



On the Obsolescence of Hate Speech Detection for Memory-based Conflicts

and Technopolitical Conflict Transformation
in the Armenian Struggle

Fatima K. Abu Salem

Wissam Saade

On the Obsolescence of Hate Speech Detection for Memory-based Conflicts

and Technopolitical Conflict Transformation
in the Armenian Struggle

Research Report

Fatima K. Abu Salem Wissam Saade

© All Rights Reserved. Beirut, February 2026.

This Analysis is published by the Issam Fares Institute for Public Policy & International Affairs (IFI) at the American University of Beirut, made possible (in part) by a grant from Carnegie Corporation of New York, and is available on the following website: <http://www.aub.edu.lb/ifi>.

The views expressed in this document are solely those of the author, and do not necessarily reflect the views of the Issam Fares Institute for Public Policy & International Affairs or that of Carnegie Corporation of New York.

About the Authors

Fatima K. Abu Salem is Professor of Computer Science at the American University of Beirut. She holds an MS in pure mathematics from AUB and a DPhil in Computing from the University of Oxford. Her former research area has been in Computer Algebra, with a focus on developing parallel and cache efficient algorithmic designs for algebraic computations at scale. Her recent research area is in data science with impact, with applications in the social, health, agricultural and environmental sciences. In her work, she incorporates advanced machinery that adheres to trustworthy Machine Learning requirements such as ML for small data, distributional robustness, probabilistic forecasting, and uncertainty quantification. She has spoken multiple times on challenges affecting women in computing and mathematicians in the developing world, and served as secretary for the special activity group on Supercomputing for the Society of Industrial and Applied Mathematics. She currently serves as associate editor for the Journal of Parallel and Distributed Computing, and member of the ACL special interest activity group on Arabic Natural Language Processing.

Wissam Saade is an IFI Associate Fellow and Lecturer of Political Science and History at Saint Joseph University since 2003. He is a regular op-ed writer in leading Lebanese and Arab newspapers. Saade's research interests span medieval and modern political thought, the social and intellectual history of modern revolutions, as well as nationalism and ethnicities in South Asia and the Middle East.

Table of Contents

Abstract	6
Memory-based Conflicts: A Comparative Lens	6
Diaspora, Digitization, and the Technopolitics of Memory	8
On the Obsolescence of Automated Hate Speech Detection in Memory-based Conflicts	10
Technopolitical Prospects for Conflict Transformation	11
Technopolitical Moral Enhancements	11
Technopolitical Mobilisation	12
Discussion and Conclusion	14
References	16

Abstract

The Armenian struggle remains an intractable memory-based conflict involving identity-driven disputes that are deeply rooted in the irreconcilable collective memories and interpretations held by different groups. Modern embodiments of associated grievances have begun permeating the digital sphere through manifestations of genocide denial, a mummified collective memory, and competitive victimhood. As automated hate speech detection methods rely on word embeddings trained on textual data with short temporal scales, it can be argued that digital platforms obscure the Armenian conflict's relational and historical dimensions, and suppress discursive practices necessary for conflict transformation, such as the proper archival of lived experience, acknowledgment of harm, and the possibilities of counter narratives. Specifically, we argue that online platforms that are governed exclusively using the conceptual framework associated with hate speech cause a stagnation in the repetitive ways that the conflict is processed and conceal the relational shifts required for sustainable peace. In this op-ed, we explore ways by which alternate sociotechnical devices can promote technopolitical mobilisations towards conflict transformation inspired by H. Arendt's characteristic of action in the "Human Condition", which renders obsolete the prevalence of modern-day, "hate speech" parlance. The alternatives that we promote pivot around moral enhancements for guiding the adoption of such technologies in light of human suffering under conditions of digital genocide, a deeper long durée understanding and archival of collective memory on digital platforms, and content generation tools that can shift the competitive victimhood narrative in favor of low-power groups.

Memory-based Conflicts: A Comparative Lens

The Armenian struggle can be understood as a protracted, memory-centered conflict in which identity and politics shape divergent collective narratives. The prolongation of this specific conflict is intensified by the fact that identity practices influence aspects of acknowledgment, recognition, and denial. Conflict resolution in this context, therefore, requires more than historical verification; it commands a transformation in the political and emotional ecosystems through which communities narrate their grievances and memory.

