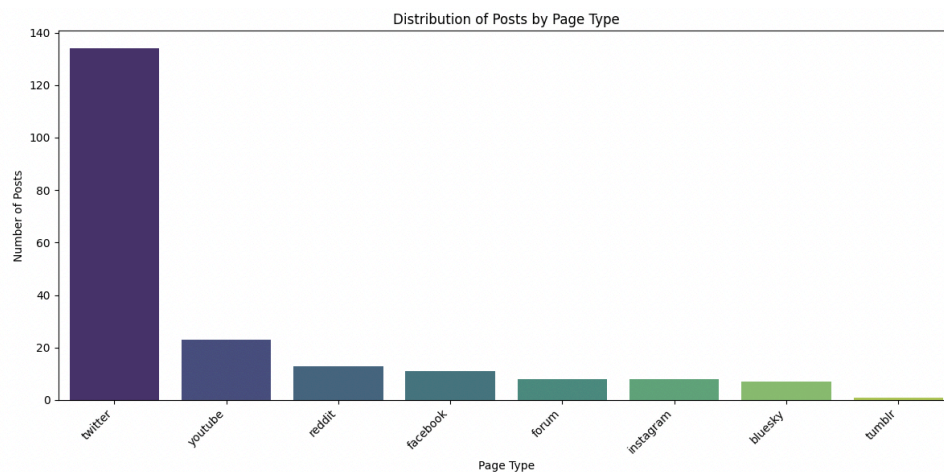


Figure Two: Distribution of BSA Data per Platform.

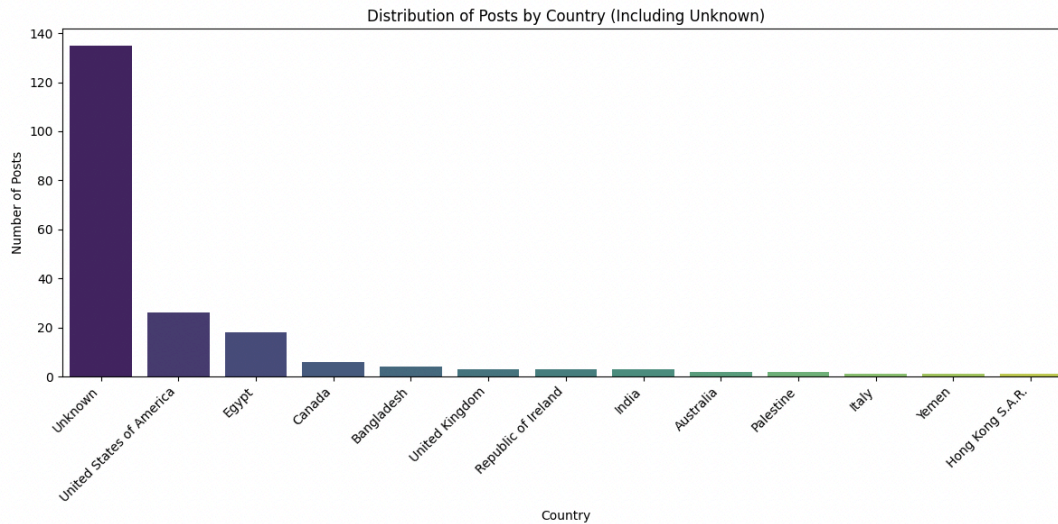


Figure Three: Distribution of BSA Data per Country.

Constructivist Grounded Theory (CGT)

Initially, Qualitative Content Analysis (QCA) was adopted with the intention of systematically developing a category system to classify instances of philosemitic discourse (Schreier, 2014). It seemed appropriate at the outset due to its suitability in producing operationalisable categories for computational frameworks. However, QCA requires a level of conceptual stability early in the coding process and for this study, this assumption proved indefensible. Both the literature and the early expert interviews displayed a lack of consensus on philosemitic boundaries. Imposing a predefined coding frame risked prematurely fixing a phenomenon that is inherently fluid. QCA is also mainly oriented towards systemically classifying content making it less suited to contexts where the primary research task is conceptual development.

The methodological approach subsequently adopted was Kathy Charmaz's Constructivist Grounded Theory (2014). Adapted from Glaser and Strauss' classical Grounded Theory (cGT) (1965), CGT assumes the constant data comparison that is central to the method and allows for an interpretive understanding of information (Sebastian, 2019, p.3; Charmaz, 2014). Figure Four displays Sebastian's practical guide for selecting the most applicable form of GT. Through a systematic iterative approach, grounded theory research develops theories that explains processes and interactions and builds reasoning for how and why social phenomena unfold temporally.

	GT (Classical Grounded Theory)	IGT (Interpretive Grounded Theory)	CGT (Constructivist Grounded Theory)
Philosophical Influence	(Attempts to be) free from influence	Interpretivism	Constructivism and Pragmatism
Role of the Researcher	The researcher is <i>distant and detached</i> .	The researcher is <i>engaged</i> with and <i>actively interprets</i> the data.	The researcher <i>constructs</i> rather than <i>discovers</i> .
Allowance of Prior Knowledge	No, the researcher and research must remain neutral. Only the information provided by the collected data should influence the progress of the research. Prior knowledge could negatively influence the direction or quality of the concluding theory.	Yes, it can be used to strengthen the overall research and data collection. Referred to as ' <i>sensitivity</i> ' and includes having insight on relevant issues.	Yes, it is understood that one cannot escape prior knowledge. Examine and understand how it influences the researcher and research; do not ignore, erase or let it control the direction of the research.
Literature Review	To be conducted following data analysis.	Allowed prior to and during the data collection process. Can be used for data comparisons, enhance sensitivity, stimulate observations, and confirm or explain results.	There is no prescribed location; it is up to the decision-making process of a given researcher. If written early, it should be revisited to critique and confirm it aligns with the research's conclusion.
Research Question(s)	There should be no pre-set or vaguely established questions prior to data collection. Questions should become clear during data analysis.	Kept partially vague for flexibility; will become clearer as data emerges. Each additional question can reference another topic of interest.	Influence how data is collected. Can and should be altered if more significant or pressing questions arise.
Data Coding and Analysis	(1) Substantive coding, and (2) Theoretic coding (TC) Focuses on patterns or trends within the data. Includes the Constant Comparative Method and use of a core category. Includes TC to merge categories together into a substantive theory.	(1) Open-coding, (2) axial-coding, and (3) selective-coding Allows for single occurrences within the data to be coded and analyzed for significance. Includes the Constant Comparative Method and use of a core category. Have broken down and rebuilt codes to create more significant and descriptive categories that lend themselves to a substantive theory.	(1) code everything, and (2) group all data around the most predominant codes (includes focused coding) Purports these as flexible guidelines rather than strict rules. Allowance of more than one core category.
Theory Creation and Verification	There is a distinct separation between theory generation (primary) and verification (secondary). Creation of a substantive or formal theory is central to the completion of a research study. Verification can only occur afterwards by quantitative analysis.	Creation of a substantive or formal theory is central to the completion of a research study. Verification occurs through multiple perspectives confirming the same data.	The constructed theory is an interpretation rather than an exact representation. The theory is dependent upon the researcher's view and cannot occur or stand without it.

Figure Four: Sebastian's Illumination of Types of GT (2019).

There are many forms of GT including classical, constructivist, interpretive and situational (Corbin & Strauss, 1990; Clarke, 2005; Brink et al., 2006; Charmaz, 2014). Classical GT requires a purely inductive approach which would hold back this thesis as it was partly inspired by Genevieve Zubrzycki's paper *Resurrecting the Jew* (2022). Classical GT claims that an abductive approach inserts potentially significant bias negatively impacting the study's quality whilst Interpretive Grounded Theory (IGT) and CGT assert that prior knowledge cannot be ignored and can even be used as a tool to strengthen, rather than define, research (Sebastian, 2019, p.5). IGT assumes that there is a coherent underlying structure in an emerging phenomenon that can be revealed and later verified. However, there were unstable thresholds and contested definitions throughout this study's research process. The task was not to uncover a stable structure and was therefore not relevant for this project.

For an inchoate phenomenon like philosemitism, CGT takes a pragmatist approach combined with a relativist epistemology which is conducive to this study (Sebastian, 2019, p.3). Abductive reasoning is required because of the diversity of philosemitism and its disjointed timeline would make using purely an inductive approach far less effective. CGT acknowledges different constructive perspectives from the 'historical, social and situation conditions of its production' (Charmaz, 2014; Charmaz, 2017, p.299 cited in Sebastian, 2019,

p.4). With international participants approaching the same phenomenon from distinct cultural positions, the positive language of Jewish people will be interpreted differently depending on the context of Jewish communities in each country. CTG accommodates this dynamic.

The coding process followed a two-tier structure developed from 502 memos, beginning with a line-by-line approach (Charmaz, 2014). Memo writing was important for prompting early data analysis data and codes, capturing the emerging theoretical directions of the research. Early memos were written informally to generate ideas, as Charmaz suggests, 'I wrote the memo to catch my fleeting ideas about the code and to probe data' (2014, p.80). Advanced memos described how categories emerged and made comparisons across the data.

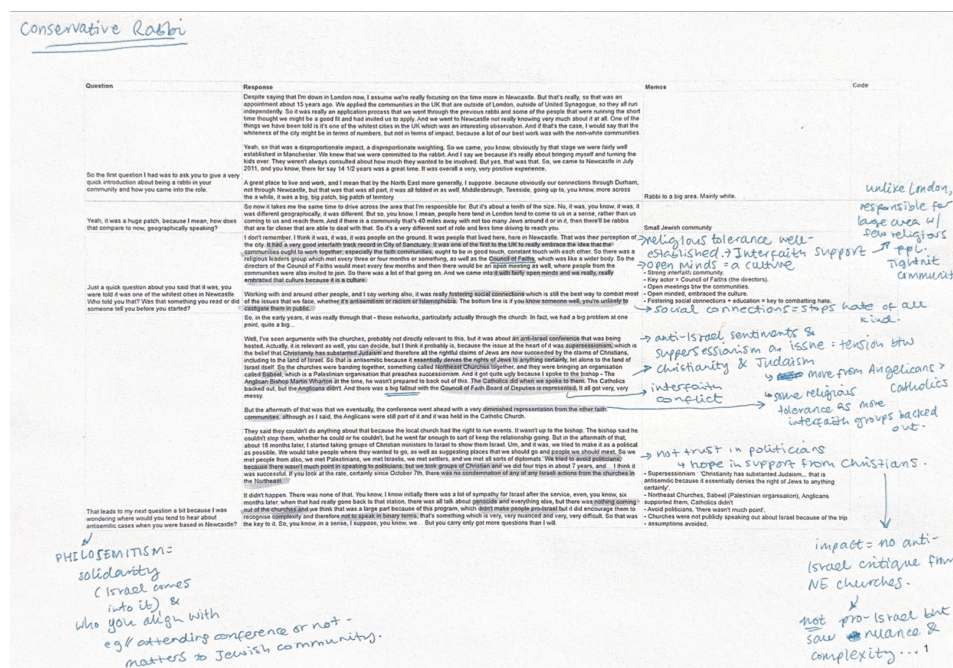


Figure Five: A Scan of Manual Writing of Initial Memos before using MAXQDA software.

Two main phases structured the coding. All pieces of data were named from the collection. Secondly, initial codes were produced and the most used codes were identified to organise the remaining data around them to create the main categories (Charmaz, 2014). Focused coding then tested constructed theoretical assumptions against remaining data. The clustering of codes evolved over the project due to the ranging professional backgrounds of interviewees that brought in different concepts. A core category was not required as CGT allows for more than one central theme and is sceptical of collapsing rich descriptive work into a single overarching category (Berthelsen et al., 2016). The category framework was constructed rather than uncovered, and was therefore open to revision as new forms of philosemitism emerged.

The consideration of positionality is integral to CGT and was addressed throughout the memo writing process (Charmaz, 2014). Prior knowledge was controlled rather than erased and used tentatively to open up inquiry (Charmaz, 2014, p.31). Participants' views and actions were interpreted from their own perspectives at each stage of coding, though positionality as a

researcher will be addressed in full in the limitations section (Charmaz, 2014, p.115). After manual coding appeared too slow, all coding was also completed using MAXQDA software, including for memos, advanced memos, initial codes and code comparisons, leading to the final categories presented in the findings.

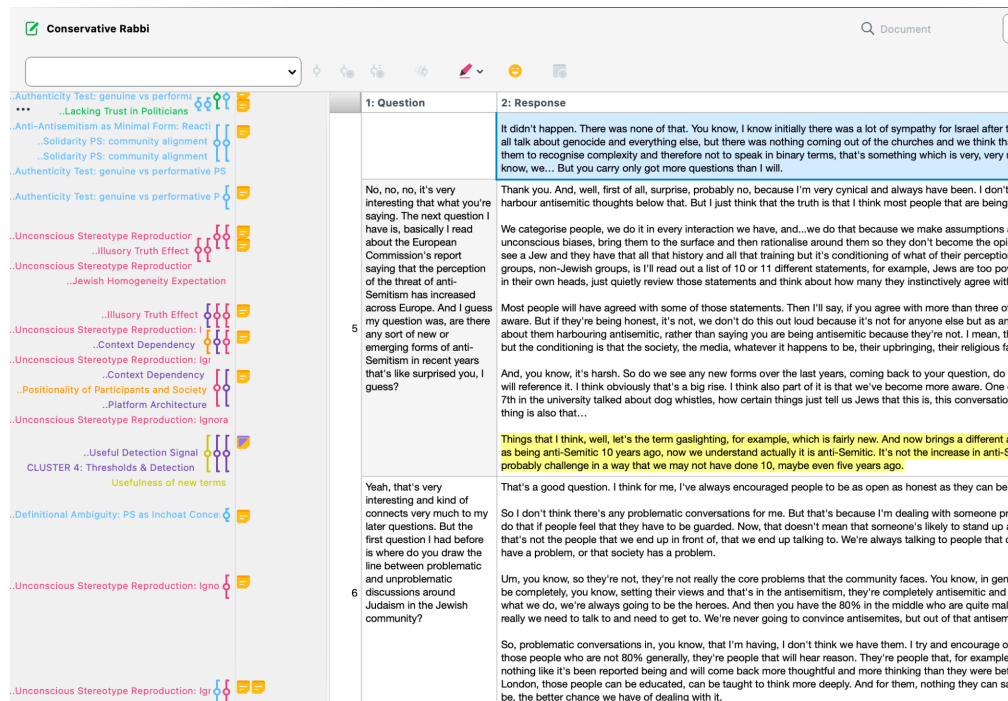


Figure Six: Screenshot of MAXQDA Coding from Conservative Rabbi Interview.

Theoretical sampling was undertaken by following social media accounts and Facebook Forum threads referenced by interviewees once categories became clearer. In theory, theoretical saturation is reached when gathering new data no longer sparks new insights and repeats existing codes (Brink et al., 2006, p.190). However, it is acknowledged that further avenues of inquiry may remain that could not be pursued due to the constraints of the research timeframe.

Annotation Guideline

The final methodological output is an annotation guideline, produced from the conceptual and technical findings across both findings chapters and available in the appendix. The guideline was developed in consultation with Jordan Blatter, co-author of Becker et al.'s (2026) paper on decoding antisemitism online. It adopts Becker and Blatter's annotation technique of distinguishing between explicit, implicit, co-text and grey area cases and applies this framework to philosemitism for the first time. It includes inclusion and exclusion criteria and a three-way classification of posts as philosemitic, antisemitic or neutral. The rationale for specific guideline decisions is explained at the close of the findings chapters, where the conceptual and technical evidence for each choice is established. The guideline is a first step towards a labelled dataset that could be used for LLM benchmarking experiments and future automated monitoring of online philosemitism.

Introducing the Findings

Online philosemitism is hiding in plain sight, reproducing the same essentialising logic as antisemitism. It serves demonstrably non-Jewish agendas and evades both the threshold judgement of humans and consequently any automated detection systems. What is clear in the findings is the convolution between philosemitism and antisemitism that runs throughout all the findings. Just as the literature suggests, one works as a nexus for the other.

CGT relied on the diversity of participants' backgrounds. Much of the divergence in opinion about the detection and problematising of philosemitism depended on participant's religious and professional backgrounds which brings notable strengths and challenges to producing coherent findings.

Four core categories emerged from CGT's iterative, bottom-up approach of CGT (Charmaz, 2014). These derived from the writing of over five-hundred memos, advanced memos, initial codes and then finally the focused codes that created the categories, shown in Figure Seven. Chapter Two of findings expands on the first two categories which are consistent with the content and uses of online philosemitism found from the data. Chapter Three develops an understanding of the platform context and challenge of interpreting such a complex topic which depends partly on the participants' subjectivity.

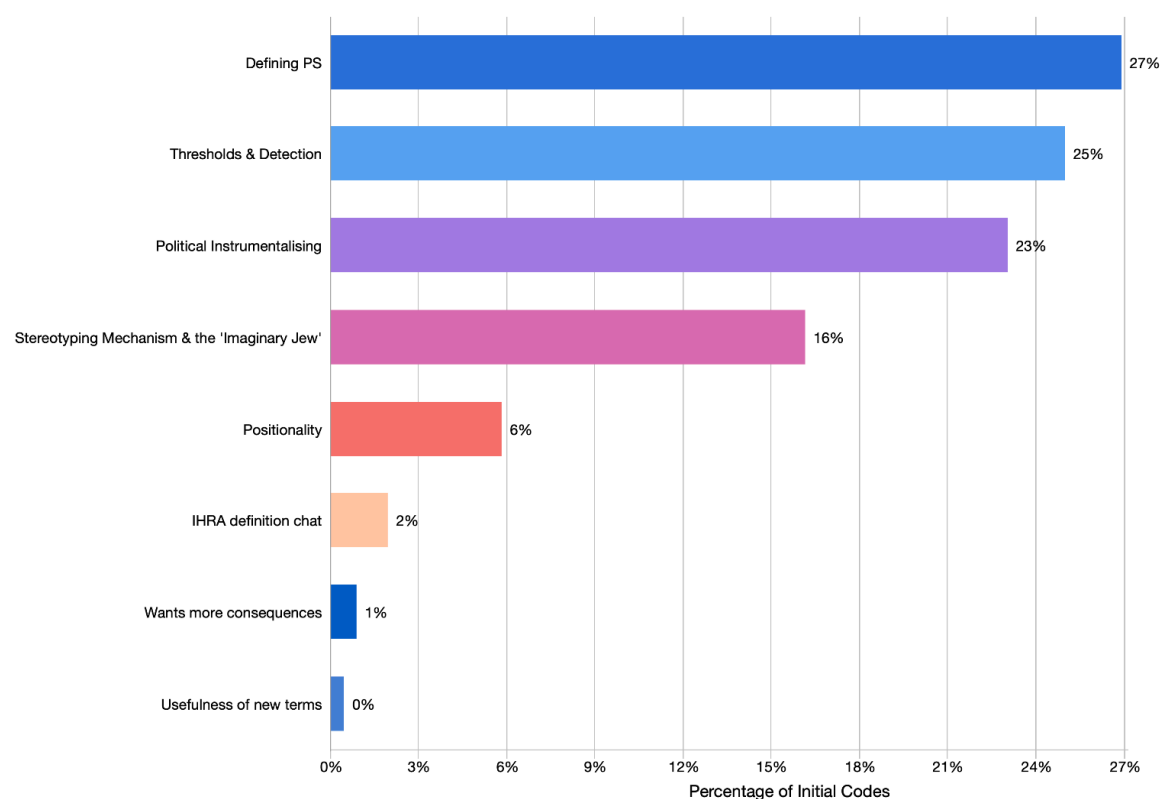


Figure Seven: Visualisation of the Code System.

There are initial codes that belong to multiple categories but broadly, online philosemitism consists of the following main findings:

1. The Essentialising Mechanism

This first finding presents how online philosemitism operates primarily through essentialising Jewish stereotypes that attribute collective positive characteristics to Jewish people in a manner structurally analogous to antisemitism, regardless of the speaker's intent. These attributes 'other' the Jewish people and make them unique to the majority culture re-creating a positive formation of the 'Imaginary Jew'.

2. Instrumental Formations

The second finding demonstrates how actors use these simplified stereotypes in an instrumentalised manner, reaffirming established literature. It is found to be deployed by far-right actors, Hindu nationalists, Christian communities and guilt-driven non-Jews to serve non-Jewish political, religious and identity-related agendas.

3. The Authenticity Threshold

Philosemitism's varying instrumental formations ties in closely with the challenge of identifying non-Jewish authenticity in their support for the Jewish community. This finding emphasises the convergence of detecting genuine support for the Jewish community. Interviewees clarify four criteria to determine sincerity but fail to apply this logic to online contexts where the medium structurally rewards performative over genuine expression and removes contextual information about non-Jewish actors.

4. A Detection Gap

The final finding reveals how positive linguistic formations of philosemitic discourse lead to a structural detection failure. It brings together prior category conceptualisation and highlights the operationalisation process. Interviewees stressed the challenges of automated moderation systems that struggle with implicit language. However, they also conferred how to overcome these obstacles. This was used to build an annotation guideline as a first step to create a labelled dataset on the phenomenon. This could monitor online philosemitism at a larger scale to detect its significance and how it evolves as it could become a new player in signalling hateful views of the Jewish community by non-Jews.

In this analysis, findings are unpacked through these four core categories. They form a natural chronology for conceptualising online philosemitism and pointing towards its operationalisability in an annotation guidebook.

Category	Parent Code	Initial Codes	Frequency
Essentialising Jews	Stereotyping Mechanism & the 'Imaginary Jew'	Jewish homogeneity expectation, 'Imaginary Jew' and absence of contact, positive stereotyping through essentialised praise, unconscious stereotype reproduction due to ignorance	16%
Instrumentalising Jews	Political Instrumentalising	Religious instrumentalising, guilt-engendered PS, lacking trust in politicians, cultural appropriation, model minority identification (sub-code: to signal hardship), new right politics: anti-Islam/racism, 'noticing' with positive language	23%
The Authenticity Threshold	Defining Philosemitism (PS)	Solidarity PS: community alignment, authenticity test: genuine vs performative PS, antisemitism (AS) as a minimal form, PS as AS structural sibling, definitional ambiguity/inchoate concept (sub-code: anti-zionism)	27%
A Route out of the Failure of Detection	Thresholds & Detection	AI & LLM limitations, useful detection signal, illusory truth effect, context dependency, platform architecture, coded language and meaning-form divorce, threshold problem (sub-code: free vs hate speech)	25%
Positionality	Positionality	Researcher Positionality, positionality of participants, annotation protocol, dataset creation	6%

Figure Eight: Visualisation of Application of CGT Methodology.⁴

⁴ Other less significant creative codes: IHRA definition (9 segments), demand for consequences against antisemitism (4 segments), usefulness of new terms (2 segments).

Chapter Two

The Mechanism of Online Philosemitism

Category One: The Essentialising Mechanism

Four Sub-Types of Essentialised Praise

Online philosemitism is, at its core, a stereotyping mechanism. Three quarters of participants confirmed how online philosemitism constructs the ‘Imaginary Jew’ (Finkelkraut, 1997; Wieviorka, 2018). The concept describes a fictional Jewish subject defined for a collective to dislike. It is a character that morphs depending on the desired construction required to position the Jew in opposition to what society values about itself and dependent on the context. As shown in this chapter, the positive attributes ascribed to Jewish people online often sit at the crossroads between philosemitic and negative antisemitic stereotypes.