While the historical event is broadly recognized in genocide scholarship, contemporary disputes, specifically on digital platforms, persist through entrenched national memories, denialist narratives, and forms of competitive victimhood. New modes of information capture are exacerbating manifestations of such forms of disputes in digital spaces. The Armenian case illustrates how memory conflicts become difficult to resolve: they are not simply about what happened, but about who communities believe they are because of what happened.

Yosef Hayim Yerushalmi's seminal work, *Zakhor: Jewish History and Jewish Memory* (1982), lends a crucial analytical lens for understanding contemporary memory conflicts. Yerushalmi distinguishes between history (the critical, evidence-based reconstruction of past events) and memory, which is inherently identity-driven, transmitted within communities, and concerned with deriving meaning behind a certain struggle, rather than empirical accuracy. This distinction is particularly relevant to the Armenian case, where debates over recognition and denial are less about the factual existence of the genocide and more about collective identity, legitimacy of the cause, and political narrative.

Within Yerushalmi's framework, memory serves to sustain group cohesion and continuity, which explains why acknowledgment or denial carries symbolic and political weight beyond historical documentation. By highlighting the ways in which memory can diverge from history, Yerushalmi illuminates why some post-genocide conflicts persist. In that line of thought, the challenge is not simply constrained to historical verification, but rather with the emotional engagement one can strike with the cultural and political economies, those through which communities narrate and negotiate controversies surrounding their past.

This interpretive lens resonates with the work of Ronald Grigor Suny, who argues that the contemporary Armenian-Turkish divide is shaped less by empirical uncertainty than by conflicting national mythologies. In accordance with this observation, each distinct narrative promotes a particular understanding of state legitimacy and historical continuity. Scholars like Taner Akçam and Fatma Müge Göçek extend this argument by demonstrating how denial operates within Turkey not simply as a rejection of archival evidence but rather as a mechanism of state-building and legal rationalization, as well as identity preservation. It follows that the persistence of dispute cannot be attributed solely to conflicting documentation or archival and scholarly research, but over what acknowledging the event would mean for collective self-understanding.

Contrasting the Armenian memory conflict against the Holocaust and the Roma genocide helps us understand the impact that the institutionalization, recognition, and marginalization of memory can have on shaping post-genocidal societies differently. The dynamic embedded in the recollection of Armenian memory stands in contrast to the post-Second World War treatment of the Holocaust. In the European context, and in the aftermath of the Holocaust, there erupted an unprecedented alignment between international acknowledgment, legal adjudication, and institutionalized memory practices. The Nuremberg Trials not only prosecuted individual actors but also produced a juridical framework that redefined mass violence under emerging legal categories such as “crimes against humanity”. Most notably, it also further informed the 1948 Genocide Convention.

This legal foundation was later reinforced and helped embed the Holocaust into public consciousness through a plethora of channels, including but not limited to educational policy, state memorialization strategies, and the creation of transnational research infrastructures (such as museums, documentation centers, and academic institutes). It is in this context that Aleida Assmann’s notion of a “normative memory regime” becomes analytically useful. For Assmann, the Holocaust did not simply enter historical records; it became a regulative framework that set standards for how states and societies should remember and respond to large-scale violence. Assmann’s regime establishes expectations regarding Holocaust commemoration versus denial, as well as historical and legal accountability, and is maintained not merely by formal law but by what Assmann identifies as an informal cultural consensus: a shared understanding that recognition is tied to democratic legitimacy, and that explicit denial is incompatible with membership into the post-war European political order. As such, the Holocaust is considered to operate within a normative memory regime, and became central to post-war memory regimes.

Contrasting the Holocaust against the Armenian case, the latter stands as a contested memory field. And the genocide of the Roma (Porajmos) exemplifies a third type of memory regime in this comparative frame. Unlike the Holocaust, the Roma experience remained significantly under-recognized and structurally silenced for decades, and only very recently did it begin to attract attention from scholars and some form of institutional recognition. The challenges associated with the Roma memory regime are not exclusive to denial, but include manifestations of historical invisibility as evident in the absence of documentation and public memorial infrastructures. This resulted in a situation where memory around the Roma genocide could not be internationally or collectively acknowledged. In this sense, the Roma case demonstrates how a genocide, whilst still historically established, can remain socially and politically under-remembered, resulting in what is known as a memory gap rather than a memory conflict.

Comparative reflection clarifies these dynamics. The Jews, Armenians, and Roma experienced twentieth-century catastrophes differently, not due to variations in scale or horror but because of unequal mediating structures. Jewish communities transformed trauma into institutionalized structures; Armenians, in 1915, lacking any formal recognition as a sovereign state, failed to preserve genocidal evidence or protect survivors. Without proper institutions to confront denial, their collective memory emerged in exile, under fragility. The perpetrator state got immersed in genocide denied, destroyed evidential archives, exerted pressure on allies and lobbied other states at the international level, thereby subjecting Armenian memory to a systematic act of organized erasure. Efforts to recognize the Armenian collective suffering since then have been slow, uneven, and reactive.

The digital age was expected to democratize remembrance but instead has amplified pre-existing asymmetries between groups with strong institutions, those with dispersed identities, and those with virtually no infrastructure. Consequently, Armenian memory remains emotionally intense and globally diffused but strategically vulnerable to state-sponsored disinformation campaigns. The Roma faced the most tragic structural conditions. The Roma memory enters the digital and AI era from the weakest structural position: no state, limited intelligentsia, marginal languages, low digital representation, minimal institutional archiving of the genocide, and almost no algorithmic authority. What the algorithm cannot index, it cannot remember.

This hierarchy of suffering, as Benbassa emphasizes, reveals that the (moral, political, and cultural) centrality of memory is not proportional to the magnitude of collective pain experienced at the level of the group. Instead, it is incumbent on the capacity of a community or state to institutionalize and politicize its suffering. As exemplified in today's digital and AI era, Jewish memory functions as a global civil religion, hyper-documented and algorithmically reinforced. Armenian memory is heavily contested: a battlefield of identity against powerful, coordinated state denial.

Diaspora, Digitization, and the Technopolitics of Memory

Over time, Armenian collective memory has gradually drifted from the full complexity of its historical context, slowly adopting features of what can be described as a mummified collective memory: one that is selectively and emotionally transmitted and relatively isolated from the realms of critical debate. From a social science and genocide studies perspective, this reflects the ways in which trauma gets embedded within ethnic and national identity, as demonstrated in Vamik Volkan's concept of chosen traumas: historical suffering becomes a persistent marker of group cohesion and intergenerational continuity. Chosen traumas produce ambivalent consequences at the heart of collective identity such as internal and intergenerational cohesion as well as external rigidity at the same time. They also result in cognitive and emotional inflexibility: the trauma becomes a canonical narrative, resistant to reinterpretation, and a lens through which all intergroup relations are perceived. This aligns with research on ethnic boundary maintenance and symbolic ethnicity, in which shared memories of victimization reinforce group distinctiveness and the perception of existential threat.

In contemporary Armenia, the state plays a focal role as guardian of collective memory. Memorials such as Tsitsernakaberd in Yerevan serve as sites of mourning as well as for political messaging and nation-building. They further serve to institutionalize remembrance through educational and civic engagement around the narrative of historical suffering. At the same time, alternative interpretations of history or dissenting perspectives around the genocide are often perceived as threats to national unity. For example, it is precisely because institutionalized memory is deeply tied to national solidarity that attempts to contextualize the Armenian genocide within broader Ottoman wartime dynamics or to critically assess decisions made by Armenian political actors can provoke intense debate (Libaridian, 2004 Suny, 2015).

In the diaspora, however, these debates take on a different character. Some peoples have experienced diaspora as a constitutive element of their collective life, arising from long cycles of exile, trade networks, persecution, imperial fragmentation, economic out-migration, or genocide and ethnic cleansing. A nation becomes diasporic when its historical continuity depends as much on networks across borders as on the population within its territory.

The Armenian nation is one of the rare historical communities for which diaspora is not a deviation from national life but one of its enduring structural conditions. For centuries, the Armenian people have existed in a dual configuration: a territorial core in the Armenian Highlands and a vast extraterritorial world extending across the Middle East, the Mediterranean, Eastern Europe, and later on, the Americas. This duality is not merely demographic; it is constitutive of Armenian identity itself.

Unlike many modern diasporas that emerged in the nineteenth and twentieth centuries, the Armenian diaspora has an early origin in commercial colonies dating to antiquity and medieval trading networks. It developed institutional autonomy through the church, schools, and guild networks, and served as a mediator between empires, including the Byzantine, Persian, Ottoman, and Russian. The 1915 genocide did not create the Armenian diaspora; rather, it radically expanded and traumatically re-encoded an already existing diasporic structure. The diaspora became the surviving body of a nearly annihilated nation, the repository of memory, and the carrier of the national narrative when the homeland was fragmented.