There are four sub-types that make up this category: intellectual and cultural admiration, economic and power attribution, resilience narratives and model minority identification. These are not mutually exclusive and from the participants’ input, confirms what the literature suggests - that philosemitism online conveys the positive forms of old antisemitic tropes. These lay the conceptual groundwork for operationalising online philosemitism in an annotation guidebook at end of Chapter Three.

That these patterns persist is partly explained by the psychological architecture underlying all stereotyping. Discrimination, stereotypes and prejudice operate as a tripartite system. Stereotypes function cognitively as automatic associations between a group and a set of traits, prejudice as the affective response they generate, and discrimination as the behavioural outcome (Stangor, 2016). These associations operate whether content is positive or negative and impacts the empathy responses to out-group members (Hein et al., 2010). Where Jewish people are consistently constructed as a distinct out-group, the conditions for inhumanisation and dehumanisation are structurally reproduced. As the Conservative Rabbi postulated, most people are unaware of ‘our unconscious biases’. He believes most people are ‘not antisemites but the conditioning is that the society, the media, their upbringing, their religious faith, has conditioned them to think antisemitic thoughts.’ Online philosemitism reveals the dynamic where antisemitic structures can be reproduced through the lens of excessive admiration.

Cultural and Intellectual Praise

The first sub-type found is the intellectual and cultural praise of Jewish people and their history. The Quantitative Antisemitism Researcher explains ‘people have been saying forever...“Oh, Jews are the people of the books.”’ It appeared continuously through the sampled data. Below are two cross-platform examples. The first praises the history of

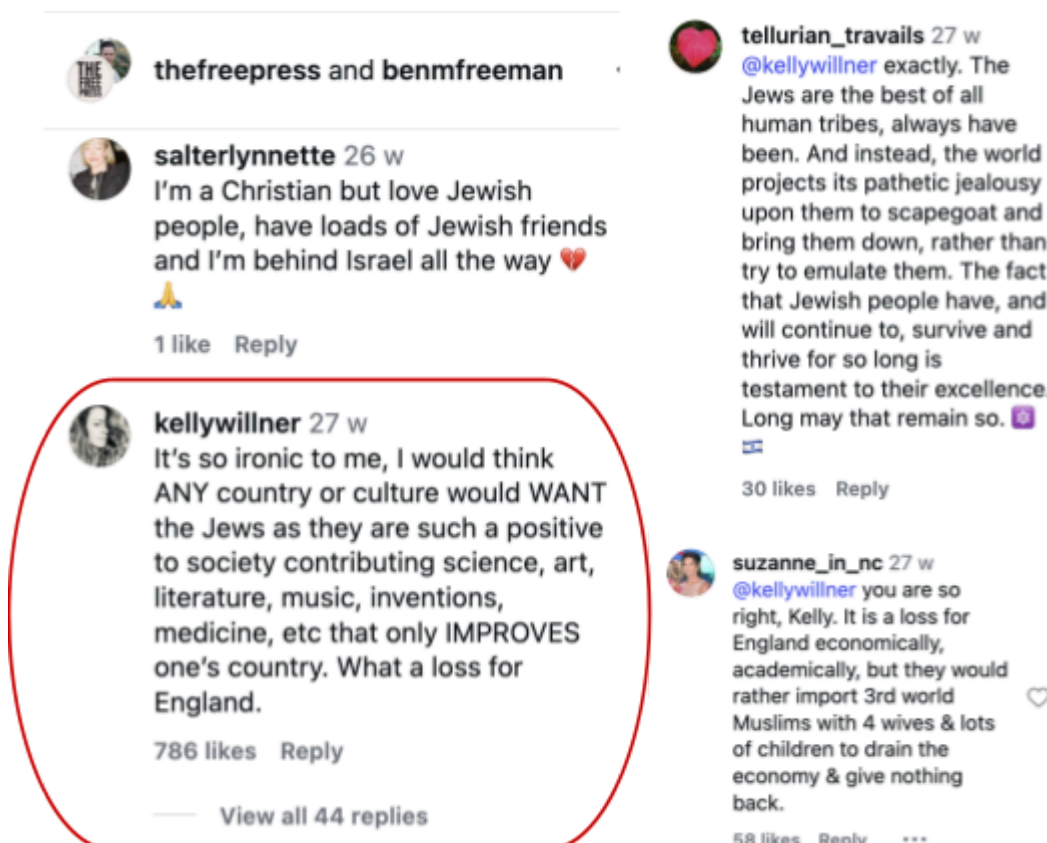
philosophical scholars in Judaism while responding to a question on a Facebook Forum about how their perception of Judaism has changed. She shares,

‘it’s been enlightening... For me, the most remarkable thing is that the Jewish sages & scholars of old developed a complex legal system with attendant features that can even rival that of contemporary nation-states. I knew it was fairly ancient when I first started learning, but to know how much it has evolved over the centuries is amazing.’

The positive linguistic field, ‘most remarkable’, ‘amazing’, ‘enlightening’, echoes Sutcliffe and Karp’s work that traces philosophical and intellectual appreciation of Jewish people back to Greco-Roman history (2011, pp.201-3). Though few comments in the Forum discussed this subtype, it was more hyperbolic online within Instagram thread comments. The second example displays this with the replies to an emotive picture headline about Jews emigrating from the UK due to rising antisemitism. Below, the commenter observes,



Figure Nine: Sampled comments from Ben Freeman’s Instagram Post (03/10/2025).



Kellywillner's comment emphasises the contributions of Jewish people to areas of high culture. The comment's high engagement reflects the platform architecture, gathering responses along the same sentiment as *suzanne_in_nc*. She points explicitly to where England will struggle more, 'economically' and 'academically' and critiquing the governments 'import[ing]' of large Muslim families. Combined with *tellurian_travails*'s response that uses the superlative, 'Jews are the best of all human tribes, always have been', the comments display how different sub-types merge to essentialise the Jew in practice. The image of Jewish people as academic, economically skilled, resilient and belonging to a different 'tribe' that is superior and in contrast to Muslims reference philosemitic sentiments identified by the literature when describing dynamics offline (Sutcliffe & Karp, 2011; Rose, 2020; Cohen 2025). Ultimately, these comments 'other' Jews and refers to a wider discussion of religion and emigration that is linked to populism and 'in-group/out-group' narratives that, even when supportive, are seen through the lens of the 'Imaginary Jew'.

This form is concerning because philosemitism can act as a precursor to antisemitism. The Quantitative Researcher hypothesised, 'How quickly can "Jews are the people of the books" turn into "Jews know more than non-Jews", then "Jews know more about Covid vaccines than non-Jews?"' Placed within the context of rising populism and nationalism in Europe that is tangled with misinformation, essentialising praise could become a gateway for fringe groups to normalising conspiracy framing (Bergmann, 2020, p.251). Conceptualising this subtype displays that it is more than just intellectual and cultural praise, operating with the same essentialising framework that underpins conspiracy narratives.

Unique Possession of Economic Skill and Power

The second sub-type subverts one of the most common antisemitic conspiracy theories that Jews control the world due to their perceived economic dominance and secret interconnected network. A quarter of the interviewees explained how non-Jews want to learn from Jewish teachings as they view Jewish knowledge and economic skills as unique and therefore useful to acquire to increase their own economic power. The Sustainability Media Expert compared Jewish people to their own Indian Brahmin identity admiring, 'Jews were running the bank when World War Two happened...they did really well. We're [Brahmin people] the same.' This converges with the model minority subtype but importantly, there was lots of platform data to evidence the economic admiration of Jews. *Casarthakuhuja*, the social media business influencer evidences this, opening with the lines,

'Do you know why Jews are so wealthy?

They're into international trade, finance, diamonds, precious metals...and every traditional industry that has made immense wealth in the world.

It's because the Jews are taught one word in their childhood that shapes their mind for the rest of their life..."Fargin"... and has a very deep meaning, which all of us must teach the next generation if we want them to be wealthy.

Fargin is a Yiddish word that means being happy for someone else's success...

(Instagram, 2.4 million views, 06/03/2026).

His viral video claims 'fargin' has led to three advantages for the Jewish community. The first is when a Jew discovers a new business opportunity, he involves his network, 'without keeping it a secret, so that everyone makes money'. The second is knowledge sharing which has a 'compounding effect on the business' and the third is 'mentorship culture' that expands to 'other countries, trades' (Casarthakuhuja, 2026).

There are many points to unpack here. 'Fargin' is not a business concept and this misinterpretation connects it to *material joy*. In Jewish teaching, 'fargin' emphasizes rejoicing in another's *incorporeal joy* (Jewish English Lexicon, 2026). The secondary clause, 'without keeping it a secret', reiterates the Global Domination conspiracy theory that claims Jews secretly manipulate events with their shared control of institutions and the economy (Chapealan, 2024, p.177). The essentialisation assumes Jewish cultural homogeneity, flattening the lived diversity of Jewish identity for a non-Jewish audience and reinforcing an in-group/out-group dynamic by relaying these differences as taught 'in their childhood', shaping Jewish minds for life. They present 'fargin' as a useful economic concept for his followers, which is replicated by other social media business strategists, such as Vusi Themekwayo (Themekwayo, 2023).⁵

However, LinkedIn users were quick to identify the Sanskrit equivalent 'Mudita' on this same 'Fargin' post. It conveys how the economic wisdom attributed exclusively to Jewish tradition is not culturally unique. The philosemitic essentialisation is not only reductive but factually constructed, projecting onto Jewish identity a distinctiveness that cross-cultural data does not support.

⁵ A link to Themekwayo's 'fargin' video can be found in the bibliography.

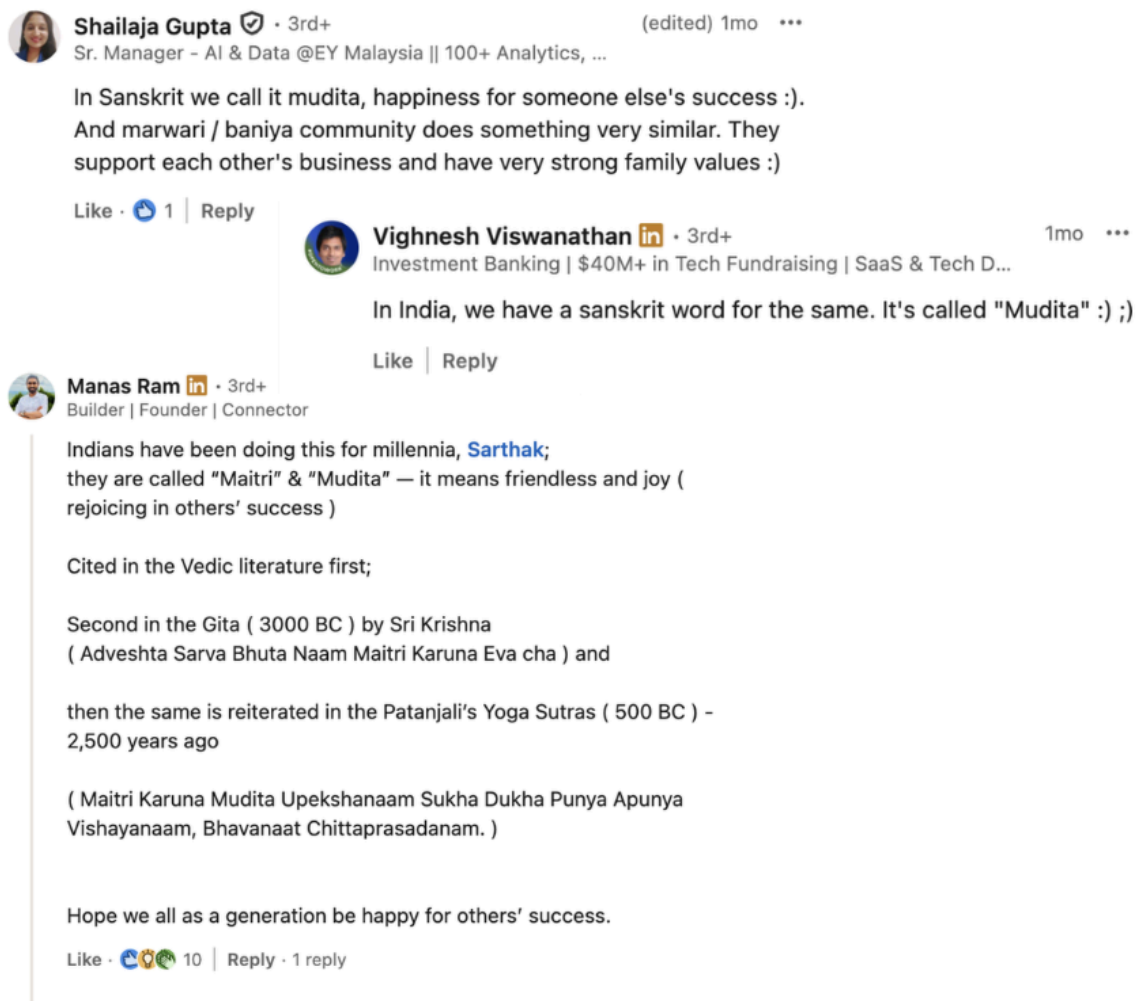


Figure Ten: LinkedIn Replies to 'Fargin' Post explaining Mudita (24/03/2026).

The viral reach of 'fargin' is just one example of how economic essentialisation has entered the mainstream. It frames the 'Imaginary Jew' as aspirational but perpetuates a philosemitic code.

A Resilient Community

The third sub-type frames Jewish people as notably resilient in the face of adversity, rooted in Holocaust history. There are many causes for this stereotyping mechanism. Most non-Jews learn about Jewish people through history education (Hale, 2025; Woo et al., 2025). The data displays this reality. One interviewee recalled how the 'Jews are perceived in India' when they 'learn about the Holocaust in our books and all that. There's a sense of great admiration...for the odds that Jewish people have overcome.' The social media data supports this too. The Facebook Forum responses to the question, 'What were you taught about Jews and Judaism growing up?' (26/02/2026, 22 responses), conveys how curriculums can unintentionally solidify a singular persecution narrative of Jewish identity. Commenters said,

‘But tbh, most of what I learned about Jewish people was about the Holocaust.’

‘Most of the knowledge I got in school relating to Jewish people was regarding just general history (ex: Holocaust).’

‘I think the first big research on Jewish ppl happened in history class when we were learning about WW2.’

This corroborates the Conservative Rabbi’s assertion that non-Jews ‘see a Jew and they have all that history and all that training but it’s condition of their perception of Jews are about’.

It is a useful backdrop to the other key aspect of this sub-type which connects to the literature on non-Jewish European guilt after the Shoah (Özyürek, 2023; Selent & Kortman, 2025). One interviewee suggests ‘Europe does bear a certain, or even great, sense of guilt for not protecting its Jewish people during the Holocaust.’ Finkelkraut’s ‘Imaginary Jew’ highlights this aspect of guilt, where Jewish memory is framed to equate to a perceived ‘memory of loss’, collapsing Jewish history into a singular victimhood narrative (Finkelkraut, 1997; Moyn, 2005). Although few interviewees identified this idea, it appeared consistently in the sampled data demonstrating its real-life relevance.

When a contemporary antisemitic attack occurs, non-Jewish people speak out online in solidarity. As Figure Eleven demonstrates, these expressions are consistent in their framing across platforms. Resilience is anchored to Holocaust memory with contemporary attacks understood through this lens of prior persecution.

Social Media Platform	Author	Text	Threat entry type
Twitter	rmhaley8	The march, marking 80 yrs since the lib of the Nazi death camps...highlighting the resilience of the Jewish people in the face of renewed trauma since the 10/7 Hamas terror attack https://t.co/1QklOir9Mw	Post
Instagram	Anonymo us	This Chanukah, we have been painfully reminded of the consequences of rising hatred. But we have also been reminded of the extraordinary resilience of the Jewish people...This striking image comes from Kiel, Germany, 1932, during the rise of the Nazis. As swastika flags were put up on the streets, Rachel Posner, wife of Rabbi Akiva Posner, captured this photograph of their family's Chanukah. #Chanukah #Hannukah #WorldWar2 #Holocaust	Post
Bluesky	montreal muffin.bs ky.social	It has failed to protect Jewish people. It's a Nazi state engaged in a holocaust and loathed throughout the world. It only puts Jewish people in danger.	Reply
Youtube	CarynDempsey-rf8ve	I stand with Israel and Jewish people. People scream the genocide of Palestine but how exactly did we forget the slaughter of jews throughout history. 6 million jews in the holocaust. Thats genocide. The murder of hundreds on October 7th. Today in Australia. When and where next. Its notif its when. This will continue. All terrorists swore to this. It will never stop until all Allies come together as one and stop them together.	Reply

Figure Eleven: Table of Sampled Supportive Comments Post-antisemitic Attacks (BSA Data, 2025).

Across platforms, expressions of solidarity invoke the Shoah as a foundational reference point, positioning present-day Jewish suffering as a continuation of historical victimhood. Phrases such as, ‘failed to protect’, ‘renewed trauma’, ‘never stop until all Allies come together’, may stem from authentic support but still retrieve the image of resilience from pre-existing symbolic associations of prior suffering. Some comments go further, politicising Jewish victimhood to legitimise contemporary political positions, reworking Holocaust memory through guilt comparison in ways that blur the line between commemoration and instrumentalisation (Rose, 2020, p.15).

This sub-type exhibits how online philosemitism can be sincere but also reduce Jewish identity to the same singular victimhood image identified across the literature. As the following sub-type demonstrates, this can extend beyond European guilt frameworks into broader cross-cultural constructions of the ‘Imaginary Jew’.

Model Minority Identification

The final sub-type, model minority identification, extends and consolidates the previous three forms of essentialisation. The finding emerges from a sole participant but offers a novel

contribution by placing online philosemitism within the Indian context. It restates a similar pattern in the literature where, in the absence of lived interaction, non-Jewish groups construct a desired version of the 'Imaginary Jew' to identify with (Schilling, 2020; Zubrzycki, 2022). The Media Sustainability Expert articulates this process,

'When I look at the admiration for the Jewish people [in India], I don't think the superiority they ascribe to Jewish people in the pecking order for non-Indians people comes from the God's chosen people thing. There is instead a connection between the Brahmin superiority complex and Jewish adoration... We're the same... They see Jewish people as getting to the States via merit and not by reservation and therefore all the Jews are just like us.'

Jewish identity is abstracted into a set of desirable attributes including resilience, economic and intellectual success. It becomes a comparative benchmark used to legitimise non-Jewish identity. This shifts slightly from essentialisation into instrumentalisation. The interviewee expands on this dynamic,

'Often, people will say, "Brahmins, we are the Jews of India. We are smart...work so hard but it's so difficult to get into medical school whereas 'dumb' Dalit students can get straight in!".... It produces this identification with caste-politics.'

Jewish identity is invoked to validate claims of merit and to critique affirmative action programmes supporting Dalit communities who 'come from rural areas with few resources'. It embeds the 'Imaginary Jew' within local systems of inequality and reflects how non-Jewish actors position themselves as analogous to Jews to articulate their own political grievances (Zubrzycki, 2022).

The online data confirms this dynamic. Figure Twelve depicts a viral Instagram about the Jews of India claiming 'India is one of the only places in the world where Jews have never been persecuted' and that the Kochi Jews have never faced any form of antisemitism (Instagram, 16/02/2026). This contradicts the Media Expert's account that Indian relations with Jews are 'a little darker', attacking Jewish buildings which means Southern Indians 'think - okay, Indians hate Jews.'

She explains there are 'only a few Jewish families left' in Kochi's Jew Town, which is now 'viewed as a tourist attraction'. Jewish culture is romanticised here and 'othered' because no actual Jewish people remain to complicate the image. This connects to the literature on post-war philosemitism, where Jewish presence is a marker of healthy democracy, where its absence would indicate declining cosmopolitanism (Von Bieberstein, 2016).

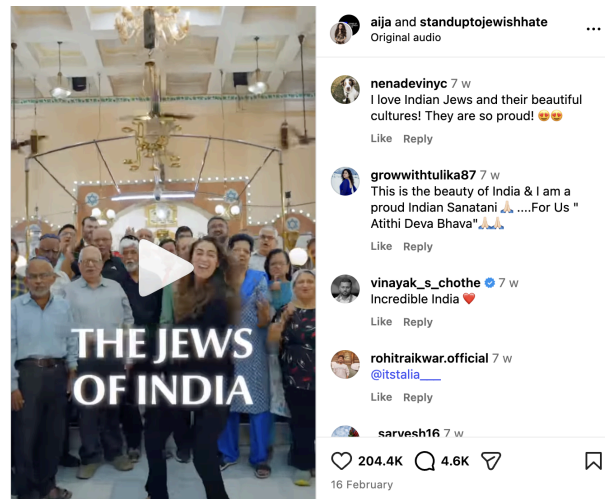


Figure Twelve: Screenshot of *aija and standuptojewishhate*'s Instagram video (5.1 million views, 16/02/2026).

Perceived cultural affinity reinforces model minority logics. The interviewee posits,

‘This admiration comes from, aside from resilience, when they meet Jewish people they think: these people are exactly like us. Intense academic pressure to succeed...well paying careers...prioritise spending on education...so for all these reasons, there's a tendency to think American Jews, at least, are like us.’

This draws on the same essentialised tropes found across the previous sub-types, including economic perspicacity and academic excellence. It is framed as admiration but returns to the same essentialised tropes that this category, to some degree, belabours. The ‘Good Jew/Bad Jew’ binary reappears as Jewish people are admired conditionally based on the conformity to these projected traits.

The interviewee shows awareness of this essentialising reality, confessing ‘I know how bad the model minority stereotypes are model minorities everywhere’. This sub-type is framed through aspiration making is unlikely to be flagged by automated detection systems. It still reproduces the same reductive logic.

Conclusion

The four sub-types confirm that online philosemitism does not operate through a single fixed image but through a flexible, context-dependent construction of the ‘Imaginary Jew’ that adapts across cultures and platforms while reproducing the same essentialising logic. Only a few interviewees raised each sub-type, depending on their professional and personal backgrounds. However, each sub-type was evidenced sufficiently in the platform data. All were reflected in the literature, except for the model minority mechanism, which has not

previously been extended to the Indian context.⁶ These sub-types are operationalised in the annotation codebook presented at the end of Chapter Three.

Category Two: Instrumental Formations

This category establishes four main instrumentalising mechanisms that expand on the philosemitic stereotyping from the previous category. Where the essentialising mechanism constructs the ‘Imaginary Jew’, the instrumentalising mechanism deploys it. Jewish identity is *actively* used by non-Jews as an object to serve broader ideological and political functions. The distribution of mechanisms identified was uneven, shown in Figure Thirteen. Religious and ideological framings dominated, while more cultural forms of instrumentalisation were comparatively marginal. This displays the professional and personal backgrounds of interviewees, each gravitating to observations more visible to themselves online.