Freed from the constraints of a state-centered national narrative, Armenian diaspora communities often emphasize identity preservation and political advocacy and mobilization. Here, the Armenian Revolutionary Federation (Dashnaktsutyun or Dashnak) plays a particularly prominent role, using the memory of the genocide to organize political activism and sustain a transnational sense of Armenian identity through campaigns for recognition (Panossian, 2006).

Armenian identity is built on an unusual tension between those Armenians who live within the homeland and those in the diaspora who retain a robust national consciousness while residing in host societies. In the twentieth century, this diasporic identity was sustained through Armenian schools, churches, political parties such as the ARF and AGBU, local newspapers, and community clubs. Today, however, digital media bypass these institutions entirely. A seventeen-year-old in Paris or Los Angeles often learns about Armenia not through parish schools or political publications but through Instagram feeds, Telegram channels, YouTube influencers, and TikTok narratives. Younger generations consume history in the form of short videos, memes, and algorithmically curated feeds. This transformation risks compressing the genocide into a distant symbol, detached from its archival complexity.

In the digital age, the complexities underlying those dynamics become more pronounced. Social media platforms, online forums, and digital archives allow diaspora communities to rapidly engage younger generations in identity politics by disseminating commemorative content and mobilizing transnational campaigns. But this unprecedented ease by which information flows nowadays comes with a caveat, as online users are instantly exposed to competing narratives and denialist content that perpetuate ideological fragmentation. Artificial intelligence further reshapes such memory politics. It amplifies all narratives and instigates a competition among victimhood claims where visibility is scarce and contested. In a global environment that is increasingly oblivious to factual accuracy, smaller nations like Armenia are particularly vulnerable to the plethora of denialist and pseudo-historical artifacts that AI can help propagate with high speed, rendering archival truth to be completely irrelevant and sidelined.

In this context, Turkey and Azerbaijan enjoy notably more structural advantages over Armenia. Turkey combines a large tech base, strong universities, and emerging AI laboratories. It has already been integrating AI into defense systems and running a robust, state-coordinated digital narrative machine. Azerbaijan, smaller but significantly well financed as an energy hub, is rapidly adopting advanced military AI technologies and synchronizing its digital propaganda with state policy. Armenia, by contrast, faces considerable constraints that hamper its ability to roll out advanced AI and other digital transformations across its institutions. This technological vulnerability renders Armenia to be specifically susceptible to narrative and technological asymmetry.

Nonetheless, digitization and diaspora activism allow for increased global visibility through expanded access to archives and oral-history projects and platforms like the Armenian Genocide Museum-Institute. Yet, while the digital sphere democratizes collective memory, it also enables algorithmic distortion that can orchestrate forms of organized denial and political pressure. This is because online visibility, even when appearing to sustain recollection and remembrance, can still fail to substitute for reparative justice or formal recognition. As memory is increasingly mobilized in contemporary conflicts, historians and civil society must remain vigilant against its instrumentalization in the digital sphere.

Armenians thus enter the digital and AI era in paradoxical conditions: they possess a literate and active diaspora memory yet confront powerful rival states who can dispatch coordinated propaganda, online bot networks, and carefully curated narratives. In this realm of competition, AI and content moderation algorithms can reinforce existing inequalities: memory that is archived, documented, and retrieved online is amplified, while unrecorded memory dissipates into oblivion. The digital preservation of history thus mirrors its structural asymmetries.

Against this backdrop of technopolitically charged entanglements, the impending question at the core of the present manuscript is the following: in what ways are automated hate speech detection technologies deemed to be obsolete, not as a technological tool, but rather as a conceptual framework for understanding such memory-based manifestations in the digital age? And what alternate sociotechnical constructs can offer a myriad of other possibilities for disrupting the stagnation around transformation of conflicts including but not limited to the Armenian conflict?

The alternatives that we promote pivot around moral enhancements for guiding the development of such technologies in light of human suffering under conditions of digital genocide, a deeper long durée understanding of collective memory on digital platforms, and a sociotechnical mobilization to counterbalance the prevalent competitive victimhood discourse in favor of low-power groups.

On the Obsolescence of Automated Hate Speech Detection in Memory-based Conflicts

Automated hate speech detection conceptualizes harm through linguistic markers that capture sentiment polarity or direct speaker-to-target hostility. Despite the fact that modern detection systems are increasingly able to capture contextual information underlying those linguistic markers, this context can only conceive of polarity and social-bias frames as a function of short-term temporal scales by borrowing from the Braudelian conceptual framework. According to Fernand Braudel (Braudel, 1980), history operates on different time scales characterized by the length of their temporal structures. The *longue durée* structure refers to very slow-moving structures that last for centuries, such as geography, economic systems, social hierarchies, and collective ways of thinking, which quietly shape historical processes. Of shorter length are conjunctures, identified as medium-term cycles that last for decades, capturing fluctuations in economic trends, demographic changes, and other forms of instabilities. Finally, the *short durée* focuses on events and individuals in the context of wars, civil upheavals, and political decisions and shifts. As they tend to be the most dramatic, visible and tractable, such short-term temporal structures form the backbone of context that underlies modern hate speech detection systems. However, they remain to be the least powerful level of history and are embedded within the deeper long durée. Consequently, they rest on the assumption that harm is explicit and can only be traced to isolated manifestations at the level of the “individual”. In contrast, discursive violence associated with memory-based conflict unfolds through other forms of manifestations such as denial, asymmetrical power relations, and narratives framed within competitive victimhood at the level of the “collective”.

Against this obsolescence, we hereafter explore ways by which alternate sociotechnical devices can promote technopolitical mobilizations towards conflict transformation that pivot around moral enhancements for guiding the adoption of such technologies in light of human suffering under conditions of digital genocide, a deeper long durée understanding and archival of collective memory on digital platforms, and content generation tools that can shift the competitive victimhood narrative in favor of low-power groups.

Technopolitical Prospects for Conflict Transformation

The contemporary use of the term citizen science emerges from two distinct epistemological perspectives rooted in their respective disciplinary origins (Haklay et al., 2021). The first perspective, developed by Alan Irwin (Irwin et al., 1994; Irwin, 1995), emphasizes the involvement of citizens as stakeholders in the outcomes of research, particularly in areas like public and environmental health. Irwin positions citizen science “at the point where public participation and knowledge production, or societal context and epistemology, meet” (Irwin, 2015). He contends that such approaches create opportunities to foster closer connections between the public and science, advancing the notion of an engaged ‘scientific citizenship’ with direct relevance to public policy. The second perspective, articulated by Rick Bonney (Bonney, 1996), highlights the role of volunteers in contributing observational data on the natural world, coordinated by professional scientists. Both conceptualizations and several analogous ones gave rise to the adjacent notion of citizen science data, more commonly known as data democratization.

Data democratization refers to the process of rendering data accessible and easily usable by all members of an organization, regardless of their role or level of expertise. In principle, it dismantles the traditional barriers that confined data access to specialized departments such as information technology (IT) or data science. The caveat is as follows: by empowering a broader spectrum of individuals with the ability to engage with data, organizations can foster a culture of data-driven decision-making that drives both innovation and operational efficiency.

Viewed from another perspective, data democratization also involves engaging users, not only experts, in the discovery, access, and sharing of data, while maintaining compliance and control. This approach aligns closely with the FAIR principles of data management: ensuring that data is Findable, Accessible, Interoperable, and Reusable (Labadie et al., 2020).

A holistic approach to data democratization discussed extensively in (Džanko et al., 2024) integrates a plethora of socio-material components involving technology, culture, education, and governance to build an efficient and sustainable data ecosystem. Technology provides the necessary tools and infrastructure to facilitate easy access to data and to support its central role in organizational decision-making. At the same time, effective governance remains a critical enabler, ensuring that data democratization is both successful and responsible. Successful democratization also depends on cultural dimensions such as leadership, transparency, and collaboration. Strong leadership sets a clear vision for data-driven practices, encouraging employees to integrate data into their daily operations and maximize its organizational value. Transparency in data practices builds trust by making performance and decision-making processes visible, fostering open communication and confidence in data use. Finally, collaboration across various organizational silos promotes innovation and improves decision-making.

We situate several recommendations put forth in this research report under the Data Maturity Framework using the reverberating theme of Data democratization.