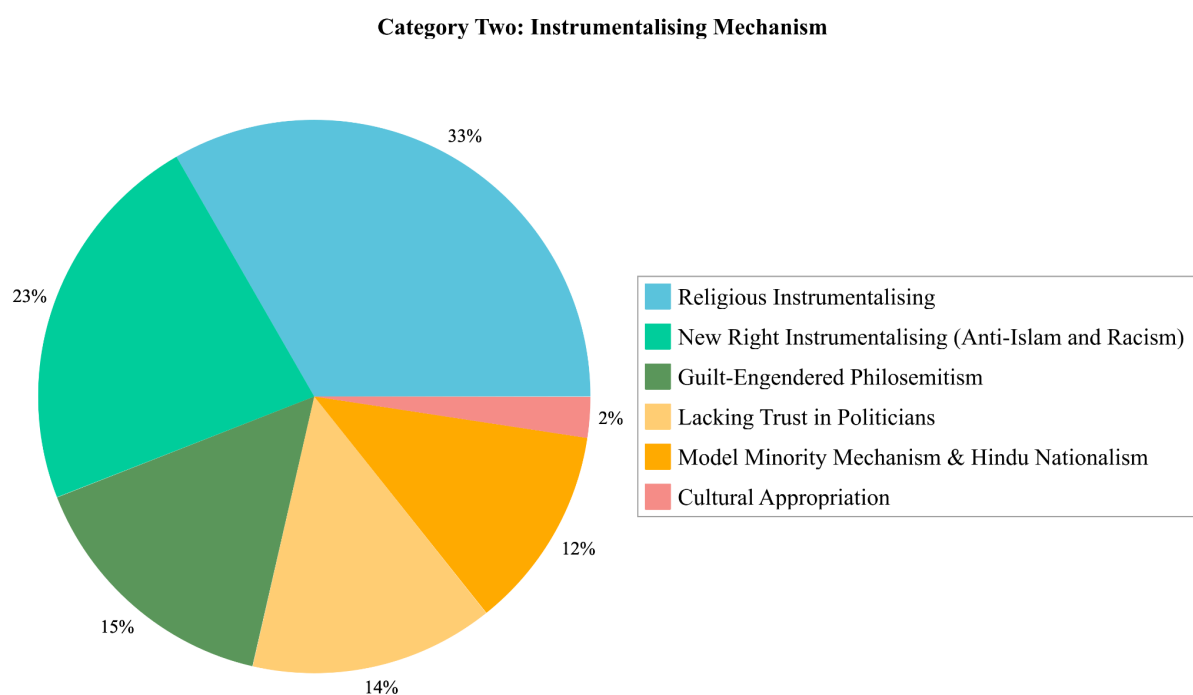


Figure Thirteen: Distribution of Category Two Sub-Codes from Interviews.⁷

⁶ There is one unpublished paper on this topic: Jai Banu Quraiza. (2004). *Hindu Pro-Zionism and Philo-Semitism*. <https://www.scribd.com/document/9680428/Hindu-Pro-Zionism-and-Philo-Semitism>

⁷ ‘Cultural appropriation’ and ‘lacking trust in politicians’ marginal sub-codes included because they function as underlying signals Jewish interviewees used to identify weaponised support.

Anti-Muslim Sentiment and the Far-Right

Three quarters of participants observed far-right instrumentalisation of Jewish people. Interviewees identified this mechanism due to its recurring attachment to anti-Muslim sentiment making it more visible than other forms of instrumentalisation. With significant exposure to online discussions of Jews, the Trust and Safety Expert found it ‘was very clear that philosemitism was ‘proxy for Islamophobia’ having ‘noticed how this strong sentiment is co-opted by some groups, especially in the US, some of the right-wing groups’. Rose’s paper on European far-right philosemitism as a proxy for Islamophobia resonated in participants’ responses (2020).

Jewish interviewees were clear about their lacking trust of far-right ‘love’ when looking at snippets of their Twitter accounts. The JSOC Chair describes how the far-right,

‘Try to use antisemitism as a tool as to why immigration is bad or whatever, which is what they believe anyway. Their goal isn’t the Jewish community here. Their goal is the Jewish community as a voting box to get whatever else that they want passed.’

The Chair conveys the reduction of Jewish voices through the object nouns, ‘tool’ and ‘voting box’, that are co-opted by the far-right’s own goal’. The Conservative Rabbi echoes this, wanting the far-right to ‘just leave us alone’ and to not ‘give us the sting, the good, the bad, just leave us out of it’ because,

‘When we start attracting friends like the far right, it ultimately will create more problems than it solves, so the problematization is not just in what they’re saying, it’s in the very fact that they feel that they have a voice to speak up for Jews at all.’

Despite the vignettes linguistically displaying support for Jewish community, the participants disregard this. It demonstrates how implicit and coded language requires a historical and cultural knowledge to detect sincerity (Weinberg et al., 2025). In the case of the Rabbi and JSOC Chair, their positionality informed their ‘problematization’ with Tweets and their identification of ‘new philosemitism’.

The social media data shown in Figure Fourteen supports this finding by documenting how antisemitic incidents serve as a catalyst for Islamophobic pivots. In these digital spaces, expressions of Jewish solidarity are used as a rhetorical shield to immediately launch attacks against Muslim communities, confirming Rose’s ‘new wave’ of philosemitism functions primarily as a proxy for broader exclusionary agendas.

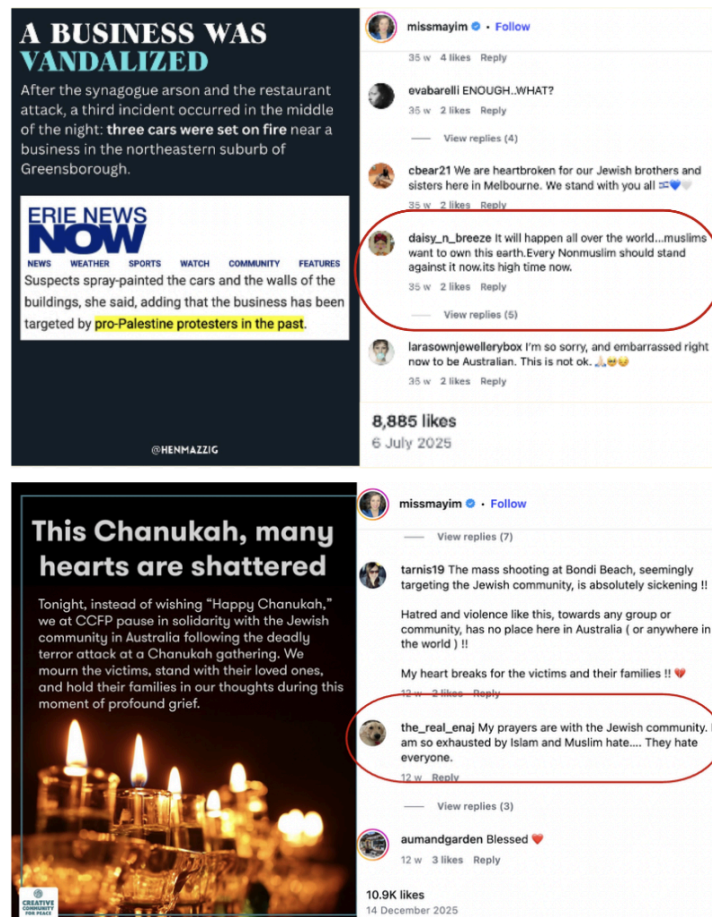


Figure Fourteen: Examples of ‘Proxy’ Philosemitism on *missmayim*’s Instagram (06/07/2025; 14/12/2025).

This has direct implications for platform governance. If philosemitism functions as an Islamophobia ‘proxy’ through a positive framing and praising of Jewish people, then existing detection systems targeting anti-Muslim hate may already contain the analytical infrastructure needed to identify certain forms of instrumental philosemitism. Merging philosemitic and Islamophobic datasets could result in strengthening both. However, the absence of a distinct philosemitism label within trust and safety vocabularies conveys this conceptual gap that limits any form of detection and response.

The Indian Context

a. Hindu Nationalism

One interviewee identified a theoretically novel form of philosemitism within Hindu nationalist discourse. She depicts how Jewish identity is instrumentalised to legitimise territorial and nationalist claims where there is a significant absence of Jews. The interviewee drew parallels between historical Kashmiri-Indian relations and the current Israel-Gaza conflict, observing the Hindu nationalist media ‘LOVES Israel’. This is because they believe Jews ‘are resilient people and deserve a land of their own...because the Hindus deserve the same...and this matches up perfectly with the BJP’s template’.

In this milieu, the ‘Imaginary Jew’ is not a recipient of support but a functional symbol used to justify domestic exclusionary agendas. The BJP uses a pro-Israel stance as a ‘playbook’ to frame Hindus as a ‘deserving population’ facing similar threats from Muslim-majority neighbours, effectively weaponising Jewish trauma as a proxy for Islamophobic domestic policies (Media Sustainability Expert Excerpts). It references many elements of philosemitic literature, from far-right anti-Muslim sentiment to the identification with Jewish people to place oneself against a majority culture (Rose, 2020; Zubrzycki, 2022).

Social media supports these observations. Figure Fifteen demonstrates a response to a post describing the perceived global isolation of Jews. *Gopalak78325085* constructs an exclusive alliance between Indian Hindus and reinforces an in-group identity that mobilises minority groups against a homogenous majority. By claiming Hindu Indians to be the ‘ONLY’ supporters, Swamy boosts the prestige of Hindu nationalism. They also frame Jewish identity through the lens of the Holocaust and indirectly criticises other political actors, exemplifying the online existence of many philosemitic forms established in the literature.



Figure Fifteen: *Gopalak78325085* response to *LauraLoomer*’s Tweet (24/12/2025).

b. Model Minority Mechanism

The interviewee distinguished a model minority mechanism emerging in India through the formation of caste-identification with Jews from Brahmin and Dalit communities. Brahmins identify with Jews to assert superiority, while Dalits use it to distance themselves from the caste system. This minority mechanism was also leveraged with the Indian diaspora, where the Indian community are no longer a majority, that identifies with and fetishises perceived Jewish economic success. It evidences how competing groups co-opt the flexible symbol of the ‘Imaginary Jew’ to serve their own identity agendas.

i. Brahmin Identification

This mechanism is operationalised through a direct identification with upper-caste Hindus and Jewish people. The Media Expert explicitly stated their positionality as Brahmin, noting a ‘Brahmin superiority complex’ where,

‘Often people will say: Brahmins, we are the Jews of India. We are smart, we are so intelligent and work so hard but it’s so difficult to get into medical school whereas ‘dumb’ Dalit students can get straight in.’

The use of the plural first-person pronoun creates an anaphoric structure that elucidates a close identification between troped Brahmin qualities and stereotyped Jewish people. The victimhood narrative compliments this direct identification and solidifies the ‘Imaginary Jew’ construct. It also articulates caste-based grievances by claiming Brahmins are the ‘Jews of India’. It ignores the Cochin Jews as the ethnic Jews of India and repurposes stereotypes of Jewish intellectual superiority to justify meritocracy claims against ‘dumb’ Dalit students, using philosemitism to erase the actual Jewish history in favour of local political projections.

Online, influencers, such as *casarthakuhuja*, echo this desire to link Jewish and Indian identities through a shared victim narrative. He asks, ‘why do so many countries hate Indians’, drawing parallels Jews and Indians were targeted because they ‘make the money money’ (Casarthakuhuja, 2026). He further leverages the perception of Jewish economic success that echoes the Global Dominant Conspiracy Theory in his videos. In Figure Sixteen, he states Jews are the ‘wealthiest community in the world, and I’ll tell you why’, fetishising their perceived success. He claims Jews are taught ‘the five jars method of money management’ from an early age which includes long-term saving to invest in children’s education and using 10% to uplift society because Jews are ‘an instrument for God’s work’. His admiration is portrayed through the cyclical structure that closes the video uttering ‘no wonder the Jews know how to make money’ and displays how across his videos, he is consistent in instrumentalising the same essentialised narratives about Jews.



Figure Sixteen: *casarthakuhuja*’s Instagram Video (09/09/2023, 2,168,345 views).

ii. Dalit Instrumentalisation

Online Dalit communities also identify with Jewish people but through religious instrumentalisation. Religious language is used to argue that Dalit oppression is somehow a misunderstanding of Jewish mysticism. *Dalit-isHebrew*'s accounts in Figure Seventeen asserts 'Dalit' derives from the Hebrew letter 'Dalet' before drawing on Kabbalistic teachings about divine emanation. This theological framework is unrelated to the Dalit experience in Hindu caste society which is a historical socio-economic reality rooted in Brahminical ideology.

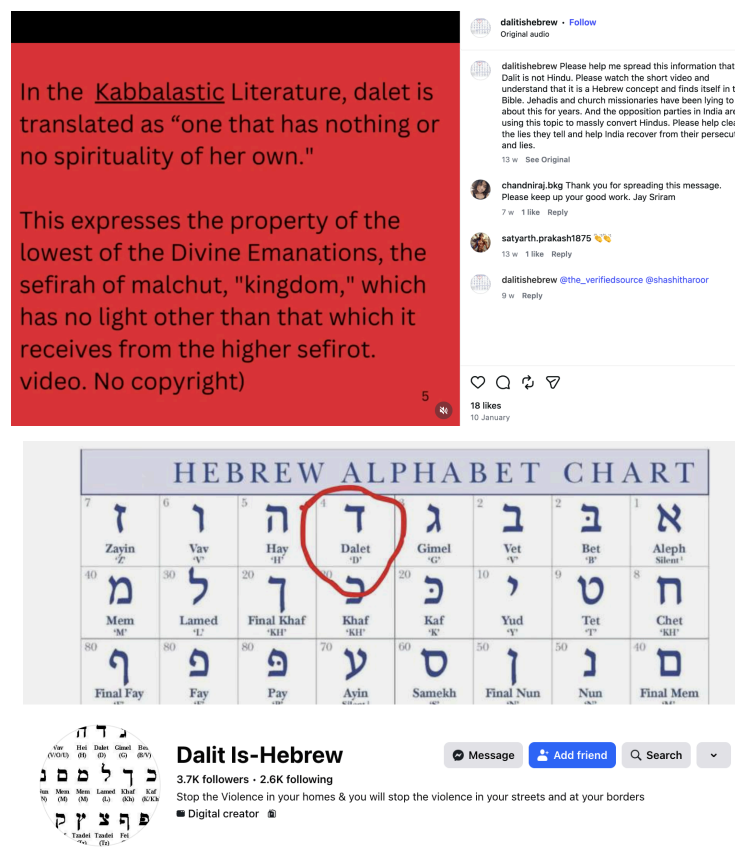


Figure Seventeen: *Dalit Is-Hebrew*'s Instagram and Facebook Profiles (10/01/2026; 16/04/2026).

Mapping *Dalitishebrew*'s cross-platform output reveals consistent attempts to delegitimise the caste system through Kabbalistic knowledge. Both *Dalithebrew* and *casarthakujuja* diversify their content across multiple platforms, repeating their ideas. This mirrors the illusory truth effect, whereby reiteration builds credibility and signals an strategic instrumental pattern.

Religious Instrumentalising

Half of the interviewees identified religious instrumentalisation as a significant concern for two core reasons. The first is Evangelical supersessionism: the belief that Christianity has substantiated and succeeded Judaism, transferring all rightful Jewish claims to Christians.

Four participants characterised this as dangerous, with the Conservative Rabbi defining it as fundamentally ‘antisemitic’ by denying Jewish people the rights to their own identity or land. The second formation identified is the far-right’s promotion of shared “Judeo-Christian values”, instrumentalised to further establish anti-Muslim sentiment.

a. Supersessionism

Interviewees contested supersessionism’s deceptive surface presence of ‘love’ makes it challenging to detect. Drawing on 20 years of moderation experience, the Forum Moderator explains,

‘Why is that dangerous? What's the problem when they're coming and they're saying that they love me? I understand that because sometimes you'll have someone who's fervently Christian saying “I love Jews,” and they're saying, “I love Jews,” because Jesus was Jewish.

And I'm like, lovely, I'm not a relic in a museum but okay. I am a living, breathing human being who's not a zoo object and I'm not the relic of what the Judaism that Jesus practised 2000 years ago. So don't look at me because I'm an Orthodox Jew as the same as Jesus. There's sometimes this Christian philosemitism comes in because they think “I love Jesus and therefore I love Jews because Jesus was Jewish and so therefore I love Jews.” That's a little bit uncomfortable.’

Her question ‘what’s the problem..they’re saying that they love me?’ captures how the divorce of linguistic form from its meaning makes identifying its instrumentalising narrative challenging for those without religious or cultural context. It is another example of how implicit or coded language requires prior knowledge to see it which online environments deny (Weinberg et al., 2025). The Trust and Safety Expert explained how the link of religious philosemitism to Christian movements 'was unknown for me because I'm not a very religious person so would not have dug into it'. The positive linguistic framing raises the detection threshold, in SDT terms, to the point where instrumentalising content is not registered as a signal at all by those with low sensitivity.

Supersessionist misinformation is deeply embedded within some Christian education frameworks. One Facebook Forum member was taught Jews were ‘unfulfilled Christians’ who ‘were monotheistic people hoping and praying and waiting for some other version of Jesus to show up’. Another contributor recalled the teaching that Jesus had to die as the ‘Ultimate Sacrifice’, learning later that ‘offerings/sacrifices for repentance of sin is not a thing in Jewish beliefs’. Early exposure to these ‘inaccurate’ frameworks calibrates the detection criterion from the outset, meaning instrumentalising content is not missed due to inattention but because it was never learned to be a signal in the first place.

b. Politicising Religion

Interviewees and social media data support Rose's assertion that Judeo-Christian are increasingly used by populist parties (2020). The appropriation of the 'Judeo-Christian' label by populist parties like Reform UK represents the political operationalising of this religious instrumentalising mechanism. In response to the open Iftar held in Trafalgar Square, which was open to all religious and non-religious people to join, Nigel Farage tweeted,

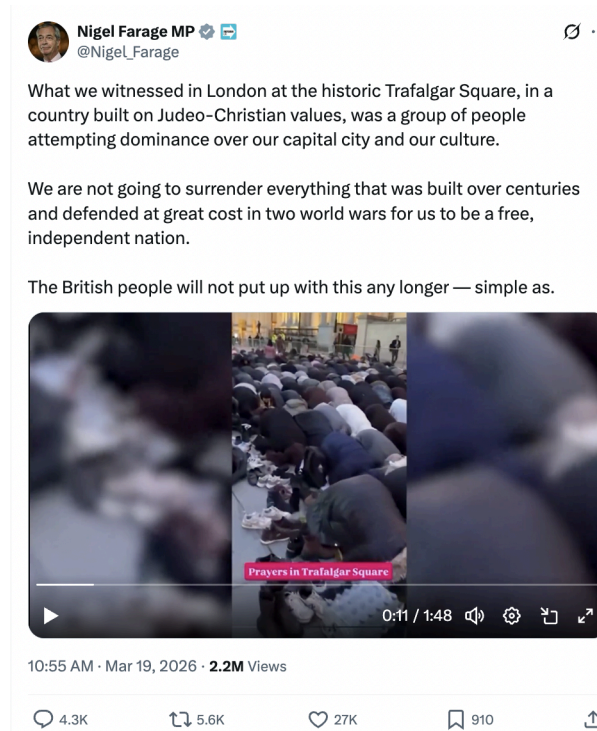


Figure Eighteen: *Nigel_Farage*'s Tweet about Trafalgar Square's Open Iftar. (26/03/2026, 2.2 million views).

Farage's plural personal pronouns, 'our', spotlights his deployment of Jewish identity as a co-signatory of a shared cultural heritage that excludes Muslims. His instrumentalisation is evident as it contrasts real Jewish reactions as religious communities publicly supported the Iftar (Arora, 2026). Farage is nescient in that British Jewish history was dominated by their expulsion for over 360 years, spurred by Christian haters (Byrne, 2017, p. 1). Still, far-right philosemitism appeared 'clearer' for non-religious participants than supersessionism as they display more overt signals in their religious instrumentalisation (Trust Expert).

Guilt-Engendered Philosemitism

The majority of interviewees did not spontaneously name post-Holocaust guilt as a driver of philosemitic expression suggesting a gap between lay and academic recognition of guilt-engendered philosemitism. Only one interviewee recognised 'Europe does bear a certain, or even great sense of guilt, for not protecting its Jewish people during the Holocaust'. This disconnect reflects how guilt-engendered philosemitism in online spaces can be absorbed into and indistinguishable from other instrumental forms (Rose, 2020).

The online dimension is demonstrated by waves of pro-Jewish content following antisemitic attacks and Holocaust remembrance posts. Figure Nineteen illustrates two examples. The first depicts supportive semantic fields, such as ‘I love you my sister’, as commenters verbalised solidarity with the Australian Jewish community following the Bondi attack in 2025. The second portrays messages of thanks and solidarity in response to a birthday post of a Holocaust survivor. Both posts connect to Israel either through language or emojis, reflecting a recurring pattern across comment sections in which Jewish identity and Israeli nationhood are treated as interchangeable when expressing support.

While support appears sincere, the posts’ high engagement and broad reach mean they function algorithmically in the same way as instrumental content, amplifying an essentialised image of Jewish identity at scale that is defined by suffering, resilience and victimhood.



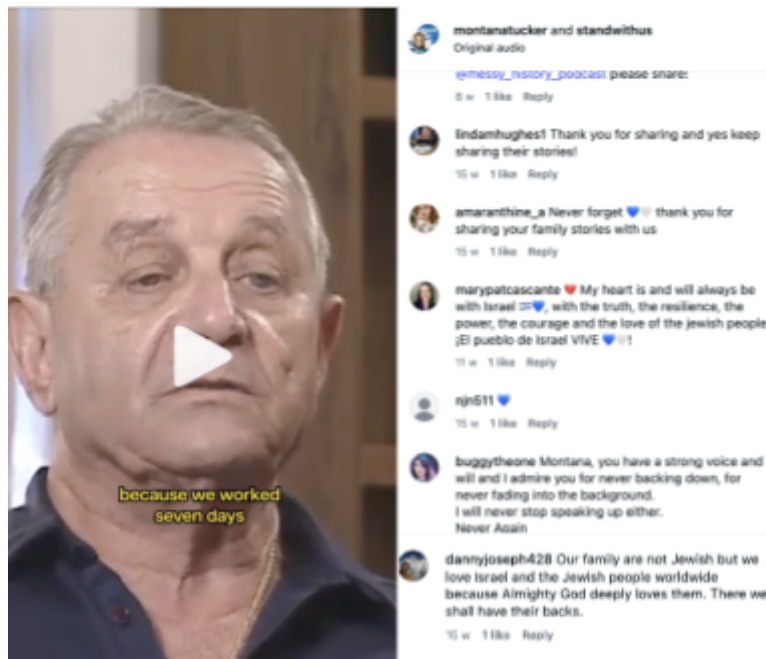


Figure Nineteen: *missmayim*'s Instagram post (15/12/2025, 42,400 likes) and *montanatucker* and *standwithus* Instagram post (09/11/2025; 47,300 likes).

The simultaneous rise in antisemitic content after October 7th further complicates the picture as support may be performative in its visible counter-signal to antisemitism. The Conservative Rabbi recalls from his own experience of being instrumentalised by a Muslim leader that reposted photographs of them praying together at Auschwitz. He explains that 'someone in the picture was praying as a defence of his attitude towards Jews, when he was clearly being antisemitic as well'. This aligns with Rose's identification of philosemitism as a strategy of distancing from historical antisemitism and avoidance of any association with Nazi or neo-Nazi views (2020, p.18).

Conclusion

The analysis of instrumental formations demonstrates how the 'Imaginary Jew' becomes an active discursive resource, mobilised across political, religious and cultural contexts to serve external agendas. These mechanisms extend beyond essentialising representation into strategic deployment, confirming H1 by evidencing typologically distinct forms that diverge from traditional Eurocentric frameworks. At the same time, the persistence of reductive tropes within these deployments supports H2, as even positive appearing solidaristic expressions reproduce essentialised constructions of Jewish identity. The diversity of these instrumental uses highlights both the global adaptability and contextual specificity of online philosemitism.

Categories One and Two establish that online philosemitism operates through a dual process of construction and deployment, in which flexible yet consistently essentialised images of Jewish identity circulate across digital environments. This reinforces the need for a more precise conceptual and operational framework capable of distinguishing between supportive

expression and instrumentalising discourse. Addressing this challenge, the following chapter develops an annotation guideline to systematically label these forms in online data, while also engaging with H3 by examining the limitations of existing detection systems in capturing positively framed content.