Technopolitical Moral Enhancements

Despite the enormous strides transforming the realms of political action and decision making, the emerging technological world order has been marred with significant limitations either in its ability to ensure fairness and justice, or its ability to reduce the suffering of individuals and communities in a manner that is commensurate with the privileges bestowed upon a few powerful actors or well-developed countries. Increasingly nowadays, technological power appears to outpace our moral psychology: beyond their apparent technosolutionism, the disruptive

sociotechnical tools of our age are acting as a geopolitical resource that we believe necessitates certain moral enhancements, a revisionary process where the ethical and responsible use of emerging technologies is assessed against a backdrop of newly evolving notions of "risk", dubbed astronomical suffering risk. So, what exactly is astronomical suffering risk? And how does it pave the way for reasoned politics, a space for pondering ulterior sociopolitical motives behind which technologies to use and why?

There may be outcomes as a result of conflicts or crises that could be worse than extinction or death, outcomes in which the universe is filled with unimaginable suffering. Those outcomes are known as suffering risks, or s-risks, "where an adverse outcome would bring about severe suffering on an astronomical scale, vastly exceeding all suffering that has existed on Earth so far" (Sotala and Gloor 2017). Just as an individual might consider prolonged torture to be more terrifying than death itself, a civilization might one day face scenarios in which survival lingers with immense and incessant misery. This is referred to as the pan-generational net suffering outcome, when, across all time, the sum total of suffering outweighs happiness for everyone who ever lived.

Technopolitical Mobilisation

The sustained manifestations of memory-based conflicts in the digital sphere serves to obstruct the process of conflict transformation, thereby prolonging a pan-generational net suffering outcome. We argue that for technologies to be deemed "ethical", they must do more than "refrain from causing harm," and be designed with the ultimate aim of reducing astronomical suffering. This is deeply connected to characteristics that Hannah Arendt has identified in "The Human Condition" (Arendt, 1958), which laid out the tripartite distinction between labor, work, and action, thereby providing an epistemology that crystallizes the entanglement of identity, memory, suffering, and action. Labor corresponds to the cyclical rhythms of life, the management of necessity, and the knowledge of repetition. Work belongs to homo faber, the fabrication of durable artifacts, governed by instrumental rationality and calculation. Action alone, in Arendt's view, is the sphere of freedom: unpredictable, relational, and inaugurating new beginnings. In juxtaposition, we argue that incorporating the lens of suffering-based ethics around the design and purpose of modern technologies would enable human action within this sphere of freedom. In the context of memory-based conflicts, this requires the deployment of technologies that are able to detect digital manifestations of intractable disputes, and to extract and mine relevant content that elucidates the root causes behind the prolongation of the conflict, and to reconcile various narratives online in favor of low-power groups.

Automated Speech Detection for Competitive Victimhood

In social science and linguistic research, hate speech is commonly defined as language that attacks, degrades, or vilifies individuals on the basis of social identity, and that carries broader social consequences, such as legitimizing discrimination or harm against target groups. A central feature of hate speech is therefore its group-directed orientation. Following (Ajil, 2022), grievances are not reducible to complaints or expressions of dissatisfaction. Instead, they are best understood as narratively embedded interpretations of injustice that persist across time and space.

Ajil shows that for grievances to endure, they often operate by what he terms a process of decomplexification, whereby interpretation shifts away from immediate, situational factors toward the global, the collective, and the past.

This perspective is highly dependent on the role of collective memory, defined as widely shared knowledge of past social events that may not have been personally experienced, but is nonetheless socially constructed through communicative practices. Such memories function as interpretive resources that structure how injustice is narrated, justified, and reactivated over time. Importantly, Ajil explicitly cautions against a deterministic interpretation of grievances; his findings "do not suggest a straightforward link" between grievance expression and violent engagement

– rather, grievances may be mobilized, instrumentalized, or promoted under certain conditions, while remaining non-violent in others.

A central limitation of automated hate speech detection lies in its decontextualization of speech in ways that render it oblivious to the nuances embedded in historical grievances. Most contemporary speech-classification models are trained on datasets where fragments of text are annotated as isolated textual units, detached from the *longue durée* of denial and memory within which they circulate. Statements that appear neutral or conciliatory at the surface level may, when situated within a specific historical trajectory of injustice, function indirectly to deny or relativize the suffering of others, a distinction in the interpretation of text that current models are unable to capture.

This limitation is not inherent to machine learning itself, but to the sociotechnical choices that govern data annotation and model design. Hate speech datasets overwhelmingly privilege short time horizons, explicit lexical hostility, and individual intent, while neglecting grievance-laden discourse that unfolds through implication and selective acknowledgment. In contexts where certain narratives are already marginalized or contested, this bias is further amplified, constraining what natural language processing models are able to “learn” as being harmful speech. Consequently, long-term trauma and intergenerational suffering remain algorithmically invisible, producing a mode of digital governance that recognizes offense but is oblivious to grievance.

The obsolescence of current systems becomes especially evident amidst manifestations of competitive victimhood. Here, harm is articulated through strategic narrative framing as opposed to explicit hostile language: by foregrounding one group’s suffering while erasing or delegitimizing that of others, this kind of discourse helps to relinquish responsibility through moral absolutism. Detecting such dynamics requires models trained not merely on hate labels, but on grievance-sensitive annotations that encode relational positions and historical narratives. It would also necessitate developing alternative labeling guidelines that distinguish simultaneously among direct hostility, grievance expression, denial, minimization, and victimhood mobilization. More technically speaking, it would also require training objective functions and adapting evaluation metrics that prioritize contextual reasoning over surface-level classification, to move decision boundaries away from the dichotomies embedded in hate speech classification and more in the direction of different labels such as “acknowledgment versus denial”, “empathy versus erasure”, or “historical reference versus strategic silence.”

Counter Narratives using LLM-RAGs and Genocide Denial

Recent work on counter-narrative generation online offers concrete methodological tools that can be repurposed to counter genocide denial, particularly in digital environments where denial circulates through fragmented or decontextualized claims. This requires both structural understanding of denial discourse, using, for example, dedicated Large Language Models (LLMs) that are trained to detect such discourse, combined with retrieval-grounded factual grounding.

LLM-RAGs (Large Language Models with Retrieval-Augmented Generation) are LLM-based systems that integrate retrieval of external knowledge with generative language modeling (Lewis et al., 2020). Technically, they combine the parametric knowledge stored inside a pre-trained language model with non-parametric memory (e.g., external databases, document corpora, or indexed texts) that is accessed at query time. This allows the system to ground its content generation using up-to-date or domain-specific information, reducing reliance on biased training data and improving factual accuracy against historical records. Sociotechnically, LLM-RAGs reshape how users interact with automated knowledge systems: by exercising control over what counts as justified knowledge, they can be seen as agents capable of altering epistemic authority, whilst redistributing responsibility for verifying information between people and machines.

Genocide denial online is a prime example where content moderation using generic condemnations or automatic censorship of speech can be entirely futile. Applied to genocide denial, LLM-RAGs can allow for an automated

intervention on digital platforms that shifts counter-narratives toward historically accountable and epistemically robust interventions (Baez Santamaria et al., 2024; Wilk et al., 2025). For example, because LLM-RAGs can generate their responses by retrieving relevant information from archival records and legal and scholarly documents, they can be adapted to capture how denial narratives systematically link selective historical claims and moral asymmetries between victims and perpetrators, thereby allowing counter-narratives to directly address the internal structure of denial.

Linked Open Data for Collective Memory

Addressing the repercussions of mummified collective memory, where the past is frozen and detached from historical complexity, helps propel societies onto the path of conflict transformation and a reduction in pan-generational astronomical suffering by allowing them to confront ambiguity and inter-group contradictions. Linked Open Data (LOD) is an under-utilized technology in the context of conflict transformation that has been recently gathering increasing importance for the proper archival of multimodal historical records (oral, written, digital, etc.). When LOD technologies are combined with natural language processing capabilities, new opportunities to structure and analyze oral histories arise in ways that align with standard archival practices. An adjacent application appears in several works that explore LOD-based data management systems for the documentation of casualties in World War I and World War II, helping establish robust methodologies for record-keeping and legal frameworks. In the aftermath of World War I and II, records for missing persons were documented primarily through government or military archives and war crime tribunals. Those records were primarily centralized in the form of highly structured data that facilitated cross-referencing and the identification of casualties by consolidating information across official war logs, prisoner-of-war records, and post-war tribunal findings. The WarSampo project (Hyvönen et al., 2016), a digital humanities initiative focused on WWII data, has successfully modernized this structured approach by utilizing LOD to interconnect historical documents into a publicly accessible knowledge system, by scaling it across multiple databases for the establishment of a significantly richer, interlinked historical narrative.

The fragmentation witnessed in the Armenian case presents a stark contrast. Instead of structured archival records, relevant information is scattered across non-governmental organizations, human rights groups, forensic teams, and family-led initiatives, hampering all attempts at large-scale data integration. A WarSampo-inspired approach could address these discrepancies by leveraging LOD to connect oral testimonies, forensic data, and historical accounts. By structuring oral histories into machine-readable formats and linking them with existing legal and forensic records, one could create a more cohesive and accessible repository, thereby bridging the gap between archival documentation techniques and oral history narratives.