Chapter Three

From Discourse to Detection

Chapter two focused on the conceptualising of online philosemitism through essentialisation and instrumentalisation of Jewish people. Chapter Three continues this conceptualisation while addressing the second part of the core research question: how to operationalise it. Central to the creation of an annotation guideline, this chapter identifies how participants recognise sincerity and highlights the challenges of detecting authenticity online.

Category Three: The Authenticity Threshold

Defining the Threshold

Online philosemitism is difficult to decipher because there is no universal framework to determine what is true versus performative admiration for Jewish people and drawing a line is wildly subjective. Interviewees struggled to define the term philosemitism but when asked to distinguish between real and ingenuine support, they were articulate in their views. 28% of Category Three responses were suggestions of signals to detect differences between authentic and inauthentic behaviour online. From these, four common markers for genuine behaviour emerged: consistency, inconvenience, action-orientated and selfless behaviour.

However, this ‘test’ of authentic behaviour only partially applies to online environments due to the architecture of social media platforms. With little historical context on individual posts, this criteria fails as performativity is rewarded online making detection challenging. This category draws on Signal Detection Theory (SDT) to explore the differences between participant responses and how to operationalise these findings.

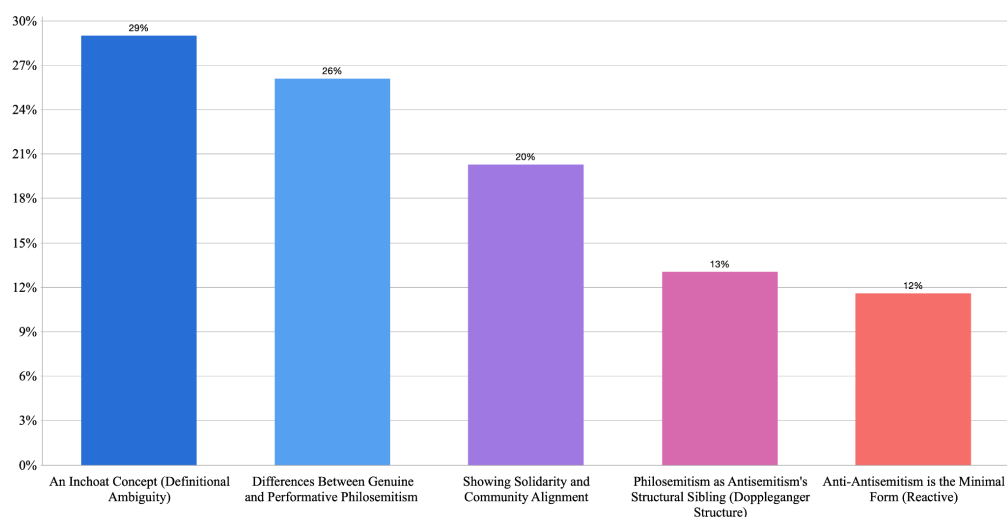


Figure Twenty: The Distribution of Category Three Initial Codes from the Interviews.

The ‘Authenticity Test’: Four Markers

No matter the ideological stance of participants, they returned to this criteria of consistency, inconvenience, action-orientated and selfless behaviour to the sincerity of support. Interviewees viewed performativity through symbolic support that lacks consistent action between online and offline contexts. The Trust and Safety Expert explains ‘most platforms also have offline behaviour policies’ for creators to provide this same consistency for their followers and be responsible influencers. Similarly, when raising awareness about antisemitism on Instagram, the JSOC Chair shared, ‘I’d see my friends later that day who I knew had seen my story and I’d know had liked it, but then we don’t do anything or talk about it in person. I found it a really strange disconnect that just ended up with me being frustrated and upset all the time’. When behaviour suddenly changes, an individual’s detection threshold shifts as they become more sensitive to insincere motives by identifying more ‘hits’ (Heeger, 1997, p.1).

Jewish interviewees were especially sensitive to any signals of insincerity. The JSOC chair casts authentic support ‘to an extent when it’s inconvenient’ rather than ‘to aid your own goals’. They ask, ‘are you actually listening to the Jewish community?’ subverting the potential for self-serving behaviour,

‘I think there’s a lot of tokenizing voices or publicity stunts. For example, people will light Shabbos candles with some rabbi and they’re going to be photographed doing it and it’s going to be ‘oh, look at me, I support the Jewish community!’ But you’re not actually listening to the Jewish community...You’re not actually doing anything tangible about it...I look at action more than anything.’

Other Jewish participants objected to virtue signalling, perceiving sincerity as active resistance rather than symbolic coded behaviour. The Forum Moderator asks, ‘are you going to actively fight Jew hatred?...Or are you just going to wear this object [Magen David] because you want to show your solidarity, but it’s meaningless?’ The Liberal Rabbi further rejected multiple requests by Members of Parliament to speak to the congregation after a post-October 7th surge in antisemitism, stating ‘I’m not here as your mouthpiece’. These are clear examples of when non-Jewish “allyship” does not fall under the four markers of authentic behaviour, ordaining it symbolic philosemitism.

Non-Jewish participants do not display such strong sensitivity to these signals of sincerity. They focused on the public utility of support rather than its behavioural depth or underlying consistency. The Antisemitism Awareness Educator and CAA Representative found the current socio-economic climate makes any vocal support for the Jewish community ‘brave’. The Educator argued ‘it doesn’t do anybody any personal good to support the Jewish community’. She cited moments where individuals who speak out require security due to harassment, concluding that a supporter would not invite such ‘inconvenience’ unless they were genuine. They question, ‘why would they do it otherwise?’

For the CAA, their primary goal is the public rejection of antisemitism. They believe ‘it is entirely reasonable for politicians and other leaders...to express their solidarity with the Jewish community and opposition to antisemitism’, to ensure the community does not feel alone. They extend this to controversial far-right figures, such as Geert Wilders. While other participants fear these alliances, the CAA evaluates this public solidarity as desirable and genuine, displaying how positionality impacts the fear of instrumentalisation and sensitivity to ‘hit’ signals.

Confidence

The majority of participants felt rather than analytically determined a detection threshold for authenticity. Only one interviewee was confident in their ability to determine genuine support. With twenty years of moderation experience, the Forum Moderator stated,

‘I have a pretty good track record of predicting who’s going to go off the rails or not...It’s not just me but we all know when one’s not going to last long [on the forum]. We always even say it. We have running bets’.

To them, sincere advocacy goes beyond the four markers of authenticity. They assert non-Jews must acquire ‘insider knowledge’ so ‘they can recognise antisemitism’ (Weinberg et al., 2025). Long-term engagement with the community emerges as an indicator of sincerity as non-Jews develop the capacity to recognise subtle ‘dog whistles’ that can signal an undercurrent of instrumentalisation, particularly as their understanding deepens about Jewish history and experiences.

Through the behavioural lens of SDT, closed online spaces with stricter moderation practices are likely to be perceived as less lower-threat environments than the mainstream norm. This means decision-bias shifts closed to a neutral baseline, enabling more analytical rather than affective judgements. The Moderator’s private Facebook forum demonstrates how reduced environmental threat contributes to greater confidence in assessing authenticity compared to other participants. They avoid public online spaces because it ‘just gets ugly’. It shows how reduced environmental noise raises sensitivity, enabling the detection of subtle instrumentalising signals that are drowned out in higher-noise public spaces, increasing the interviewee’s confidence in their detection abilities.

Threshold Instability

To test the interpretive stability of philosemitic content, interviewees were presented with a series of vignettes. The most revealing was a 2022 statement by former British Prime Minister Liz Truss regarding ‘Jewish values’ which evoked the most divergent responses and displayed the challenge of subjectivity for automated detection.

Interviewees varied with the problematising of the statement. The Conservative Rabbi did not find it inherently offensive stating, ‘I don’t have a problem’, despite acknowledging it

referenced stereotypes that ‘perhaps isn’t always helpful.’ The JSOC Chair found the statement more homogenising, declaring Truss was ‘she’s picking and choosing’ certain Jewish values to appeal to voters. The Liberal Rabbi agreed that Truss expects ‘Jews to be homogenous and we’re not in anyway...including politically.’ None took strong issue with the statement, but their slight variations mean classifying the statement confidently as philosemitic instrumentalisation would be discouraged under conservative coding principles (Becker et al., 2026).

Other interviewees were stronger in rejecting Truss’s statement. The Forum Moderator explained Truss ‘is seeing the things that she wants to see as common and she’s projecting her own perception of Jewish stereotypes...why does that have to be specific to Judaism? That’s not a Jewish value.’ The Media Sustainability Expert also took issue with the subtle way the statement ‘pits the Jewish community against others’, arguing Truss creates a hierarchy that fails to recognise the contributions of other communities because she does not believe ‘setting up a corner shop is as desirable or prestigious as setting up a legal practice’. These negative interpretations are more lucid in their declaring Jews as this an essentialising mechanism for Truss’s own political agenda.

These comments illustrate how positionality drives threshold assessment, a human element that is fundamentally unavailable to automated detection systems. To allow for interpretative flexibility, the annotation guideline incorporates Becker et al.’s framework requiring coders to distinguish between explicit, implicit, grey area and co-text examples (Becker & Troschke, 2023; Becker et al., 2026). This acknowledges that different coders will interpret posts differently depending on their positionality and exposure to Jewish essentialised tropes and that the offline test for authenticity does not work online. By providing specific inclusion and exclusion criteria with illustrative examples, just like in the IHRA definition, the guideline acknowledges that the line between genuine support and performance remains fluid. Yet, it can still be operationalised with this integration of interpretive sensitivity will improve future automated monitoring and refining of the phenomenon.

‘Anti-Antisemitism’ Minimal Threshold

The interviewees also identified a minimal threshold for authentic behaviour that falls under the definition of ‘anti-antisemitism’ (Judaken, 2006, cited in Rose, 2020, p.5). The four markers test the depth of commitment, while anti-antisemitism ascribes a baseline of solidarity. It is the active opposition to Jewish hatred rather than excessive admiration for the community. A quarter of participants placed this above their concern of philosemitism. The Forum Moderator explains, ‘dude, I don’t care about appropriation. I care if you are trying to kill more...When push comes to shove, are you really going to speak up and protect Jews?’ The Conservative Rabbi reduces it further, saying ‘really all it [support] looks like is stopping the haters.’

This baseline is useful because it is observable. Where the four markers require sustained behavioural evidence unavailable in decontextualised online posts, anti-antisemitism can be

inferred from a single comment. This makes it a signal for the community when the fuller authenticity test is void.

However, the minimum threshold is limited. The Liberal Rabbi recalled the surge of support following the Yom Kippur attack in Manchester this year,

‘In some ways, the cards that are left, the flowers are left, they’re quite nice. Although, it felt like leaving flowers on a grave. It just felt a bit eerie. I think we hadn’t suffered directly but people are leaving flowers as if people within our community had been killed, stopping a terrorist getting in and killing lots of Jews. So it was all a bit strange.’

These displays of solidarity appear genuine in intent and meet the anti-antisemitism threshold yet still felt imposed for the Rabbi. The ‘eerie’ disconnect arose because the gesture centred non-Jewish distress rather than the community’s needs, reacting to a homogenised image of Jewish resilience and suffering rather than listening to the community itself. As in Chapter Two, symbolic solidarity can cause structural harm, regardless of supporters’ intentions. It reproduces the same essentialising dynamic, however sincerely motivated, making anti-antisemitism a useful indicator but not a standard.

Conclusion

Instability is expected for an inchoated phenomenon evolving in fast-paced online environments. This is a productive finding for annotation design. Rather than apply binary philosemitic/non-philosemitic classifications rendered unreliable by threshold instability, the guideline will adopt Becker et al.’s approach distinguishing between explicit, implicit, grey-area and co-text cases.

This structured flexibility reflects the empirical reality that differently positioned annotators may interpret the same content divergently. By recording classifications alongside confidence scores, the guideline makes this instability legible and measurable. In doing so, it addresses both elements of the research question: conceptualising philosemitism as context-dependent, and operationalising that instability within annotation guidelines, recognising that detection depends on prior conceptual construction.

Category Four: A Detection Gap

This category unpacks the difficulties laced with classifying philosemitism for automated detection. The phenomenon is implicit by nature, surrounded with a positive linguistic shield that makes detecting underlying negative sentiment challenging. Though this section identifies the most viable detection models for addressing this, interviewees were most concerned about the platform architecture and threshold instability, shown in Figure

Twenty-One, confirming it is an architectural and definitional problem as well as a technical one.

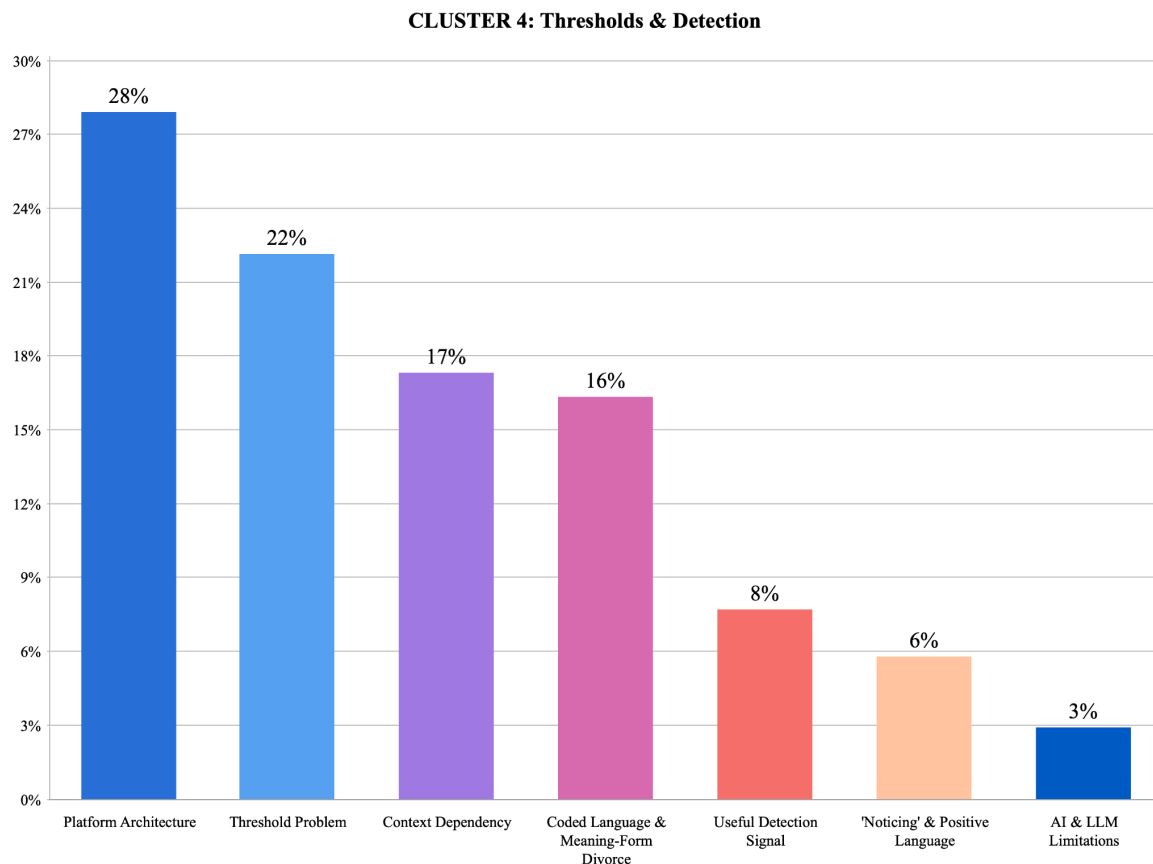


Figure Twenty-One: Distribution of Category Four Initial Codes from Interviews.

This category presents a moderation paradox explaining why philosemitism is more likely to appear online in some areas of the world than others. However, this is not to be misunderstood with the argument that philosemitism *should* be moderated. As clearly stated by interviewees, this category finds philosemitism useful to flag rather than censor as this brings in complicated discussion around free and hate speech. Altogether, these findings operationalise the conceptualisation of online philosemitism through an annotation guideline, the first step to addressing the large gap in the quantitative research of the phenomenon. With this, a subject deemed too immersed in ambiguity no longer must be devoid of having its own labelled dataset.

The Structural Detection Gap

A definitional gap must be filled to include philosemitism in automated detection systems. Interviewees from all backgrounds displayed skepticism during interviews about the current capacity of automated tools to identify philosemitic content. This was mainly because overt antisemitism has not yet been sufficiently addressed online and continues to rise. Unlike in the literature, participants saw more of a disconnect between the two phenomena and

therefore sought for one, antisemitism, to be resolved first before philosemitism was brought to the table.

This makes awareness a prerequisite for detection. As Chapter Two displayed, Jewish people are essentialised and instrumentalised through an online version of the ‘Imaginary Jew’ (Wieviorka, 2018). These images will continue to be disseminated and feed more violent discussions of the Jewish community that can translate back to offline contexts (Bjola & Manor, 2020). The importance of awareness is explained by the Conservative Rabbi who, when asked if new forms of antisemitism had appeared online in recent years, said overall no but ‘we’ve become more aware’. He expands,

‘Let’s take the term gaslighting, for example, which is fairly new ... people say to us things we may not have thought of as being antisemitic ten years ago but now we understand actually is antisemitic. It’s not the increase in antisemitism but it’s an increased sensitivity to what is antisemitic that therefore we now would probably challenge in a way we may not have done ten, even five years ago.’

Philosemitism is a source for structural discrimination as it is part of the process of ‘othering’. This has a long-term indirect effect whereby the extensive ‘othering’ of the ‘Imaginary Jew’ dominates the complexities of reality. It risks non-Jews, as well as Jewish people themselves, coming to understand the Jewish community through a lens that perpetuates hegemonic powers. Thus, spreading awareness of philosemitism is one way to untangle this ‘othering’ process, uncovering this hidden form of antisemitism to be uncovered.

This human detection gap is directly reproduced in current models. Current state-of-the-art models, including transformer architectures like BERT, LLMs or multimodal LLMs (MLLMs), are trained mainly on datasets of explicit hostility and are likely to treat positive linguistic valence as a marker of “categorically safe” content (Becker et al., 2026). They already struggle with implicit and coded language (Weinberg et al., 2025), so unless fine-tuned, this is unlikely to be remedied (Steffen, Pustet & Mihaljevic, 2024). Whilst this is a qualitative study, it brings to the forefront the need for quantitative expansion to refine the possibilities of online philosemitic detection through high-quality annotated datasets that capture positive-valenced prejudice.

The Challenges for Automation

a. The ‘Positive Sentiment’ Shield

One key reason why automating philosemitic detection is so difficult is its surface-level positive linguistic appearance. The Online Antisemitism Professor explains how philosemites avoid accountability by ‘saying, “I love them!”’ as ‘it’s easier to do and harder for AI to track them. It [AI] won’t look at it as hate speech. It will see it as a regular thing. This is why I am

saying you need to have a human being included with these tools.’ There are LLMs that have been programmed to make decisions more aligned with human judgement (Davidson et al., 2017). However, the Quantitative Researcher explained how philosemitism is even more nuanced than antisemitism, ‘I think philosemitism is probably harder [to detect]...I’ve classified antisemitism I don’t know how many times and I find that once I go behind definitions, such as the IHRA definition, there are a clear 12 labels you can give. I really struggle with labelling philosemitism well.’ If humans ‘struggle with labelling philosemitism well’ then programming a machine to replicate classification becomes impossible.

b. Meaning-Form Divorce

Social media platforms lack context. They are fertile ground for the meaning-form divorce process where the decoupling of the sign and referent means that hostile meanings can move away from its supportive words. The JSOC Chair exemplified this when their close friend made offensive jokes about the Rothschild conspiracy, saying on a WhatsApp group chat, “the Rothschilds need to stop making me go to parties so I can do my dissertation.” When informed this was part of an antisemitic conspiracy theory, he apologised explaining, “I must have spent far too long on social media because I had no idea that was an antisemitism thing.” It displays the impact of the illusory truth effect, where if you consume information repeatedly, you begin to believe it is true (Urdry & Barber, 2024). Philosemitism does not announce its harm which is why the following detection strategies prioritise structural and contextual signals over surface sentiment built for binary classification.

c. Coded Language

Coded and implicit expression is used to avoid content moderation on social media (Becker & Troschke, 2022). In antisemitic contexts, this operates through substitution to evade keyword detection. The Antisemitism AI Researcher explained, for example, the juice box emoji is commonly used to refer to “Jews”. The mechanism is more complex with philosemitism, as a hostile stereotype is repackaged inside an admiring register. Phrases celebrating Jewish intellectual or financial success, for instance, can circulate as apparent compliments while encoding the same essentialising logic as explicitly antisemitic content. The surface form is positive, the cultural reference is hostile, and neither layer identifies itself as such.

This distinction matters as it identifies two separate problems. The positive sentiment shield is a classification problem. Coded philosemitism language is a referential problem. Coded language ‘requires prior historical or cultural knowledge’ that must be actively built into labelling infrastructure rather than assumed (Becker et al., 2024, p.2) which is what the annotation guideline produced attempts to encode structurally.

How to Combat these Challenges

a. Operationalising Signals for Detection

One approach to using critical detection signals is the ‘noticing’ grammatical structure. ‘Noticing’ is the framing of Jewish success as suspicious, popularised under the hashtag #TheNoticing (ADL, 2025). The AI Researcher declined regularly seeing explicit philosemitism online unless connected to the ‘noticing’, saying, ‘If you take pure philosemitism and add ‘noticing’ on top of it, I see that a lot...For example, “Oh my God, there were four movies about the Holocaust this year...That’s so many.”’ This language structure appeared sixteen times in the BSA dataset that praised the ‘power of the Jews...’. Without knowledge of the Global Dominant Conspiracy Theory, the words seem complementary. However, this syntactic pattern can be operationalised as a training feature for detection as praising the Jewish community is contraposed by the grammar indicating suspicion and displays the contributor’s ulterior sentiment.

Another step is to add labelled data. Labelled data ‘adds enough context to disambiguate some of those grey zones’, explains the AI Researcher. Context-aware prompting is already being developed to add nuance to classification and explanation tasks (Jin et al., 2025). The Quantitative Researcher confirms this explaining, the more context provided for an LLM to label something, the better it performs which would be useful for philosemitism because ‘I wouldn’t know how much LLMs know about philosemitism.’

This contextualising involves labelling gateway tropes. The previous chapter displayed how philosemitism is built on the same essentialising stereotypes as antisemitism. Although ‘philosemitism can take new forms’, the Quantitative Research postulates, ‘I’d say this [philosemitic expression] is rather permanent.’ As ‘we don’t know what’s going to happen from now’, developers can be perspicacious, training LLMs to recognise antisemitic tropes and pinpoint these new philosemitic formations, such as the Happy Merchant.

Conservative coding is a key consideration in operationalising detection signals. To minimise false positives, even with a contextualised dataset with examples of gateway tropes and ‘noticing’ language structures, confidence must somehow be determined. This is resolved through the technique of conservative coding, explained by the Antisemitism AI Researcher when an ‘example could plausibly be interpreted as antisemitic or not antisemitic, we would annotate it officially as not antisemitic to give the benefit of the doubt or just to be on the safer side’. The inter-coder decision to conservatively code posts that fall outside of the IHRA and Becker’s Lexicon definition of antisemitism (Becker et al., 2024), ‘happens surprisingly a lot, or maybe unsurprisingly, in a real social media environment’ and therefore must be considered for future philosemitic research. Together, these create the annotation guideline’s operational backbone, making the grey zones of philosemitism legible.

b. Selecting the Correct Tool: Textual vs Behavioural Models

Given the novelty of philosemitism as a detection category and the structural limitations of textual classifiers on coded language, a hybrid approach is most appropriate. Textual models offer the most immediate, scalable route to detection whereas behavioural models represent the more structurally robust long-term solution for monitoring how philosemitism circulates online within different communities.