Discussion and Conclusion

In this research report, we have argued that memory-based conflicts as exemplified by the Armenian struggle cannot be meaningfully transformed using existing mainstream digital governance paradigms that reduce political violence and historical grievance to the narrow vocabulary of hate speech. Such approaches, when operationalized through automated detection systems with limited temporal and relational depth, risk reproducing a prolonged stagnation in conflict transformation whilst projecting a deluded impression that hate speech detection systems can prevent harm online. By detaching violent or intractable discourse from its historical continuities, digital platforms contribute to the mummification of collective memory, the intensification of competitive victimhood, and the normalization of genocide denial. Those discursive patterns continue to prevail online and are not adequately captured by automated hate speech detection, if at all, which demands moral enhancements and reasoned political responses against the proliferation of hate speech parlance in the context of memory-based conflicts.

Drawing on Arendt's conception of action in *The Human Condition*, we have proposed a shift away from purely regulatory, content-policing models toward sociotechnical interventions that enable the possibility of new beginnings. Conflict transformation in memory-based disputes requires more than the suppression of harmful language; it demands infrastructures that enable acknowledgment, that preserve archival complexity, and that give voices to low-power groups whose suffering is routinely rendered invisible using asymmetrical power in the digital sphere. Ethical technologies, in this sense, must be reoriented toward reducing pan-generational astronomical suffering through practices of recognition and reconciliation.

By embedding long-durée understandings of collective memory and fostering counter-narratives that resist competitive victimhood, sociotechnical systems can move beyond the constraints of hate speech parlance and contribute to conditions more conducive to sustainable peace. Such a reorientation positions technology not as an arbiter of permissible speech, but as an enabler of reasoned politics that can interrupt cycles of denial of suffering and opens pathways toward conflict transformation.

References

- Ajil, A. (2022). Politico-ideological violence in Lebanon: The narrative embeddedness of grievances. *Frontiers in Human Dynamics*, 4, 988999. <https://doi.org/10.3389/fhumd.2022.988999>
- Arendt, H. (1958). *The human condition*. University of Chicago Press.
- Baez Santamaria, S., Gomez Adorno, H., & Markov, I. (2024). *Contextualized graph representations for generating counter-narratives against hate speech*. In Findings of the Association for Computational Linguistics: EMNLP 2024. Association for Computational Linguistics.
- Braudel, F. (1980). *On History*. Translated by S. Matthews. Chicago: University of Chicago Press.
- Hyvönen, E., Heino, E., Leskinen, P., Ikkala, E., Koho, M., Tamper, M., Tuominen, J., & Mäkelä, E. (2016). *WarSampo knowledge graph: A linked data service for digital humanities research on the Second World War*. *Semantic Web*, 7(4), 1–13. <https://doi.org/10.3233/SW-160215>
- Lederach, J. P. (2003). *The Little Book of Conflict Transformation*. Intercourse, PA: Good Books.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Riedel, S. (2020). *Retrieval-augmented generation for knowledge-intensive NLP tasks*. In *Advances in Neural Information Processing Systems*, 33, 9459–9474. <https://arxiv.org/abs/2005.11401>
- Sotala, K., & Gloor, L. (2017). Superintelligence as a Cause or Cure for Risks of Astronomical Suffering. *Informatica*, 41, 501–505.
- Wilk, B., Shomee, H. H., Maity, S. K., & Medya, S. (2025). *Fact-based counter narrative generation to combat hate speech*. In *Proceedings of the Web Conference 2025* (pp. 3354–3365). ACM.



Issam Fares Institute for Public Policy & International Affairs
American University of Beirut P.O. Box 11-0236 Riad El-Solh /
Beirut 1107 2020 Lebanon



961-1-350000 ext. 4150



+961-1-737627



ifi.comms@aub.edu.lb



www.aub.edu.lb/ifi/



[Issam Fares Institute for Public Policy &
International Affairs IFI](#)



[aub.ifi](#)



[@ifi_aub](#)



[@ifi_aub](#)



**ISSAM FARES INSTITUTE FOR PUBLIC
POLICY & INTERNATIONAL AFFAIRS**

معهد عصام فارس للسياسات العامة والشؤون الدولية

AMERICAN UNIVERSITY OF BEIRUT