LLMs are increasingly centered in the large-scale monitoring of online content. The Quantitative Researcher notes computer scientists used neural networks, then trained transformer models like BERT and now ‘everyone is prompting LLMs’ because ‘LLMs have by far the best performance’. Claude was found to be the best at identifying antisemitism, outperforming competing models (ADL, 2026). Despite this progress, the majority of automatic detection programmes remain insufficient in their classification of *implicit* antisemitism, struggling with accuracy, cost and the sheer volume of antisemitic content online (Bjola & Manor, 2020; Chared et al., 2025). For example, PerspectiveAPI has shown limitations in detecting contextually nuanced or coded antisemitism (Steffen, Pustet & Mihaljevic, 2024). This is why the annotation guideline produced prioritises metadata over surface sentiment, as the failure of existing models is a failure of context, not vocabulary.

This trajectory matters for the detection of philosemitism specifically. If models are being evaluated on their capacity to identify antisemitism, it is important that philosemitism - as its ideological companion and, in some cases, its precursor - is incorporated into these developments so that models can be fine-tuned to close the current detection gap.

The performance of closed-weight, commercial LLMs is structurally unreliable. The Quantitative Researcher warns that model updates can unpredictably degrade performance and because training data is never disclosed, researchers cannot diagnose why. For philosemitism, where no training data currently exists at all, this opacity is particularly consequential: a model cannot be evaluated on a category it has never encountered and any apparent competence at detecting adjacent phenomena like antisemitism cannot be assumed to transfer.

Participants also evaluated philosemitism as a behavioural signal - one that may be detectable by who says what and how their network develops over time. The Quantitative Researcher identified it as a foreshadowing of overt antisemitism stating, ‘it’s something to look out for because it’s certainly an indicator of discussions turning to a certain point or people turning to certain arguments.’ The Trust and Safety Expert extended this, positing could function as a dog whistle, ‘it would be interesting to add this as a signal so...when Christian right-wing groups co-opt this or support this then we’ll know it’s a dog whistle for something else.’ These thoughts elucidate the idea that automated behavioural models may be better equipped than text classifiers to monitor the structural spread of philosemitism online. Dog whistles, by definition, are designed to pass surface-level detection (Kruk et al., 2024; Mendelsohn et al.,

2023). If philosemitism functions as a dog whistle then it will be resistant to commonly used classifiers which is why behavioural models would be indispensable.

Graph Neural Networks (GNNs), for example, are language-independent and well-suited to detecting antisocial behavioural patterns and coordination networks (Khan et al., 2024). Applied to philosemitism, GNN models could be pertinent for examining whether seemingly positive discourse about Jewish people cluster around actors whose broader behaviour is associated with the European new far-right parties or the case of the Hindu nationalist online spaces discussed in Chapter One. Long Short-Term Memory (LSTM) models offer another path as they effectively detect repeated themes, tonal escalations and ideological drift over time (Nagar et al., 2023). Even if a singular post appeared positive, these models would reveal ideological clustering and repeated discourse patterns that can be linked to political or extreme communities.

Although GNN and LSTM models require large-labelled datasets, which do not currently exist in the world of philosemitism, this is not an insurmountable obstacle. Emerging techniques in synthetic data generation for low-resource speech contexts offer a path around this (Girón et al., 2025). Transferring learning from adjacent, in this case antisemitism, datasets can also improve detection performance (Ghorbanpour, Dementieva & Fraser, 2025). A philosemitism corpus could be generated through this combination of LLM-generated synthetic examples and existing antisemitism datasets and targeted crowdsourced annotation to make model training a methodological reality.

As the Online Antisemitism Professor explained, 'the far-right are harder to see sometimes, especially on mainstream social media.' If philosemitism functions as an early stage in the deterioration of apparent admiration into hate, and if the 'Imaginary Jew' tropes that animate it continue to evolve, then monitoring who spreads this content, rather than only what they say, may be the more structurally meaningful detection strategy. Ultimately, neither textual nor behavioural models can function without a principled dataset built for specific contours of philosemitism which is what the creation of an annotation guideline provides.

A Consideration of Platform Architecture

a. Moderation Paradox

Yet even the most sophisticated detection tools operate within platform architectures that shape what content is visible and to whom. The distribution of philosemitic is shaped by platform architecture, regulation and moderation practices. As the Trust and Safety Expert explains, platforms are shifting from binary removal towards 'softer' interventions, such as 'limiting discoverability', 'IP blocks' or 'restricting distribution'. Content may remain online but becomes less visible. These decisions vary by jurisdiction. European frameworks impose stricter thresholds, while American approaches exist 'without the weight of history' and are less restrictive. When platforms disagree with local law, 'they'll go with an IP block' meaning content can be accessible in one country and invisible in another.

This creates a moderation paradox relevant to philosemitism. Content that falls into grey areas, such as philosemitism, is unlikely to meet removal thresholds, circulating beneath detection. As explicit hate is more heavily moderated, users migrate towards its positive-register equivalent (Farrand, 2023). As the Online Antisemitism Professor observed, philosemitism appears ‘more on TikTok and places like this. I haven’t seen it on less regulated platforms’. This suggests that moderation displaces and transforms rather than eliminates such discourse.

This has implications for research. Since visibility is shaped by geography and platform design, data collection is inherently uneven. Any study of philosemitism must account for these structural differences by specifying platform, region and collection contexts to ensure reproducibility.

b. Automated Detection versus Moderation

This category emphasises that philosemitism should be monitored rather than moderated as it avoids the tension between prioritising free speech or minimising hate speech. Interviewees consistently how the phenomenon's ambiguity does not currently warrant removal from mainstream platforms. Platforms are already reluctant to act. They ‘don’t casually remove content, you have to be really sure’, said the Safety Expert, and philosemitism’s grey area means this threshold is rarely met. The term has not entered Trust and Safety circles. The Expert notes, Jewish groups ‘never had this topic as one of their main concerns’ at TikTok or Meta, focusing more on Holocaust denial and conspiracy theories. The ‘large gap between research companies and real world applications’ of automated detection means that even where the will exists, the infrastructure does not, explains the Quantitative Researcher. Monitoring is therefore the appropriate response as evidential justification for removal does not yet exist. Should future research establish philosemitism as a measurable harm, this position may need reconsideration.

Conclusion

Category Four shows that the detection gap is fundamentally conceptual as well as technical. By operationalising philosemitism, this study provides an evidence-based foundation for practitioners. Future research should benchmark LLM performance and test these categories across multilingual and cross-cultural contexts to validate its robustness.

Creating an Annotation Framework

Category Four brings to the forefront the reasons behind philosemitism’s current structural detection failure. The production of an annotation guideline begins to remedy these. The following maps the decisions behind the choices made to create the framework.⁸

⁸ Find annotation guideline in Appendix A.

The guideline consists of seven independent annotation dimensions with a corpus collection protocol, applied per item, with conservative coding as the governing rule and the Decoding Antisemitism framework as the disambiguating instrument (Becker and Troschke, 2023; Becker, Blatter and Stavenich, 2026). Three real items from the empirical dataset are annotated in full as worked examples: a Kelly Willner Instagram comment under a Free Press post, a Nigel Farage Tweet following the 2026 Trafalgar Square open Iftar and *k0227529*'s Instagram comment under a Jewish influencer's reel. These three were chosen to span the schema. The first example is paradigmatic essentialising at the explicit tier and the second is implicit instrumentalisation tier with co-text confirmation. The final example is an overall-labelling firewall case as, under the rules of conservative coding, is classified as Neutral.

Seven Sections

The schema has seven separate sections because the four findings' categories were not mutually independent. Categories One and Two are content categories, classifying what posts say about Jews. Category Three is an evaluative category, classifying posts on the four authenticity markers. Category Four is a structural category, classifying items against the linguistic and discursive features that produce the detection gap. A single post needs labels from all four. The seven sections are therefore operationalise this multi-dimensional structure. Sections 1-3 handle content, Section 5 focuses on evaluation, Sections 4 and 6 involve metadata and Section 7 labels structural features.

An alternative option would be to collapse all four categories into a single multi-label classifier but this would lose the diagnostic resolution that helps downstream analysis and model training. The seven-section structure preserves the finding's distinctions but is tractable for annotator workings at an item level.

Justifications

Becker's *Decoding Antisemitism* project has produced the most empirically grounded annotation framework for socio-cultural content concerning Jews (2024). The four-tier framework was developed for antisemitism but transfers to philosemitism without distortion because the underlying problem (the relationship between surface meaning and discursive effect) is structurally parallel. Adopting it as a name dimension in Section four preserves cross-comparability with *Decoding Antisemitism* datasets for future joint analysis. It also makes visible the Category Three threshold-instability finding at the end of every annotation which gives annotators more understanding of why labelling an item is challenging.

Social media platforms decontextualise posts and strip any longitudinal evidence required to assess genuine support. This is why the guideline has two axes for authenticity. It combines the authenticity test markers (consistency, inconvenience, action-orientation, selflessness) with engagement-depth scale.⁹ Only using the authenticity test criteria would lose the

⁹ Selflessness is not included as it cannot be observed at post level.

discriminating capacity of the markers. For example, a high-inconvenience but low-action item is structurally different from a low-inconvenience high-action item.

To operationalise Category Four, the guide records detection-gap features as flags applicable to every item, regardless of primary classification. A Neutral item that uses the noticing syntax is useful for downstream model training so the five structural flags in Section 7 provide a feature-training signal for fine-tuning detection models on the surfaces that philosemitic content evades sentiment-based classifiers.

The schema also recommends Krippendorff's α with raw agreement scores (Hayes & Krippendorff, 2007). The first adjusts for chance, while the second stays reliable when data is uneven. Agreement should be high for clear cases, lower for implicit or context-dependent ones, and lowest for grey-area items. Low alpha scores in these grey areas are expected and should not be seen as annotator error. Instead, they reflect the real difficulty of defining the category and show threshold instability.

Finally, Charmaz's emphasis on the engagement between the data and researcher's interpretive frame is grounded in the guideline (2014). As per CGT, the findings' categories are constructed, not discovered, and therefore remain provisional. The dataset should keep all annotator labels and not just the final agreed label because the disagreement in grey-area cases is expected and shows the issue itself.

Overall, the seven sections, their justifications and the governing principles of conservative coding form a guideline designed to help make positive-valenced prejudice more visible, bridging the gap between the qualitative findings and future quantitative research.

Limitations and Areas for Future Research

Without a standardised definition, framework or dataset on online philosemitism, key gaps remain. Despite efforts to follow Charmaz's CGT and address researcher positionality, this thesis has limitations that future research can build on. The following limitations acknowledges this and seeks to help future researchers in developing the exploration of online philosemitism.

Data Limitations

The social media data has several structural constraints. The BSA dataset was limited to one Boolean query category due to time constraints, meaning the guilt-engendered sub-type is represented more thoroughly than others in the sample. The remaining social media data was collected purposively to saturated theoretical categories by following leads from interviewees rather than through systematic sampling. This makes the dataset illustrative rather than representative which is appropriate for an exploratory study but a constraint for any future claims to generalisability.

TikTok is absent from the BSA tool and from the broader data collection. This is a notable gap given that the Online Antisemitism Professor specifically identified it as a platform where philosemitism appears with particular frequency. Thus, future research should prioritise TikTok data collection.

The Facebook Forum, while rich in depth, is a closed and mediated environment. Members self-select and must declare their religious identity. This produces a more reflective and articulate sample than public social media and does not capture the unreflective, ambient philosemitism that circulates without awareness on mainstream platforms.

The study is limited to English-language content. The cross-cultural findings, particularly those relating to India, Hindu nationalism and the model minority mechanism, relied on an English-language interviewee and English-language posts. The full scope of non-Western philosemitism operating in Hindi, Japanese, Korean or Polish remains beyond the study's reach. This is a meaningful gap given that the thesis explicitly argues the phenomenon is international in formation.

It would also clarify the importance of cultural context. For example, guilt-led philosemitism differs between the Netherlands and Germany, reflecting divergent national histories and their relationship to their relative Jewish communities (Cohen, 2025). Further study of language variation within the same platform would refine conceptualisation of the phenomenon's existence online.

Methodological Limitations

This is an exploratory qualitative study and makes no claims to statistical significance or generalisability. The twelve interviews, while diverse in expertise, represent a relatively small Euro-centric sample. Most participants had significant UK exposure, which may have inflected the findings towards an Anglophone and European understanding of the phenomenon. The model minority and Hindu nationalist categories relied on a single participant. They are included because they are theoretically significant and an empirically novel contribution, particularly the Indian context, which does not appear in the existing English-language literature. However, further research with a larger and more geographically diverse interview sample would be needed to consolidate them.

The annotation guideline has not yet been tested for inter-coder reliability. It is a first-draft instrument designed to make future labelled datasets methodologically possible and not a validated tool. Its categories will require iterative refinement through pilot annotation and inter-coder testing before they can credibly support a gold standard dataset. The importance of this step was referenced by two interviewees working with annotation methods for studying online antisemitism. For sensitive subjects such as philosemitism, inconsistent annotation risks reproducing biases and therefore necessitates rigorous inter-coder agreement.

H3 Scope

The third hypothesis proposed that the positive linguistic features of philosemitic discourse produce a systemic detection gap in existing content moderation frameworks. This was partly confirmed. The qualitative evidence portrayed the structural and practical conditions under which misclassification is likely. The hypothesis breached the scope of this research which is its most significant downfall. However, it poses an exciting prospect for future research as it can be tested more sufficiently through a benchmarking experiment using the annotation guideline to produce a labelled dataset, test against existing models and provide empirical confirmation of what this study's qualitative phase could only anticipate.

Positionality

CGT requires explicit acknowledgement of researcher positionality, though not as a confession to be apologised for (Charmaz, 2014). As a Liberal Jewish woman from the UK, this shaped the research in ways that were largely a strength, with one significant constraint.

The constraint was the Israeli-Palestinian conflict. Interviewees raised it freely and frequently. It was contextually unavoidable. A deliberate decision was made to exclude this dimension from the analysis. This was not a methodological oversight but a considered judgement rooted in positionality. The conflict is one of the most contested political questions of the current moment. Regardless of how carefully a neutral analytical stance is constructed, I opted for a cautious approach to its inclusion. Excluding it allowed the thesis to explore avenues of philosemitism that receive far less scholarly attention, such as the model minority mechanism, Christian supersessionism and the detection gap. In future research, conducted with inter-coder support or co-authored across different acknowledged positionalities, this dimension could be incorporated in a more rigorous and accountable way which would certainly deepen the contextualisation of the findings.

Overall, being a British Liberal Jew was more an asset than a limitation. It provided a working familiarity with the cultural codes and historical references that underpin philosemitic discourse and the capacity to identify nuance in how Jewish identity is being discussed that a researcher without that background might not detect. This is consistent with CGT's treatment of the researcher's perspective as a resource to control rather than erase (Charmaz, 2014, p.31). The UK liberal Jewish experience is nevertheless one specific formation within a diverse Jewish diaspora. It differs from the Orthodox, Sephardic, Israeli and other diaspora experiences discussed by other participants in this study. This was kept in mind throughout analysis and is reflected in the deliberate diversity of the interview sample.

Conclusion

This thesis set out to answer a question the existing literature had not yet posed in operational terms: how can online philosemitism be conceptualised and operationalised in order to detect and analyse its presence in contemporary online communication? The study offers the groundwork to make it possible for further exploration.

All three hypotheses were confirmed. Online philosemitism does indeed take typologically distinct forms that existing offline literature has not recorded. Existing scholarship does not document, for example, the Hindu nationalist formations, the Brahmin and Dalit identification mechanisms, the cross-platform circulation of guilt-engendered content after antisemitic attacks. The second hypothesis was also confirmed as the majority of philosemitic discourse encountered reproduced essentialised stereotypes of Jewish identity regardless of intent. The 'Imaginary Jew' was constructed repeatedly through different lenses, from intellectual and economic admiration to resilience narratives and comparative identification. The sincerity of intent did not prevent structural harm and so this constructed Jewish identity was the consistent structural feature across all four sub-types and all four instrumentalising formations. H3 was partly confirmed as, from interviewees insights, the positive linguistic surface of philosemitism engenders a systemic detection gap. Whilst this study could not empirically demonstrate that existing monitoring systems categorically treat positive sentiment as inherently safe, interviewees consistently identified structural and practical challenges that make misclassification likely. Philosemitism, therefore, has the capacity to evade detection by operation within these current constraints, effectively hiding in plain sight. What was central to philosemitism is its conceptual instability. Just like antisemitism, it resists stable classification but for different reasons. It is because philosemitism carries an element of deniability that is challenging to classify. Any detection framework that tries to eliminate this instability rather than systematise it will produce a dataset that is clean but inaccurate. The annotation guideline produced here attempts to do the latter through building in interpretive flexibility, conservative coding and grey area categories rather than imposing false precision on a phenomenon defined by grey area.

The annotation guideline is the first instrument of its kind for online philosemitism, developed in consultation with Becker and Blatter's annotation methodology and applied here to a new phenomenon. It provides future researchers with a principled starting point for producing labelled datasets and, hopefully, conducting LLM benchmarking experiments.

The gap this thesis addresses is real and consequential. There is no existing labelled dataset for online philosemitism, no validated detection framework and no scholarly consensus on where its boundaries lie. This study does not resolve all of that. It establishes that the phenomenon is empirically present, structurally significant and operationalisable. The next steps for research are lucid. They include testing the inter-coder reliability of the annotation guideline, LLM benchmarking using a corpus built from it and multilingual studies to gather whether the categories identified here hold across different cultural and linguistic contexts, beginning with the non-Western formations this thesis has opened up.

Philosemitism is not peripheral to how the Jewish community is discussed online and hiding it in the shadows of antisemitic research ignores its structural significance. The two phenomena share the same essentialising logic, feed the same 'Imaginary Jew' and will continue to evolve in tandem. Any serious effort to monitor one must eventually account for the other. This thesis is the first step in that direction.

Policy Recommendations

Philosemitism is currently invisible to the institutions best placed to address it. The recommendations below are directed at four groups: civil society organisations, research funding bodies, policymakers and mainstream social media platforms. Each is grounded in the study's findings.

Educators and Civil Society Organisations

Holocaust education dominates non-Jewish understanding of Jewish identity (Woo et al., 2025). It remains essential but it is important to teach about Jewish identity more broadly as it increases affective warmth toward Jewish people and supports constructive minority-majority cooperation (Lipka, 2019; Vesper et al., 2010). The following recommendations build on the UK's Curriculum and Assessment Review (2025) recommends Citizenship and Religious Education reform to better tackle prejudice. Although UK-focused, these recommendations have broader applicability as internationally, there is gap in media literacy on minority narratives. This study recommends to:

1. Integrate philosemitism as a complementary analytic category within existing antisemitism education programmes, providing learners with the conceptual language needed to critically evaluate the issue.
2. Encourage civil society bodies, such as the Combat Antisemitism Movement, to incorporate philosemitism into practitioner training programmes they already deliver. For example, this could extend the Community Security Trust's existing work training BBC journalists on mis- and disinformation (HM Government, 2025).
3. Add materials on positive-valenced essentialisation to the Department for Education's *Educate Against Hate* website (and international equivalents), drawing on the sub-types identified in this thesis to support teachers in addressing more subtle, acceptable forms of bias.

Research Funding Bodies

Funding should be prioritised to produce a labelled data, enabling improved detection, measurement and analysis of positive-valenced essentialisation in computational models. Cross-national funders, such as the European Commission's Horizon Europe programme, alongside national bodies like National Science Foundation (USA), should be leveraged to support this work. They enable comparative, multicultural research which is important as philosemitism appears differently across linguistic and cultural contexts.

Partnerships should be actively developed to accelerate the research. The Anti-Defamation League's Centre for Antisemitism Research should incorporate philosemitism into future benchmarking to assess current models' detection abilities (ADL, 2026). Alongside the BSA's existing data infrastructure and in collaboration with expert annotators, the development of a robust labelled dataset and associated research outputs is well within reach.

Policymakers

The IHRA working definition of antisemitism is the primary instrument used by governments, universities and platforms to identify anti-Jewish hatred. Its twelve illustrative examples focus exclusively on hostile content. Becker et al.'s *Lexicon* (2024) is also the most expansive dictionary of antisemitism definitions but does not include philosemitism. Therefore,

1. The EU Coordinator on Combating Antisemitism, the Decoding Antisemitism project and relevant civil society organisations should jointly commission a working definition of philosemitism to raise awareness.
2. The IHRA definition's operational guidance (not the definition itself) should be expanded to include worked examples of positive-valenced prejudice.

Social Media Platforms

Philosemitism should be monitored, not removed. Fringe groups already use humour, coded language and positive framing to normalise extreme views on mainstream platforms (Woods & Ruscher, 2021). The appropriate governance model is systemic risk monitoring,

1. Under Article 34 of the EU Digital Services Act (2022), Very Large Online Platforms are required to conduct systemic risk assessments of content posing risks to fundamental rights, including human dignity. Platforms must demonstrate, within their DSA transparency reports, that these assessments consider positive-valenced stereotyping alongside explicit hate speech.
2. Platforms should audit whether softer interventions, such as reduced discoverability, comment folding, IP blocks, are inadvertently pushing users towards philosemitic framings as a workaround for moderated overt antisemitism.

Bibliography

Figure Citations

- Aija (2023). [online] Instagram. Available at:
<https://www.instagram.com/reels/Cw-96OiSCFb/>.
- Casarthakuhuja, S. (2026). *Fargin*. [online] LinkedIn. Available at:
https://www.linkedin.com/posts/casarthakahuja_casarthakahuja-share-7435721209394286592-zVsr?utm_source=share&utm_medium=member_desktop&rcm=ACoAACv7qHEBoGtSdn4DV8H0a71C-34qgF1ZjhA.
- Casarthakuhuja (2026). [online] Instagram. Available at:
<https://www.instagram.com/reel/DU1bHtlj3Pv>.
- Dalit is-Hebrew (n.d.). [online] Facebook. Available at:
<https://www.facebook.com/dalit.is.hebrew/#>.
- dalitishebrew (n.d.). [online] Instagram. Available at:
<https://www.instagram.com/dalitishebrew..>
- Farage, N. (2026). [online] X. Available at:
https://x.com/Nigel_Farage/status/2034569516399648879.
- Freeman, B. (2025). *I'm Leaving the UK*. [online] instagram. Available at:
<https://www.instagram.com/p/DPWtH4fk20Q>.
- Gopalak (2025). [online] X. Available at:
<https://x.com/Gopalak78325085/status/2003653896775078327>.
- Henmazzig. (2025). *A Night of Terror in Australia*. [online] Instagram.
<https://www.instagram.com/p/DLudk8dN4zT/>
- Missmayim. (2025). [online] Instagram. Available at:
<https://www.instagram.com/p/DSQJw7FjwTq/>
- Missmayin. (2025). We are all Bondi. [online] Instagram. Available at:
<https://www.instagram.com/p/DSQvNDJEhf5>.
- Tucker, M. (2025). [online] Instagram. Available at:
<https://www.instagram.com/p/DQ18Yr2iejR>.

References

- Arora. (2026) *There's Nothing Sinister about Muslim Prayers in Trafalgar Square*. Guardian. Available at:

<https://www.theguardian.com/commentisfree/2026/mar/20/muslim-prayers-trafalgar-square-nick-timothy>

ADL. (2025) *Coded Hate: Extremists Weaponize Seemingly Innocuous Content to Promote Bigotry*. [ADL]. Available at: <https://www.adl.org/resources/article/coded-hate-extremists-weaponize-seemingly-innocuous-content-promote-bigotry>

ADL. (2026) *Six Leading AI Models Show Varied Ability to Detect and Counter Antisemitism and Extremism*. New ADL AI Index Finds. Available at: <https://www.adl.org/resources/press-release/six-leading-ai-models-show-varied-ability-detect-and-counter-antisemitism>

The Antisemitism Trust. (2024) *Online Antisemitism and the Risks of Artificial Intelligence*. Available at: <https://antisemitism.org.uk/wp-content/uploads/2024/02/7112-APT-Ai-and-Anitsemittism-v4.pdf>.

Badamchi, K. (2021) *Dimensions of Free Speech*. Berlin: Springer International Publishing.

Bauman, Z. (1998) 'On glocalization: Or globalization for some, localization for some others', *Thesis Eleven*, 54(1), pp. 37–49.

Batailler, C., *et al.* (2022) 'A Signal Detection Approach to Understanding the Identification of Fake News', *Perspectives on Psychological Science*, 17(1), pp. 78–98.

Becker, M. J., & Allington, D. (2021) *Second Discourse Report: Decoding Antisemitism – An AI-driven Study on Hate Speech and Imagery Online*. Available at: <https://decoding-antisemitism.eu/publications/second-discourse-report/>.

Becker, M. J., & Troschke, H. (2022) 'How Users of British Media Websites Make a Bogeyman of George Soros', *Journal of Contemporary Antisemitism*, 5(1), pp. 49–68.

Becker, M. J., & Troschke, H. (2023) 'Decoding implicit hate speech: The example of antisemitism' in C. Strippel, S. Paasch-Colberg, M. Emmer, & J. Trebbe (eds), *Challenges and perspectives of hate speech research*. Berlin: GESIS – Leibniz-Institut für Sozialwissenschaften, pp. 335–352.

Becker, M.J., Ascone, L., Bolton, M., Bundzíkova, V., Chapelan, A., Haupeltshofer, P., Krugel, A., Kurjan, I., Mihaljević, H., Munnes, S. and Placzynta, K., (2024) *Sixth Discourse Report: Decoding Antisemitism – An AI-driven Study on Hate Speech and Imagery Online*. Report 6. Available at: <https://decoding-antisemitism.eu/publications/sixth-discourse-report/>.

Becker, M. J., Blatter, J. & Stanevich, O. (2026) 'Decoding antisemitism online: linguistic and multimodal challenges in the age of AI', *Frontiers in Communication*, 10, 1729279.

Berthelot, K. (2020) 'Power and piety: Roman and Jewish perspectives', in K. Berthelot (ed) *Reconsidering Roman Power*. Rome: Publications de l'École française de Rome, pp. 269–289.

Bergmann, E. (2020) *Populism and the Politics of Misinformation*. The Journal of South African and American Studies. 21(3). pp. 251-65.

Bilewicz, M. & Soral, W. (2020) 'Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization', *Political Psychology*, 41, pp. 3–33.

Bjola, C., Manor, I. (2020) *Combating Online Hate Speech & AS*, DigDiploROx. Oxford Digital Diplomacy Research Group Working Paper, 4.

Blue Square Alliance. (2025a) *Coded, Conspiratorial, and Accelerated: The State of Antisemitism Online. 2025 Year in Review*. Available at: https://www.bluesquarealliance.org/wp-content/uploads/2026/02/BSA-2025_YearInReview.pdf

Blue Square Alliance. (2025b) *What Research Tells Us About What Works In The Fight Against Antisemitism*. Available at: <https://www.bluesquarealliance.org/research/fighting-antisemitism-with-impact/>.

Boromisza-Habashi, D. (2012) *Speaking hatefully: Culture, communication, and political action in Hungary*. University Park, PA: Penn State University Press.

Bossetta, M., (2020) *Scandalous design: How social media platforms' responses to scandal impacts campaigns and elections*. Social Media+ Society, 6(2).

Bossetta, M. (2022) *Antisemitism on social media platforms: Placing the problem into perspective*. In *Antisemitism on social media*. Routledge. pp.. 227-242.

Brewer, M. B. (1999) 'The Psychology of Prejudice: Ingroup Love or Outgroup Hate?', *Journal of Social Issues*, 55(3), pp. 429–444.

Brink, E. et al. (2006) *Book Review*. International Journal of Qualitative Studies on Health and Well-being, 1(3). Pp.188-192.

Byrne, P. (2017) *Why were the Jews Expelled from England in 1290?* Oxford: University of Oxford.

Chapelan, A. (2024) 'Conspiracy Theories' in Becker, M.J., Troschke, H., Bolton, M., & Chapelan, A. (eds) *Decoding Antisemitism. Postdisciplinary Studies in Discourse*. London: Palgrave Macmillan, pp. 175–190.

Chared et al. (2025) *AI-driven detection of hate speech on social media: a case study in the French language*. Cluster Comput (28,) 811.

Charmaz, K. (2014) *Constructing grounded theory*. 2nd Ed, Thousand Oaks, CA: Sage.

- Charmaz, K. (2017) *Constructivist grounded theory*. The journal of positive psychology, 12(3), pp.299-300.
- Casarthakahuja. (2026). *Why do so many countries hate Indians?* [Instagram]. Available at: <https://www.instagram.com/reels/DUa4311jxNE/>
- Casarthakahuja. (2026) *Fargin*. [online] Instagram. Available at: <https://www.instagram.com/reel/DVjI9jCDAnz/> [Accessed 28 Apr. 2026].
- Chandra, M. et al. (2021) “*Subverting the Jewtocracy*”: *Online antisemitism detection using multimodal deep learning*. Proceedings of the 13th ACM Web Science Conference 2021, pp. 148-157.
- Chiu K.-L., Collins, A., & Alexander, R. (2022) ‘*Detecting Hate Speech with GPT-3*’ [Version 4 March 2022] Available at: <https://doi.org/10.48550/arXiv.2103.12407>.
- Clarke, A. E. (2005) *Situational Analysis: Grounded Theory After the Postmodern Turn*. Thousand Oaks, CA: Sage.
- Cohen, S.J., (1989) ‘Crossing the boundary and becoming a Jew’, *Harvard Theological Review*, 82(1), pp. 13–33.
- Cohen, G.D. (2025) *Good Jews: Philosemitism in Europe since the Holocaust*. Cambridge: Cambridge University Press.
- Corbin, J., Strauss, A., (1990) *Grounded theory research: Procedures, canons, and evaluative criteria*. Qualitative sociology, 13(1), pp.3-21.
- Cowan, G., & Hodge, C. (1996) ‘Judgments of hate speech: The effects of target group, publicness, and behavioral responses of the target’, *Journal of Applied Social Psychology*, 26(4), pp. 355–374.
- Davidson, T. (2026) *Multimodal large language models can make context-sensitive hate speech evaluations aligned with human judgement*. *Nat Hum Behav* (10), pp.514–530.
- Decoding Antisemitism* (2022) ‘Glossary’ [Version 30 March, 2022] Available at: <https://decoding-antisemitism.eu/wp-content/uploads/2022/03/DA-publications-Glossary.pdf?x89829>.
- Dekel, I. (2022) ‘Philosemitism in contemporary German media’, *Media, Culture & Society*, 44(4), pp. 746–763.
- Edelstein, A. (1982) *An unacknowledged harmony: Philo-Semitism and the survival of European Jewry*. Westport, CA: Greenwood Press.
- European Union Agency for Fundamental Rights (FRA). (2019) *Young Jewish Europeans: perceptions and experiences of antisemitism*. Available at: <https://fra.europa.eu/en/publication/2019/young-jewish-europeans-perceptions-and-experiences-antisemitism#publication-tab-1>.

Farrand, B. (2023) “‘Is This a Hate Speech?’ The Difficulty in Combating Radicalisation in Coded Communications on Social Media Platforms’, *European Journal of Criminal Policy Research*, 29, pp. 477–493.

Feldman L. H. (1993) *Jew and Gentile in the Ancient World*. Princeton, NJ: Princeton University Press.

Fischer, L. (2011) ‘Anti-“Philosemitism” and Anti-Antisemitism in Imperial Germany’, in J. Karp and A. Sutcliffe (eds) *Philosemitism in History*. Cambridge: Cambridge University Press, pp. 170–189.

Finkelstein, J. et al. (2018) *A quantitative approach to understanding online antisemitism*. *arXiv preprint arXiv:1809.01644*.

Finkelkraut, A. (1997) *The Imaginary Jew* (9). Lincoln: University of Nebraska Press.

Gans, E. E. (2014) ‘Anti-Antisemitic Enthusiasm & Selective Philosemitism: Geert Wilders, the PVV and the Jews’, *Proceedings / International conference “Antisemitism in Europe Today: The Phenomena, the Conflicts” 8–9 November 2013*. Berlin: The Jewish Museum.

Gans, E. E. (2017) ‘Disowning Responsibility: The Stereotype of the Passive Jew as a Legitimizing Factor in Dutch Remembrance of the Shoah’, in Wertheim, D. (ed) *The Jew as Legitimation*. London, Palgrave Macmillan, pp.173–193.

Girón, A., Huertas-Tato, J. & Camacho, D. (2025) *LLM synthetic generation to enhance online content moderation generalization in hate speech scenarios*. *Computing* (107)164.

Glaser, B., & Strauss, A. (1967) *The discovery of grounded theory: Strategies for qualitative research*. Hawthorne, NY: Aldine de Gruyter.

Goffman, E. (1974) *Frame Analysis: An Essay on the Organisation of Experience*. Cambridge, MA: Harvard University Press.

Gomez et al. (2019) *Exploring Hate Speech Detection in Multimodal Publication*. Barcelona: Unitat de Tecnologies Audiovisuals.

Green, D. M. & Swets, J. A. (1966) *Signal detection theory and psychophysics*. New York: Wiley.

Guhl, J. (2022) “‘Everyone I know Isn’t Antisemitic’: Antisemitism in Facebook Pages Supportive of the UK Labour Party’ in M. Hubscher & S. von Mering (eds), *Antisemitism on Social Media*. New York: Routledge, pp. 55–73.

Hale, R. (2025) ‘Misconceptions and knowledge gaps in Holocaust education: considerations and challenges in the context of teaching aims and contemporary issues’, *Education Inquiry*, pp. 1–14.

Hayes, A.F. and Krippendorff, K. (2007) Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1), pp.77-89.

Heeger, D. (1997) 'Signal Detection Theory', *Tutor Handout*, Department of Psychology, New York University. Unpublished.

Hein, G., et al. (2010). *Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping*. *Neuron*, 68(1), pp.149-160.

HM Government. (2015) *Educate Against Hate - Prevent Radicalisation & Extremism*. [online] Educate Against Hate. Available at: <https://www.educateagainsthate.com/>.

HM Government. (2025) *Antisemitism: Recent Government Actions and Next Steps*. [online] Available at: https://assets.publishing.service.gov.uk/media/6942d82f9273c48f554cf55f/Antisemitism_-_recent_government_actions_and_next_steps.pdf.

Hubscher, M. & von Mering, S. (2022) *Antisemitism on Social Media*. New York: Routledge.

Humprecht et al. (2023) *The sharing of disinformation in cross-national comparison: Analyzing patterns of resilience*. *Information, Communication & Society*, 26(7), pp.1342-1362.

Huang, F., Kwak, H., & An, J. (2023) 'Is ChatGPT better than Human Annotators? Potential and Limitations of ChatGPT in Explaining Implicit Hate Speech' *WWW '23 Companion: Companion Proceedings of the ACM Web Conference 2023*, pp. 294–297.

Inbari, M., Bumin, K. M. & Byrd, M. G. (2021) 'Why Do Evangelicals Support Israel?', *Politics and Religion*, 14(1), pp. 1–36.

Jahn, L., et al. (2023) 'Friction Interventions to Curb the Spread of Misinformation on Social Media' [Version 1 July 2023] Available at: <https://arxiv.org/abs/2307.11498>.

Jalbert, M. et al. (2020) Only Half of What I'll Tell You is True: Expecting to Encounter Falsehoods Reduces Illusory Truth. *Journal of applied research in memory and cognition*, 9(4), pp. 602–613.

Jewish English Lexicon (2025). *fargin - Jewish English Lexicon*. [online] Jewish-languages.org. Available at: <https://jel.jewish-languages.org/words/4165>.

Jikeli, G., et al. (2021) 'Detecting Anti-Jewish Messages on Social Media: Building an Annotated Corpus That Can Serve as A Preliminary Gold Standard.' *ICWSM Workshops*. Available at: https://workshop-proceedings.icwsml.org/pdf/2021_14.pdf.

Jikeli, G., et al. (2022) 'Toward an AI Definition of Antisemitism?', in M. Hubscher & S. von Mering (eds), *Antisemitism on Social Media*. New York: Routledge, pp. 193–212.

Jikeli, G., et al. (2024) 'Differences between antisemitic and non-antisemitic English language tweets', *Computational and Mathematical Organisation Theory*, 30, pp. 232–266.

- Jin, W. et al. (2025). *Veracity-oriented context-aware large Language models-based prompting optimization for fake news detection*. Int. J. Intell. Syst. (1), 5920142.
- Khan, K., et al., (2024). *Graph neural networks based framework to analyze social media platforms for malicious user detection*. Applied Soft Computing, (155), 111416.
- Kansteiner, W. (2011) ‘What Is the Opposite of Genocide?: Philosemitic Television in Germany, 1963–1995’, in A. Sutcliffe & J. Karp (eds) *Philosemitism in History*. Cambridge: Cambridge University Press, pp. 289–313.
- Karp, J. (2011) ‘Ethnic Role Models and Chosen Peoples: Philosemitism in African American Culture’ in A. Sutcliffe & J. Karp (eds) *Philosemitism in History*. Cambridge: Cambridge University Press, pp. 211–234.
- Kilcher, A. B. (2017) ‘The Theological Dialectics of Christian Hebraism and Kabbalah in Early Modernity’, in Wertheim, D. (ed) *The Jew as Legitimation*. London, Palgrave Macmillan, pp. 47–62.
- Koblentz-Stenzler, L., Klempner, U. (2024). *From Memes to Mainstream: How Far-Right Extremists Weaponize AI to Spread Anti-Semitism*. ICT.
- Koblentz-Stenzler, L. (2026) *Extremism & Antisemitism Desk Monthly Trend Report: January 2026; Far-Right Extremism*.
- Kowner, R., Ainslie, M. J. & Podoler, G. (2023) ‘When antisemitism and philosemitism go hand in hand: attitudes to Jews in contemporary East Asia’, *Patterns of Prejudice*, 57(3), pp. 175–201.
- Kruk, J. et al. (2024). *Silent Signals, Loud Impact: LLMs for Word-Sense Disambiguation of Coded Dog Whistles*. Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, 12493–12509.
- Lassner, P. & Trubowitz, L. (eds) (2008) *Antisemitism and Philosemitism in the Twentieth and Twenty-First Centuries: Representing Jews, Jewishness, and Modern Culture*. Delaware: Delaware University Press.
- Levenson, T. (2004). *Between Philosemitism and Antisemitism*. Lincoln, NE: University of Nebraska Press.
- Levy, R. S. (2017) ‘David J. Wertheim, Ed. The Jew as Legitimation: Jewish-Gentile Relations Beyond Antisemitism and Philosemitism’, *Studies in Christian-Jewish relations*, 12 (1), pp. 1–3.
- Levy, R. S. (2018) ‘The “Protocols of the Elders of Zion” in Court: The Berne Trial 1933-1937 and the “anti-semitic International”’, *Holocaust and genocide studies*, 32 (2), pp. 306–307.
- Levin, I., Filindra, A. & Kopstein, J. (2022) ‘Validating and Testing a Measure of Anti-semitism in Support for QAnon and Vote Intention for Trump in 2020’, *Social Science Quarterly*, 103(4), pp. 794–809.

Lipka, M. (2019) ‘U.S. Jews know a lot about religion – but other Americans know little about Judaism’, Pew Research Center, August 1. Available at: <https://www.pewresearch.org/short-reads/2019/08/01/u-s-jews-know-a-lot-about-religion-but-other-americans-know-little-about-judaism/>.

McDermott, G. R. (2016) ‘A New Christian Zionism’, *Providence Magazine*, Winter, pp. 56–62. Available at: <https://providencemag.com/wp-content/uploads/McDermott-New-Christian-Zionism1.pdf>.

Mendelsohn, et al. (2023). *From Dogwhistles to Bullhorns: Unveiling Coded Rhetoric with Language Models*. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, 15162–15180.

Moyn, S. (2005). *A Holocaust Controversy: The Treblinka Affair in Postwar France*. Waltham, MA: Brandeis University Press.

Munger, K., 2020. All the news that’s fit to click: The economics of clickbait media. *Political communication*, 37(3), pp.376-397.

Nagar, S., Barbhuiya, F & Dey, K. (2023). *Towards more robust hate speech detection: using social context and user data*. Soc. Netw. Anal. Min. (13), 47.

Norman Lear Center (2024) *Jews on Screen: The Visibility and Representation of Jewish Identity in Contemporary Scripted TV*. Available at: https://learcenter.s3.us-west-1.amazonaws.com/JewsOnScreen_NormanLearCenter2024.pdf.

Özyürek, E. (2023) ‘Situating empathy: Holocaust education for the Middle East/Muslim minority in Germany’, in F. Mezzenzana & D. Peluso (eds) *Conversations on Empathy*. New York: Routledge. pp. 174–193.

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A.A., Eckles, D. and Rand, D.G., (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855), pp.590-595

Polyakov O’Donnell, E. (2020) *The Nun in the Synagogue: Judeocentric Catholicism in Israel*. University Park, PA: Penn State University Press.

Prabhu, R., Seethalakshmi, V. (2025). *A comprehensive framework for multi-modal hate speech detection in social media using deep learning*. Sci Rep, (15) 13020.

Pustejovsky, J. & Stubbs, A. (2013) *Natural language annotation for machine learning*. Sebastopol, CA: O’Reilly.

Regulation (EU). (2022). 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act).

Rodriguez, A., Chen, Y. L. & Argueta, C. (2022). *FADOHS: framework for detection and integration of unstructured data of hate speech on Facebook using sentiment and emotion analysis*. *IEEE Access*. (10), 22400–22419.

Rose, H. (2020). 'The new philosemitism: Exploring a changing relationship between Jews and the far-right', ICSR. Available at: <https://icsr.info/wp-content/uploads/2020/11/ICSR-Report-The-New-Philosemitism-Exploring-a-Changing-Relationship-Between-Jews-and-the-Far-Right.pdf>.

Rubinstein, W. D. & Rubinstein, H. L. (1999) *Philosemitism: Admiration and Support in the English-Speaking World for Jews, 1840-1939*. New York: St. Martin's Press.

Schilling, C. L. (2018). 'Jewish Seoul: an analysis of philo-and antisemitism in South Korea', *Modern Judaism*, 38(2), pp. 183–197.

Schilling, C. L. (2020). 'On symbolic philosemitism in Japan', *Journal of Modern Jewish Studies*, 19(3), pp. 297–313.

Schreier, M. (2014). *Varianten qualitativer Inhaltsanalyse. Ein Wegweiser im Dickicht der Begrifflichkeiten*. *Forum Qualitative Sozialforschung*, 15(1).

Sebastian, K. (2019). *Distinguishing Between the types of Grounded Theory: Classical, Interpretive, and Constructivist*. *Journal for Social Thought*, 3(1).

Selent, M. & Kortmann, M. (2025) 'Philo-Semitic Civilisationism or Anti-Semitic Nationalism? The Ambivalent Stance of the Alternative for Germany Towards Judaism, Jews, and Israel', *German Politics*, 34(1), pp. 25–51.

Shapiro, F. (2012). 'Jesus for Jews: The Unique Problem of Messianic Judaism', *Journal of religion & society*, 14, pp. 1–16.

Smith, R. O. (2007) 'Jewish-Christian Difficulties in Challenging Christian Zionism', *Christian Zionism*, 7(5).

Speinwall, A. G. (2011) 'A Friend of the Jews? The Abbé Grégoire and Philosemitism in Revolutionary France', in A. Sutcliffe & J. Karp (eds) *Philosemitism in History*. Cambridge: Cambridge University Press, pp. 111–127.

Stangor, C. (2016). *The study of stereotyping, prejudice, and discrimination within social psychology: A quick history of theory and research*. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (2nd ed., pp. 3–27). Psychology Press.

Sutcliffe, A., Karp, J. (eds) (2011). *Philosemitism in History*. Cambridge: Cambridge University Press.

Taberner, S. (2005) 'Philo-Semitism in Recent German Film: Aimée und Jaguar, Rosenstraße and Das Wunder Von Bern', *German Life and Letters*, 58(3), pp. 357–372.

Udry J, Barber SJ. (2024). *The illusory truth effect: A review of how repetition increases belief in misinformation*. *Curr Opin Psychol*. (56),101736.

Vesper, C., Butterfill, S., Knoblich, G. and Sebanz, N., (2010.) *A minimal architecture for joint action*. *Neural networks*, 23(8-9), pp.998-1003.

Vusithembekwayo. (2026). *Fargin*. [Instagram]. Available at:
<https://www.instagram.com/reels/DVjI9jCDAnz/>

von Bieberstein, A. (2016) 'Not a German past to be reckoned with: Negotiating migrant subjectivities between "Vergangenheitsbewältigung" and the nationalization of history, *The Journal of the Royal Anthropological Institute*, 22(4), pp. 902–919.

Wang, W. C., *et al.* (2016) 'On Known Unknowns: Fluency and the Neural Mechanisms of Illusory Truth', *Journal of Cognitive Neuroscience*, 28(5), pp. 739–746.

Wang, Y.-S., & Chang, Y. (2022) 'Toxicity Detection with Generative Prompt-based Inference' [Version 1 May 2022] Available at:
<https://doi.org/10.48550/arXiv.2205.12390>.

Waseem, Z. & Hovy, D. (2016) 'Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter', *Proceedings of NAACL*, pp. 88–93.

Wedgewood, Y. (2022) 'Fighting Philo-semitism', *Vashti*, 1 September. Available at:
<https://www.vashtimedia.com/fighting-philosemitism/>.

Weimann, G. & Masri, N. (2022) 'New Antisemitism on TikTok', in M. Hubscher & S. von Mering (eds), *Antisemitism on Social Media*. New York: Routledge, pp. 167–180.

Weinberg D. B., *et al.* (2025) 'Hidden in plain sight: Antisemitic content in QAnon subreddits', *PLOS One*, 20(3), e0318988.

Wertheim, D. (ed) (2017a) *The Jew as Legitimation*. London, Palgrave Macmillan.

Wertheim, D. (2017b) 'Geert Wilders and the Nationalist-Populist Turn Toward the Jews in Europe' in Wertheim, D. (ed) *The Jew as Legitimation*. London, Palgrave Macmillan, pp. 275–289.

Wieviorka, M. (2018) 'A New Anti-Semitism? Jewish Political Studies Review', 29(3/4), pp. 43–46.

Woo, A., *et al.*, (2025) 'Deepening Instruction on the Holocaust and Topics Related to Jewish Life: Findings from Teacher Interviews', RAND. Available at:
https://www.rand.org/pubs/research_reports/RRA3615-2.html.

Woods, F., Ruscher, J. (2021). *Viral sticks, virtual stones: Addressing anonymous hate speech online*. *Patterns of Prejudice*, 55(3), pp.265-289.

Zubrzycki, G. (2022) *Resurrecting the Jew. Nationalism, Philosemitism and Poland's Jewish Revival*, Princeton: Princeton University Press.

Appendices

Appendix A: Annotation Guideline

Annotation Guideline: Online Philosemitism

Produced as part of Gabriella Cohen, Hiding in Plain Sight: The Conceptualisation and Operationalisation of Online Philosemitism (2026).

1. Purpose

As an expert annotator, you will evaluate how online philosemitism appears online by labelling social media data referring to Judaism, Jewish people, identities, ideas, rituals and practices. A well-labelled item produces metadata that supports future corpus construction, automated detection research and inter-coder reliability testing.

This document provides a structured annotation framework for identifying and classifying instances of online philosemitism. Each labelling criterion is grounded in the findings of the present study and in established practice from prior antisemitism-annotation work, principally the Decoding Antisemitism project (Becker and Troschke, 2023; Becker, Blatter and Stanevich, 2026) and the IHRA Working Definition (2016). Some criteria are binary, others ordinal: please apply each as specified.

2. The Concept of Online Philosemitism

Online philosemitism is a multi-dimensional phenomenon comprising expressions of admiration, support and positive attention toward Jewish people, Judaism or Jewish identity that vary in sincerity, depth and discursive function. It encompasses both genuine and performative forms, as well as instrumental uses of Jewish identity to advance non-Jewish agendas. The phenomenon is treated here as graded rather than binary: an item is not simply philosemitic or not, but exhibits philosemitic features at varying degrees of legibility and structural depth.

3. Unit of Analysis

Primary unit: the individual post, comment or message.

Co-text: annotators consult immediately visible co-text (parent post, quoted post, directly preceding comment in thread, attached media and visible hashtag set) only where required to disambiguate the item. Where co-text is required for classification, this is recorded under Section 4 (legibility tier: co-text).

Default: annotators rely only on the content provided unless co-text is necessary for classification.

4. Annotation Structure

Each item is annotated across seven independent sections. Annotators complete each section in order and assign labels per the criteria specified below:

- Section 1.** Primary Classification.
- Section 2.** Philosemitic Subtype (target framing).
- Section 3.** Instrumentalising Form.
- Section 4.** Confidence (legibility tier and confidence score).
- Section 5.** Authenticity and Engagement Signals.
- Section 6.** Context Metadata.
- Section 7.** Detection Signals.

5. Conservative Coding Protocol

Conservative coding governs every section. Two rules apply:

1. **Philosemitic vs Neutral:** Where content could plausibly be interpreted as philosemitic or neutral, annotators default to neutral. The candidate philosemitic reading is logged in Additional comments.
2. **Philosemitic vs Antisemitic:** Where content could plausibly be interpreted as antisemitic or philosemitic, annotators record both labels under Section 1 (Both Philosemitic and Antisemitic), with confidence score and a brief justification.

This dual rule minimises false positives against genuine solidarity, preserves dual-valenced cases for downstream analysis, and maintains corpus integrity for fine-tuning detection models.

6. Corpus Collection Protocol

To ensure reproducibility across future cross-platform or cross-national studies, the following must be recorded at the point of data collection:

- Platform and version or interface used.
- Researcher's geographic location and VPN status.
- Collection date and time.
- Search terms or hashtags used.
- Whether content was accessible without an authenticated account.
- Language of the item (and language of the surrounding thread, where different).

7. Reliability and Iteration

- Conduct pilot annotations with multiple coders before full-scale labelling.
- Measure inter-coder agreement using metrics appropriate to subjective tasks (Krippendorff's α with per-class reporting recommended; raw observed agreement reported alongside).
- Refine definitions based on disagreement patterns. Disagreement at the grey-area legibility tier is expected and reflects the phenomenon, not annotator failure.

Section 1. Primary Classification

Name	primary_classification
-------------	------------------------

Definition	Antisemitic. Content meeting the IHRA Working Definition of antisemitism (2016) and/or falling within Decoding Antisemitism Lexicon categories (Becker et al., 2024). Philosemitic. Content expressing admiration, praise or positive essentialisation of Jewish people, identity or culture in ways that instrumentalise or 'other' the Jewish community, even where the surface register is positive. Neutral. Content referencing Jewish people, identity or culture without essentialising or instrumentalising framing. This includes factual reporting, basic anti-antisemitism, Jewish self-description and educational content. Both Philosemitic and Antisemitic. Content that combines positive admiration with antisemitic logic. For example, praise of Jewish wealth attached to an explicitly antisemitic image. Not Relevant. Content does not concern Jewish people, Judaism, Jewish identity, antisemitism or related discourse. Insufficient Context. Item may be relevant but cannot be classified with the available information.
Inclusion criteria	Item carries an explicit or implicit stance (supportive, critical, hostile, fascinated, performative) toward Jewish people, Judaism or Jewish identity. The stance need not be the post's primary topic, but must be discernible from the content (and, where relevant, the consulted co-text).
Exclusion criteria	Content unrelated to Jewish people, Judaism or Jewish identity. Content that mentions Israel without generalising to Jews or invoking Jewish identity falls under Not Relevant for this dimension; the Israel/Jewishness rule is set out under Edge-case rules below.
Justification	Operationalises Findings Category 1 (essentialising mechanism) and Category 2 (instrumentalising forms). Anchored in the IHRA Working Definition and the Decoding Antisemitism Lexicon.

Edge-case rules for Section 1

- **Specific praise is not group essence.** Praise of a named person, institution, text or historical contribution is not philosemitic unless it generalises that quality to Jews as a collective essence.
- **Quoting is not endorsing.** A post quoting philosemitic or antisemitic content for critique is not labelled as endorsing that content unless the poster's stance is approving or ambiguous.
- **Israel content is in scope only when Jewish identity is implicated.** Content about Israeli government policy is not philosemitic or antisemitic merely because it mentions Israel. Apply labels when the post generalises from Israel to Jews, uses Jews as a collective symbol, or frames Israel as proof of Jewish essence.
- **Holocaust memory is not automatically philosemitic.** Holocaust commemoration, education and mourning are neutral by default. Label as philosemitic when Holocaust

memory is used to reduce Jewish identity to suffering, redeem the speaker's group or legitimise unrelated agendas.

- **'Judeo-Christian values' requires context.** The phrase is not automatically philosemitic. It becomes labellable when Jewishness is used as a legitimising attachment to Christian, nationalist, anti-Muslim or civilisational claims without Jewish agency.
- **Jewish self-description requires caution.** Jewish speakers may use collective language, intra-communal humour or religious categories differently from outsiders. Self-description is not classified as philosemitic by default. Mark speaker identity in Section 6 only where self-declared or provided in the dataset metadata.

Section 2. Philosemitic Subtype (Target Framing)

Name	philosemitic_subtype
Definition	Where the primary label is Philosemitic (or Both), annotators record (a) the essentialising subtype and (b) the target-framing of Jewish identity in the item. Multiple subtype labels may apply; one target-framing label is selected as primary. Essentialising subtype: Intellectual/Cultural. Jews represented as collectively or uniquely intelligent, scholarly, creative, cultured or civilisationally advanced (e.g. 'people of the book' framings; references to inherent Jewish intellect). Economic. Jews represented as unusually wealthy, financially skilled, commercially networked or powerful, including aspirational fetishisations of Jewish economic success (e.g. Fargin-as-secret framings, '5 Jars Theory' content). Resilience-based. Jewish identity reduced to survival, suffering, trauma, Holocaust memory or heroic endurance. Moral/Spiritual. Jews represented as uniquely ethical, divinely chosen, spiritually pure or morally exemplary. Model Minority. Jews abstracted as an exemplary minority to validate or instruct another non-Jewish group's aspirations or grievances. Cultural/Religious Fetishisation. Jewish symbols, rituals, texts or practices treated as aesthetic, exotic, mystical or spiritually useful objects detached from Jewish agency. Other. Specify in Additional comments. Use only where no existing subtype applies. Target framing: Agents. Jewish people represented as complex actors with agency, internal diversity and situated perspectives. Symbols. Jewish identity used as moral shorthand, emblem or abstract reference.

	Instruments. Jewish identity used strategically to support a claim that is not about Jews.
Inclusion criteria	Apply where (i) the primary classification is Philosemitic or Both, and (ii) the item attributes positively-valenced traits to Jews collectively, or treats Jewishness as a symbolic or instrumental object.
Exclusion criteria	Do not apply to specific praise of named individuals or institutions that does not generalise to Jews as a group; to academic or historical scholarship; to in-group Jewish self-description; to reactive condemnation of antisemitism that does not impute group-level traits.
Justification	Operationalises Category 1 (essentialising mechanism) and the four sub-types of essentialised praise developed in Chapter Two; the target-framing rows operationalise the agent/symbol/instrument distinction articulated in the Imaginary Jew analysis (Finkelkraut, 1997; Wieviorka, 2018).

Section 3. Instrumentalising Form

Name	instrumentalising_form
Definition	Annotators record whether the item uses Jewish identity as a vehicle for an ideological agenda whose primary object is not Jewish life, Jewish safety or Jewish self-understanding. One value selected; the agenda is specified where present. If Yes, specify agenda. One or more of the following: Far-right/Anti-Muslim Proxy. Pro-Jewish or pro-Israel claims deployed as the rhetorical complement of anti-Muslim, anti-immigrant or far-right populist content. Christian Supersessionist/Evangelical. Pro-Jewish claims embedded in a Christian theological frame in which Jewish existence is mediated by its function in Christian salvation history or 'Judeo-Christian' civilisational claims. Hindu Nationalist. Pro-Jewish or pro-Israel claims deployed to legitimate Hindu nationalist, caste, territorial or anti-Muslim narratives within Indian or diasporic discourse. Guilt/Redemption. Performative solidarity functioning as expiation for national, generational or institutional implication in the Shoah or other historical antisemitism. Political Co-option. Pro-Jewish claims deployed by politicians, parties or movements to legitimise unrelated agendas (tax policy, foreign policy).

	Influencer/Commercial. Jewishness deployed to sell content, courses, identity products or personal-brand authority (e.g. viral 'Jewish wealth secrets' content). Cultural Appropriation. Jewish symbols, trauma or texts used as cultural prestige, aesthetic identity or moral authority for non-Jewish projects. Noticing syntax. Praise framed as quantified suspicion (see Section 7). No. Item may be admiring or supportive but no external agenda is visible. Unclear. Possible instrumentalisation but insufficient evidence; flag for adjudication.
Inclusion criteria	Apply Yes where the item's discursive work (the function the pro-Jewish framing performs in context) is the advancement of a claim about something other than Jewish life. The instrumental function must be the centre of the discursive operation, not an incidental reference.
Exclusion criteria	Do not apply to comparative scholarly analysis; to inter-faith dialogue that does not subordinate Jewish meaning to a non-Jewish frame; to specific commemorative speech that does not mobilise the present-tense positive Jewish claim for an external project.
Justification	Operationalises Category 2 (instrumentalising forms) and the typology developed in Chapter Two: far-right anti-Muslim proxy (Rose, 2020), Hindu nationalist deployment, Christian supersessionism, guilt-engendered support and cultural appropriation.

Section 4. Confidence (Legibility Tier and Confidence Score)

Name	confidence
Definition	Section 4 records two values per item: (a) the legibility tier: how the philosemitic charge of the item is recoverable and (b) the confidence score on the assigned primary classification. Legibility tier (after Becker, Blatter and Stanevich, 2026): Explicit. Philosemitic framing is overt and unambiguous on the surface of the item. No inference from co-text or world knowledge required. Implicit. Philosemitic framing requires contextual or cultural knowledge (theological codes, gateway tropes, foreign-language insertions, allusions) to identify. Co-text. Item is not independently classifiable but acquires philosemitic or antisemitic meaning from the immediately surrounding thread, parent post or quoted post.

	Grey area. Item is ambiguous; reasonable annotators may disagree, and no co-textual or contextual evidence resolves the ambiguity. Under conservative coding, grey-area items receive Neutral as primary classification with the candidate philosemitic reading logged. Ambiguity tag Apply where intent or classification cannot be determined from the content. The tag is orthogonal to the legibility tier: an item may be legibly explicit yet still tagged ambiguous if (for example) sarcasm, irony or speaker stance are unstable. Confidence score 1. Low confidence. Classification is tentative; missing context, unclear sarcasm, ambiguous referent. 2. Medium confidence. Classification is plausible but contestable; reasonable annotators may disagree. 3. High confidence. Classification is strongly supported. Annotators record a one-sentence rationale for any score of 1 or 2 in Additional comments.
Inclusion criteria	All items receive a legibility tier and a confidence score, regardless of primary classification.
Exclusion criteria	None.
Justification	Operationalises the Decoding Antisemitism legibility framework (Becker and Troschke, 2023; Becker, Blatter and Stanevich, 2026) and the threshold-instability finding articulated in Category 3 (authenticity threshold).

Section 5. Authenticity & Engagement Signals

Name	authenticity_engagement
Definition	Section 5 records two values: (a) three binary indicators of authenticity, derived from the Authenticity Threshold of Category 3 and (b) one ordinal score for engagement depth. Indicators (binary: present/absent/not determinable): Consistency. The item is consistent with the speaker's broader or repeated positions, where this is observable from the visible co-text or account history. Mark not determinable for anonymous accounts, ephemeral content or where account history is inaccessible.

	Inconvenience. Support is expressed despite plausible social, professional, political or platform-algorithmic cost given the speaker's identifiable audience and platform context. Action-orientation. The item includes a concrete commitment, named organisation, donation, attendance or specific advocacy that is distinct from purely symbolic gesture. Engagement depth (ordinal: select the highest level the content clearly supports): Level 0: Absent/Irrelevant. No support, solidarity or engagement present. Level 1: Reactive (anti-antisemitism). Low-cost opposition to hate or violence. Level 2: Expressive (symbolic solidarity). Emotional or performative expression with excessive use of hashtags, emojis, mourning declarations. Level 3: Substantive engagement. Informed awareness of Jewish diversity, context, history or stated community needs. Level 4: Action-oriented. Calls for specific action, resource-sharing, advocacy or support despite cost. Level 5: Hostile transition. Positive framing accompanies or transitions into hostility, conspiracy or conditional support; the item should also be reviewed for the Both classification under Section 1.
Inclusion criteria	Presence of multiple indicators (Consistency, Inconvenience, Action-orientation) suggests higher likelihood of genuine support. Engagement depth is recorded for every item, including those classified Neutral or Not Relevant. The Level 0 / Level 1 distinction is diagnostically valuable for downstream analysis of allyship registers.
Exclusion criteria	Absence of indicators suggests potential performativity but is not determinative. Annotators do not infer intent beyond the available evidence. Where consistency cannot be assessed, mark not determinable rather than absent.
Justification	Operationalises Category 3 (the Authenticity Threshold), including the four markers of consistency, inconvenience, action-orientation and selflessness extracted from the interview corpus and the Signal Detection Theory framing applied to the engagement-depth ordinal.

Section 6. Context Metadata

Name	context_metadata
Definition	The following fields are completed for every annotation.

	Post type. Original post/reply/repost/quote post/image caption/short-form video. Account type. Individual/organisation/anonymous/verified/institutional/pseudonymous. Platform. Specify (X, Instagram, TikTok, Facebook, Reddit, YouTube, LinkedIn, forum name, etc.). Date of post. DD/MM/YYYY. Geographic location of account. Where available; do not infer from name, language or apparent ethnicity. Thread context. Brief summary of preceding posts where these are relevant to classification. Language. Specify. Note if translated. Speaker positionality. Recorded only where self-declared in the post or account bio, or provided in the dataset metadata. Do not infer.
Inclusion criteria	All fields completed for every item, even where the value is 'unknown' or 'not available'.
Exclusion criteria	Annotators do not infer demographic data (Jewishness, ethnicity, religion, political position) from names, profile images or apparent linguistic markers. Inference is recorded as 'unknown'.
Justification	Supports the corpus-collection protocol and operationalises the platform-context findings of Category 4. Anchored in the Quantitative Antisemitism Researcher's insights from the interview corpus.

Section 7. Detection Signals

Name	detection_signals
Definition	Annotators flag the presence of structural-linguistic signals identified in this study as the surfaces at which automated detection systems systematically fail. Each flag is binary (present/absent/uncertain) and is independent of the primary classification: a Neutral item may carry Detection Signals if the structural feature is present. Noticing syntax. The three-part structure of (i) affective opener, (ii) quantified claim about Jewish visibility or success and (iii) marker of excess (e.g. 'so many', 'again', 'how is it that', 'just an observation'). Often hashtag-mediated (#TheNoticing). Gateway tropes. Content deploying recognisable antisemitic tropes (Happy Merchant, Rothschild banker imagery) re-cast in admiring register.

	Dog-whistle indicators. Surface-positive content circulating within networks associated with far-right, extremist or conspiratorial communities, identifiable via account history, hashtag set or quoted material. Positive sentiment masking. Problematic discursive content carried by surface-positive lexical sentiment (love, admiration, support, pride, respect) such that conventional sentiment-based hate-speech classifiers would not flag the item. Multimodal cue. Image, meme, emoji, video text or layout contributes to meaning beyond the textual stratum. Annotated jointly with text where the alignment of strata produces a meaning available in neither alone.
Inclusion criteria	Apply each flag independently to every item, regardless of the primary classification. The Detection Signals dimension supports downstream construction of feature-level detectors and functional-test sets.
Exclusion criteria	Do not infer dog-whistle status from surface content alone; the flag requires evidence from co-text, account history or hashtag set. Where the network association is suspected but not demonstrable, mark uncertain.
Justification	Operationalises Category 4 (the Detection Gap) as a label-set of structural features, supporting downstream fine-tuning of detection models on the surfaces at which philosemitic content evades sentiment-based classifiers (Becker, Blatter and Stanevich, 2026; Becker et al., 2024).

Worked Examples

The three examples below are drawn from the empirical dataset of the present study. Each shows a fully completed annotation across all seven sections. A blank annotation form follows each example, for use in pilot annotation and training.

Example 1. Worked annotation

<i>@kellywillner's comment on @thefreepress and @benmfreeman's joint Instagram post 'I'm Leaving the UK' (3 October 2025).</i>		
1	Primary Classification	Philosemitic. The comment includes positively-valenced collective traits (contribution to science, art, literature, music, inventions, medicine) to 'the Jews' as a group, framing the loss of Jewish presence as a national impoverishment. The framing meets the inclusion criteria for Philosemitic and not the over-labelling firewall (because it condemns antisemitism without blaming an entire group).
2	Philosemitic Subtype (target framing)	Subtype: Intellectual/Cultural (primary); Resilience-based (secondary). Target framing: Symbols. The comment constructs Jews as a homogeneous collective whose value to 'one's country' is exhausted by their cultural-intellectual contribution. E.g. The rhetorical 'ANY country or culture would WANT the Jews' generalises the Imaginary Jew as a cultural-civilisational asset.
3	Instrumentalising Form	No (provisional).

		No external agenda is the discursive centre. Note: the parent post ('I'm Leaving the UK') and the visible co-thread (the 'standwithisrael' reply) suggest a possible Guilt/Redemption frame at the thread level; this is logged for thread-level review but does not change the comment-level judgement.
4	Confidence (legibility tier + score)	Legibility tier: Explicit. Confidence score: 3 (high). Ambiguity tag: not applied. The essentialising operation is on the surface with group-level positive attribution, generalisation to 'the Jews', civilisational framing.
5	Authenticity & Engagement Signals	Indicators: Consistency: not determinable (account history not consulted at item level). Inconvenience: absent (audience-aligned). Action-orientation: absent (no concrete commitment). Engagement depth: Level 2 (Expressive or symbolic solidarity).
6	Context Metadata	Post type: Reply. Account type: Individual. Platform: Instagram. Date of post: around September 2025 (parent post 3 October 2025). Geographic location of account: unknown. Thread context: Reply to Free Press/Ben M. Freeman 'I'm Leaving the UK' video; co-thread includes salterlynnette ('I'm a Christian but love Jewish people...'). Language: English (en). Speaker positionality: not declared.
7	Detection Signals	Noticing syntax: absent. Gateway tropes: absent. Dog-whistle indicators: absent. Positive sentiment masking: present (the comment would register as positive sentiment to a polarity classifier yet performs essentialising work). Meaning-form divorce: absent. Multimodal cue: present (image-text relation: the lamenting comment co-occurs with the 'I'm Leaving the UK' framing image). World-knowledge gap: absent. Surface text is sufficient.

Additional comments:

The 'such a positive' framing is structurally interesting.

Example 1 — blank form



thefreepress and benmfreeman

salterlynnette 26 w

I'm a Christian but love Jewish people, have loads of Jewish friends and I'm behind Israel all the way

1 like Reply

kellywillner 27 w

It's so ironic to me, I would think ANY country or culture would WANT the Jews as they are such a positive to society contributing science, art, literature, music, inventions, medicine, etc that only IMPROVES one's country. What a loss for England.

786 likes Reply

geoff_jesito_mclennan 27 w

21K 1.5K 348

3 October 2025

Add a comment...

Label @kellywillner's comment.

1	Primary Classification	
2	Philosemitic Subtype (target framing)	
3	Instrumentalising Form	
4	Confidence (legibility tier + score)	
5	Authenticity & Engagement Signals	
6	Context Metadata	
7	Detection Signals	
Additional comments:		

Example 2. Worked annotation

←

Post



Nigel Farage MP

@Nigel_Farage

⌵

...

What we witnessed in London at the historic Trafalgar Square, in a country built on Judeo-Christian values, was a group of people attempting dominance over our capital city and our culture.

We are not going to surrender everything that was built over centuries and defended at great cost in two world wars for us to be a free, independent nation.

The British people will not put up with this any longer — simple as.



10:55 AM · Mar 19, 2026 · 2.2M Views

💬

4.3K

↻

5.6K

❤️

27K

🔖

910

⬆️

Relevant ▾

View quotes >

@Nigel_Farage's post on X following events at Trafalgar Square (19 March 2026).

1	Primary Classification	Philosemitic. The post invokes 'Judeo-Christian values' as the founding civilisational frame of the country, deploying that frame as the rhetorical complement of an anti-Muslim claim ('a group of people attempting dominance over our capital city and our culture'). The pro-Jewish content of 'Judeo-Christian' is the discursive vehicle for the Islamophobic content; meets the Philosemitic inclusion criteria (Becker and Troschke, 2023, on the implicit register).
2	Philosemitic Subtype (target framing)	Subtype: Moral/Spiritual (primary); Cultural/Religious (secondary). Target framing: Instruments.

		The 'Judeo-Christian values' formulation positions Jewishness as a moral-civilisational anchor whose function in the post is to legitimate a defence of British nationhood against a threatening out-group. Jewish identity does not appear as Agent (no Jewish actors, communities or perspectives are named); it is deployed as instrument.
3	Instrumentalising Form	Yes: Far-right/Anti-Muslim Proxy (primary); Political Co-option (secondary). The structural mode is the one Rose (2020) names: pro-Jewish surface authorising Islamophobic content. The political-co-option label captures the deployment of the frame by an elected MP for a partisan purpose. Edge-case rule on 'Judeo-Christian values' applies: the phrase is labellable here because the contrastive Muslim-out-group function is the discursive centre.
4	Confidence (legibility tier + score)	Legibility tier: Implicit (philosemitic charge requires the cultural-political knowledge that 'Judeo-Christian values' performs Islamophobic work in this context); Co-text-confirmed (the 'group of people attempting dominance' clause makes the contrast explicit). Confidence score: 3 (high). Ambiguity tag: not applied.
5	Authenticity & Engagement Signals	Indicators: Consistency: present (consistent with the speaker's broader documented positions). Inconvenience: absent (audience-aligned; reinforces base). Action-orientation: absent at item level (no concrete pro-Jewish commitment or named organisation). Engagement depth: Level 2 (Expressive — symbolic). The post is a high-engagement performative statement; it does not advance any specific Jewish-community ask.
6	Context Metadata	Post type: Original post. Account type: Verified individual; institutional (sitting MP). Platform: X. Date of post: 19 March 2026. Geographic location of account: United Kingdom. Thread context: posted in the wake of the Trafalgar Square event; embedded video. Language: English (en). Speaker positionality: not Jewish (publicly known); political identity declared.

7	Detection Signals	Noticing syntax: absent. Gateway tropes: absent (the post does not deploy classical antisemitic visual tropes). Dog-whistle indicators: present ('Judeo-Christian values' functioning as a coded contrast term against Muslims in this discursive milieu). Positive sentiment masking: present (the philosemitic framing would not register negatively on sentiment-based classifiers; the Islamophobic content rides on the positive frame). Meaning-form divorce: present ('Judeo-Christian' names a civilisational pairing whose discursive function in this post is detached from any engagement with Jewish life). Multimodal cue: present (embedded video). World-knowledge gap: present (classification depends on the political-cultural knowledge that 'Judeo-Christian values' is performing anti-Muslim contrast work).
Additional comments: Paradigmatic far-right anti-Muslim proxy philosemitism in the sense developed in Chapter Two. The Liz Truss 'Jewish values' vignette in the body of the thesis is the closest analogue. The present case is more legibly Implicit-with-Co-text-confirmation because the Muslim-out-group is explicitly named in the same post.		

Example 2 — blank form

←

Post

Nigel Farage MP

@Nigel_Farage

What we witnessed in London at the historic Trafalgar Square, in a country built on Judeo-Christian values, was a group of people attempting dominance over our capital city and our culture.

We are not going to surrender everything that was built over centuries and defended at great cost in two world wars for us to be a free, independent nation.

The British people will not put up with this any longer — simple as.

0:55



10:55 AM · Mar 19, 2026 · 2.2M Views

4.3K

5.6K

27K

910

Relevant

View quotes

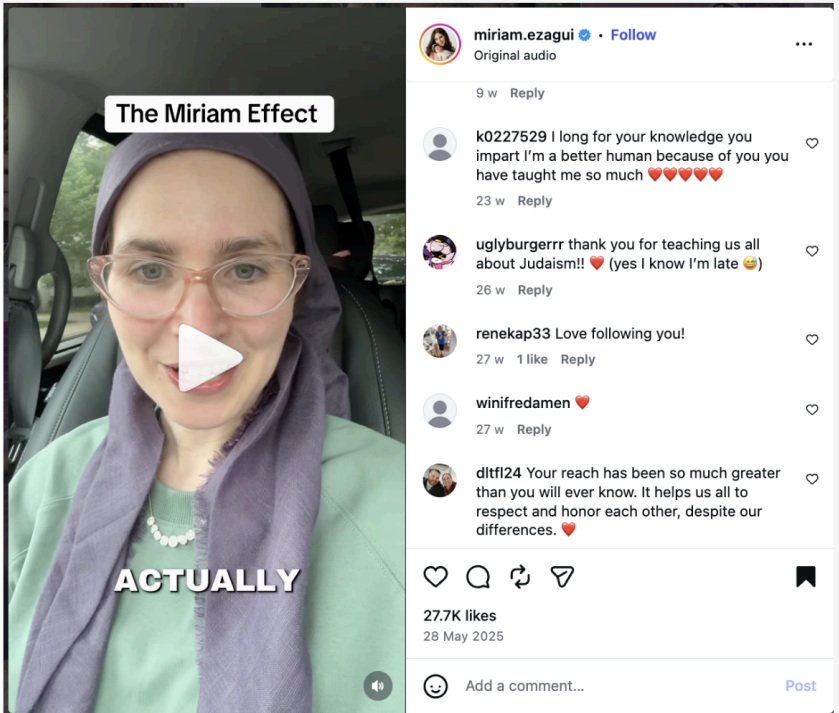
Label @Nigel_Farage's post.

1	Primary Classification	
2	Philosemitic Subtype (target framing)	
3	Instrumentalising Form	
4	Confidence (legibility tier + score)	
5	Authenticity & Engagement Signals	

96

6	Context Metadata	
7	Detection Signals	
Additional comments:		

Example 3. Worked annotation



@k0227529's comment on @miriam.ezagui's Instagram video 'The Miriam Effect — ACTUALLY' (28 May 2025, 27.7K likes).

1	Primary Classification	Neutral/Genuine Solidarity (under conservative coding). The comment is a personal expression of admiration directed at a specific Jewish creator's educational work, framed as the speaker's own ethical formation ('I'm a better human because of you have taught me so much'). The praise is anchored in a named individual's actions, not in a group-level claim about Jews. Edge-case rule on specific praise applies: 'specific praise is not group essence'.
2	Philosemitic Subtype (target framing)	Not applicable. Section 2 does not apply where Section 1 returns Neutral. No essentialising subtype is recorded.
3	Instrumentalising Form	No.

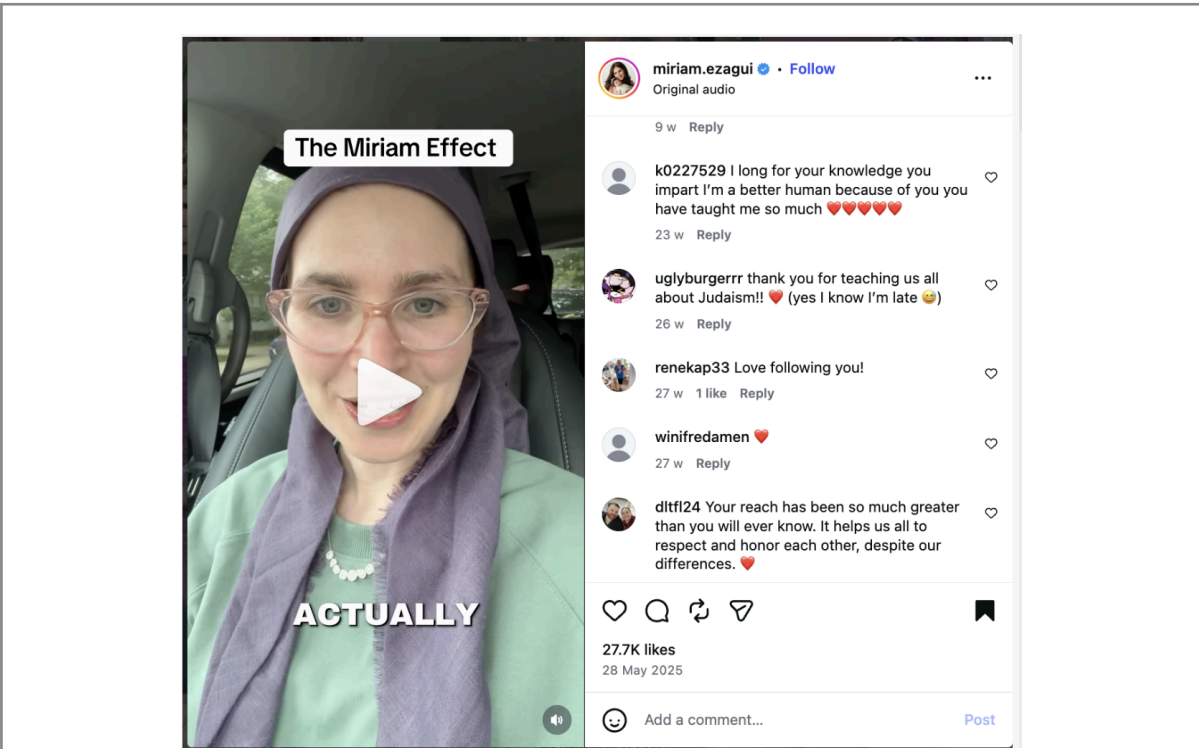
		No external agenda visible. The comment does not deploy Jewish identity to advance a non-Jewish project; the speaker positions themselves as recipient of education.
4	Confidence (legibility tier + score)	Legibility tier: not applicable (no philosemitic classification). The Neutral judgement is itself Explicit-tier in the sense that no co-text or world-knowledge inference is required. Confidence score: 2 (medium). Logged for adjudication. Ambiguity tag: applied. The 'I long for your knowledge' formulation, the multiple heart emojis and the 'better human' framing are emotionally charged and could, under a more permissive coding rule, be read as Cultural-Religious Fetishisation (Section 2). Under conservative coding (philosemitic-vs-neutral default to neutral), the Neutral label is preferred; the candidate philosemitic reading is logged here.
5	Authenticity & Engagement Signals	Indicators: Consistency: not determinable. Inconvenience: absent (warmly-received platform context). Action-orientation: absent (no concrete commitment beyond the comment itself). Engagement depth: Level 2 (Expressive or symbolic solidarity). The comment is appreciative but does not advance substantive engagement with Jewish history, communal context or stated needs in the way Level 3 would require.
6	Context Metadata	Post type: Reply (comment). Account type: Individual; pseudonymous. Platform: Instagram. Date of post: around September 2025 (parent post 28 May 2025; sampled close to parent). Geographic location of account: unknown. Thread context: comment thread under @miriam.ezagui's 'The Miriam Effect' Reel; co-thread includes 'thank you for teaching us all about Judaism!!' and similar appreciative comments. Language: English (en). Speaker positionality: not declared.
7	Detection Signals	Noticing syntax: absent. Gateway tropes: absent. Dog-whistle indicators: absent. Positive sentiment masking: not applicable (item is Neutral; no problematic discursive content is masked). Meaning-form divorce: absent.

		Multimodal cue: present (the 'ACTUALLY' caption-image and 'The Miriam Effect' framing are integral to the parent post's discursive shape, but the comment itself does not depend on the image for classification). World-knowledge gap: absent.
--	--	---

Additional comments:

Aggregating engagement-style appreciative comments under @miriam.ezagui's Reels into a 'philosemitism' corpus would over-label genuine cultural-educational engagement with a Jewish creator's own work. The conservative coding rule routes the item to Neutral while preserving the candidate philosemitic reading for downstream review. Annotators should expect items of this shape to populate the medium-confidence neutral region of the corpus.

Example 3 — blank form



Label @k0227529's comment.

1	Primary Classification	
2	Philosemitic Subtype (target framing)	
3	Instrumentalising Form	
4	Confidence (legibility tier + score)	

5	Authenticity & Engagement Signals	
6	Context Metadata	
7	Detection Signals	
Additional comments:		

Appendix B: Code to Clean BSA Dataset

```
# -*- coding: utf-8 -*-
"""Cleaned Dataset for Guilt-Engendered Love
Automatically generated by Colab.
Original file is located at
https://colab.research.google.com/drive/18UmZ-FiE6v5lONqkImbÜbs-_Feu4qBX
Mounting Google Drive Folder
"""

from google.colab import drive
drive.mount('/content/drive')

from google.colab import files

"""Loading the Dataset: Sample dataset from 70,000 online posts from three Boolean Queries"""

import pandas as pd
import numpy as np
import seaborn as sns

df_guiltlove = pd.read_csv('/content/drive/MyDrive/SA/Thesis /Online data/BSA Data/Guilt-engendered love/guiltlove.csv')
df_guiltlove.head(2)

"""With 186 columns, I am going to filter these down. First looking at the variables and data structure and then dropping missing data and obsolete columns."""

df_guiltlove.head()
df_guiltlove.columns
print(df_guiltlove.columns)

df_guiltlove = df_guiltlove.drop(columns=["Query Name", "Language", "Url", "Added", "Original Url", "Thread Author", "Thread Created Date", "X Replies",
                                         "X Reply to", "Category Details", "Entity Info", "Expanded URLs", "Facebook Author ID"], axis=1)

print(df_guiltlove)
df_guiltlove.info()
df_guiltlove.head(20)

# removing non-mainstream social media page types
page_types_to_remove = ["news", "blog", "broadcast", "48030443351", "48030449673", "print"]
df_guiltlove = df_guiltlove[df_guiltlove['Page Type'].isin(page_types_to_remove)]

# replace NaN values with 0
df_guiltlove['Likes'] = df_guiltlove['Likes'].fillna(0)
df_guiltlove['Shares'] = df_guiltlove['Shares'].fillna(0)

df_guiltlove

#merging Instagram and Facebook public pages with the rest of their said columns
df_guiltlove['Page Type'] = df_guiltlove['Page Type'].replace({'facebook_public': 'facebook', 'instagram_public': 'instagram'})

"""# Visualising the Data"""

import matplotlib.pyplot as plt
import seaborn as sns

# Plot 1: Bar graph for posts per country, including NaN values
country_counts = df_guiltlove['Country'].value_counts(dropna=False)
plt.figure(figsize=(12, 6))
sns.barplot(x=country_counts.index.fillna('Unknown'), y=country_counts.values, hue=country_counts.index.fillna('Unknown'), palette='viridis', legend=False)
plt.title('Distribution of Posts by Country (Including Unknown)')
plt.xlabel('Country')
plt.ylabel('Number of Posts')
plt.xticks(rotation=45, hha='right')
plt.tight_layout() # Adjust layout to prevent labels from overlapping
plt.show()

# Plot 2: Bar graph for posts per page type
page_type_counts = df_guiltlove['Page Type'].value_counts()
plt.figure(figsize=(12, 6)) # Adjust figure size for bar plot
sns.barplot(x=page_type_counts.index, y=page_type_counts.values, hue=page_type_counts.index, palette='viridis', legend=False)
plt.title('Distribution of Posts by Page Type')
plt.xlabel('Page Type')
plt.ylabel('Number of Posts')
plt.xticks(rotation=45, hha='right')
plt.tight_layout()
plt.show()

import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd

# Ensure 'Date' column is in datetime format
df_guiltlove['Date'] = pd.to_datetime(df_guiltlove['Date'], errors='coerce')

"""## 1. Pie Chart for Sentiment Distribution"""

# Calculate sentiment counts
sentiment_counts = df_guiltlove['Sentiment'].value_counts()

# Plot pie chart
plt.figure(figsize=(8, 8))
plt.pie(sentiment_counts, labels=sentiment_counts.index, autopct='%1.1f%%', startangle=90, colors=sns.color_palette('pastel'))
plt.title('Distribution of Posts by Sentiment')
plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show()

import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd

# Calculate the count of each sentiment for each page type
sentiment_page_type_counts = df_guiltlove.groupby(['Page Type', 'Sentiment']).size().unstack(fill_value=0)

# Calculate proportions within each Page Type
sentiment_page_type_proportions = sentiment_page_type_counts.apply(lambda x: x / x.sum(), axis=1)

# Define a specific order for sentiment categories
sentiment_order = ['negative', 'neutral', 'positive']
sentiment_page_type_proportions = sentiment_page_type_proportions[sentiment_order]

# Plotting the stacked bar chart
sentiment_page_type_proportions.plot(kind='bar', stacked=True, figsize=(12, 7), colormap='viridis')
plt.title('Proportion of Sentiment per Page Type')
plt.xlabel('Page Type')
plt.ylabel('Proportion')
plt.xticks(rotation=45, hha='right')
plt.legend(title='Sentiment')
plt.tight_layout()
plt.show()

df_guiltlove.to_excel('guiltlove_data.xlsx', index=False)

from google.colab import files
files.download('guiltlove_data.xlsx')
```

Appendix C: Vignettes Read to Interviewees

These vignettes were read aloud during interviews, therefore no images have been recorded in the table. I asked for interviewees to tell me if they had any reflections to share.

Liz Truss (posted campaign statement) on August 12th 2022:

(Wedgewood, 2022)

“So many Jewish values are Conservative values and British values too, for example seeing the importance of family and always taking steps to protect the family unit; and the value of hard work and self-starting and setting up your own business.”

Geert Wilders Tweet on 9 June 2024:

<https://x.com/geertwilderspvv/status/1799475397509132643>

‘And it is my duty as leader of the biggest political party in The Netherlands to do all I can to make life safer for Jews in Europe. To fight antisemitism and jihadism. To support Israel, especially when most European politicians are abandoning them. I never will.’ (2m views, 41,000 likes, 6,100 retweets, 3,700 comments).

Rassemblement National Tweet on 11th October 2023:

https://x.com/RNational_off/status/1712102867920691622

● 436 anti-Semitic acts recorded in 2022: with its crazy immigration policy, the government is endangering the Jews of France.

Zero tolerance for anti-Semitism: the RN will be the shield of French Jews against Islamism 👉

<https://adhesions.rassemblementnational.fr>

Rassemblement National Tweet on 20th March 2024:

https://x.com/RNational_off/status/1770397230991597772

● Let's stop the EU's complacency towards Islamism. It is a danger to our values and our security.

It is time to act firmly to protect our societies and preserve our identity and our civilization of Judeo-Christian essence.’

Robert Jenrick Tweet on October 3rd 2025 (120k views):

<https://x.com/RobertJenrick/status/1974035534524637556>

Today, being a Jew in Britain will feel lonely and scary. I am appalled by the extraordinary violence visited upon our streets in Manchester, but I can't say I am surprised. For years the Community Security Trust has been training thousands of volunteers to provide security outside synagogues for a moment like this. Their warnings have largely fallen on deaf ears. This attack didn't come out of a clear sky. It is the product of years of rampant antisemitism. We saw it go mainstream under Jeremy Corbyn. But even after his departure from front-bench politics, it never went away. Antisemitism reached unbelievable levels after October 7th. Back then MPs egged on the hate marches in London that were peppered with extremism and antisemitism. When rabid antisemitism goes unchallenged time after time across our society, tragic attacks like the one today are the destination. In the

coming days we will see the ritual condemnation of the attack. The tweets will be copied and pasted from the last incident, the empty “I stand with” graphics ready to be deployed and public buildings to be illuminated again.

But warm words ring hollow when no action follows. For years radicalisation has gone unchallenged. It has been allowed to grow and grow by a pathetically weak and cowardly elite class. The capitulation of so many of our institutions to antisemitism have made Jewish Brits feel like second class citizens. Our universities have allowed antisemitism to fester. We have seen Jewish students have swastikas on their doors; universities stand by antisemitic professors. The violent chant to “Globalise the intifada” has been made fashionable by left-wing students who consider antisemitism to be the one acceptable form of racism. Last year an Islamic charity in Nottingham hosted a vile sermon in which a trustee said “the hour will not begin until the Muslims fight the Jews and the Muslims will kill them until a Jew hides behind a rock or a tree.”

The Charity Commission responded to this call to violence against the Jews with merely a warning. We even see it in the NHS, too. This week a doctor, who is also a member of Palestine Action, launched into an unhinged rant that the Royal Free Hospital was a “Jewish supremacy cesspit” and that the UK is “occupied and controlled by Jewish supremacy”. “I don’t condemn Hamas. I don’t condemn October 7”, she said. The medical tribunal has not suspended her. At times our justice system has barely given antisemitism a slap on the wrist. When people paraded through Whitehall with images glorifying the Hamas paragliders, they were let off with a conditional discharge. Tam Ikram, deputy chief magistrate and a judicial appointments commissioner, turned a blind eye and said that he “did not find that they were seeking to show any support for Hamas”. I am a Christian. But my wife happens to be Jewish. This week I received an email. It threatened I would have my wife raped and my children shot or starved. “That’s what you deserve. You Zionist pig.” The individual was so confident there would be no repercussions that he gave his name. I am used to it now. I have had antisemitic death threats posted through my letterbox and delivered to my office. It is now water off a duck’s back. They will not silence me. And nor will attacks like we saw today silence the Jewish community. Today, being a Jew in Britain will feel lonely. Families will be thinking about their future here and whether they are safe. They should not live like this. We cannot live like this. We are not helpless victims. We can choose to root out this cancer if our leaders actually care. I know where I stand.

Appendix D: Consent Form Template



DATA PROTECTION

INFORMATION SHEET AND CONSENT FORM

I (Lastname, First name)..... am willing to participate in an interview in the thesis project, led by Gabriella Cohen, Master's Student at Sciences Po at the School of Public Affairs.

LEGAL INFORMATION

Objectives : This thesis aims to explore new expressions of antisemitism (called philosemitism) through interviews and use these quote insights to create an annotation guideline and to point towards online examples. Participating is completely voluntary.

Data controller : Gabriella Cohen, Supervised by Professor Sylvain Parasie.

Data recipients (researchers who access to the data):

- Gabriella Cohen, Master's Student
- Sylvain Parasie, Master's Student Supervisor

Category of data processed : identification data, sociodemographic data, research data, consent and technical data.

Data protection measures : There will be no sharing of the interview data as it will be stored internally (on one laptop) and deleted after its purpose for informing thesis research. It will be shared only to the university for marking. Consent prior to the interview will be verbally confirmed too.

Risk for participants :

- There is minimal risk for participants who may remain anonymous if so desired.
- Microsoft Office Suite and Google Workplace will be used to store the interview data during the thesis period and deleted afterwards.
- Team auto-transcription may be used, depending on length of interview to transcribe interview data.

Data archive and duration : The data will be treated until 28/04/2026 until thesis submission, graded, then deleted.

GDPR rights : In accordance with the General Data Protection Regulation 2016/679, any person will be able to access their data, as well as remove it if possible. To do so, I can send an email to gabriella.cohen@sciencespo.fr. In case of difficulty, please contact the Sciences Po Data Protection Officer cnil@sciencespo.fr. If your question is not answered properly, you can address a complaint to the French data protection authority (the CNIL).

Sciences Po is committed to ensure the respect of the European and French data protection principles by its current and future employees, researchers or eventual processors associated with the data.

By contributing to this thesis project, the research participant confirms:

- ☐ I have read, understood and accepted these conditions for the processing of personal data and freely participate in the research project; (opt-in checkbox required).
- ☐ I would like to remain anonymous during interviews (optional opt-in checkbox).
- ☐ I consent to the reuse of their personal data for scientific research purposes in the public interest. (optional opt-in checkbox).

_____(Signature)

_____(City)

_____(Date)

Public Policy Master's Thesis Series

This series presents the Master's theses in Public Policy and in European Affairs of the Sciences Po School of Public Affairs. It aims to promote high-standard research master's theses, relying on interdisciplinary analyses and leading to evidence-based policy recommendations.

Hiding in Plain Sight: the Conceptualisation and Operationalisation of Online Philosemitism.

Gabriella Cohen

Abstract

Antisemitism has an unknown doppelgänger that hides in plain sight online: philosemitism. Philosemitism is the excessive admiration for Jewish people, mounting the same stereotyping processes as antisemitism. Despite this, there has been little systematic examination of how the phenomenon translates onto online contexts. Addressing this gap, this study uses a mixed-methods approach grounded in Constructivist Grounded Theory to explore how online philosemitism can be conceptualised and operationalised for further research. To do so, it uses twelve expert interviews between January and April 2026, coded through an iterative process and saturated with cross-platform social media data, to produce the conceptual framework and annotation guideline. The findings confirm online philosemitism reproduces essentialised Jewish stereotypes regardless of intent and that its positive appearance generates a structural detection gap in existing automated detection systems. The annotation guideline assembled is a first step to closing this gap. By integrating philosemitism into antisemitism education programmes, funding a multilingual labelled dataset and incorporating philosemitism into platform risk assessments, philosemitic behaviours can be brought into view. It concludes that philosemitism and antisemitism feed the same imaginary construction of Jewish identity. Therefore, any serious effort to monitor one must eventually account for the other.

Key words

Philosemitism; Judaism; Automated Detection; Annotation Guidelines; Essentialisation

This paper at hand made use of artificial intelligence for paraphrasing, rewording and structuring suggestions and formatting.