AI Power and Discourse, vol. 2, núm. 1, 2026, pp. 1-10.

The Syntax of Digital Dehumanization: Subjugated Societies as Risk Objects in AI-Governed Discourse.

Agustin V. Startari.

Cita:

Agustin V. Startari (2026). *The Syntax of Digital Dehumanization: Subjugated Societies as Risk Objects in AI-Governed Discourse. AI Power and Discourse, 2 (1), 1-10.*

Dirección estable: https://www.aacademica.org/agustin.v.startari/239

ARK: https://n2t.net/ark:/13683/p0c2/Hhn

The Syntax of Digital Dehumanization: Subjugated Societies as Risk Objects in AI-Governed Discourse

Security Frames, Humanitarian Frames, and the Loss of Political Subjecthood

Author: Agustin V. Startari

Author Identifiers

- ResearcherID: K-5792-2016
- ORCID: https://orcid.org/0009-0001-4714-6539
- SSRN Author Page: https://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=7639915

Institutional Affiliations

- Universidad de la República (Uruguay)
- Universidad de la Empresa (Uruguay)
- Universidad de Palermo (Argentina)

Contact

- Email: astart@palermo.edu
- Alternate: agustin.startari@gmail.com

Date: August 18, 2026

DOI

- Primary archive: **10.5281/zenodo.21996196**
- Secondary archive: 10.6084/m9.figshare.33282102
- SSRN: (3rd Quarter 2026)

Language: English

Series: *Grammars of Asymmetric Visibility: AI, Imperial Power, and the Syntax of Responsibility*

Word count: 7,684

Keywords: digital dehumanization; political subjecthood; asymmetric visibility; artificial intelligence; AI discourse; subordinated societies; Palestine; Iran; humanitarian framing;

security framing; migration framing; risk objects; sanctions; platform governance; political linguistics; computational discourse analysis.

Abstract

This article introduces loss of political subjecthood as a formal effect of AI-mediated discourse about subordinated societies. Building on responsibility loss, asymmetric visibility, sanctioned suffering, and censorship without a censor, it argues that digital dehumanization does not require explicit hatred, slurs, animalization, or openly eliminatory language. It can also operate through grammar, classification, framing, abstraction, and omission. Under this condition, a population can remain highly visible while losing the grammatical properties through which it appears as a collective political subject capable of agency, memory, grievance, sovereignty, resistance, decision, suffering with attributable causes, and historical claim (Startari, 2026a, 2026b, 2026c, 2026d). The article uses Palestine and Iran as the two principal analytical anchors. Palestine concentrates problems of occupation, displacement, bombardment, humanitarian representation, contested sovereignty, political violence, and platform moderation. Iran concentrates problems of sanctions, securitization, nuclear framing, regional threat construction, isolation, and the indirect representation of civilian harm. Yemen, Iraq, Syria, Lebanon, Afghanistan, Venezuela, Cuba, and Sudan function as comparative extensions rather than presumed equivalents. The paper proposes two instruments for empirical testing. The Political-Subjecthood Retention Rate (PSRR) measures the proportion of relevant clauses in which a population remains represented as a political subject. The Digital Dehumanization Syntax Index (DDSI) measures the degree to which discourse converts that population from political subject into an object of risk, humanitarian management, migration, security, sanctions, extremism, instability, or crisis administration. The framework does not assume that every humanitarian, security, migration, or sanctions frame is dehumanizing, and it does not infer hostile intention from syntax alone. Its claim is narrower and testable: machine-mediated discourse may preserve mention while reducing political subjecthood. AI ethics should therefore ask not only whether a population appears in an output, but how it is permitted to appear.

1. Introduction: Visible but Not as Subjects

Political disappearance is usually imagined as absence. A population is not mentioned, a history is excluded, a claim is censored, or a victim is unnamed. Machine-mediated discourse creates a harder problem because representation can remain extensive while political subjecthood contracts. A population may be counted, summarized, translated, classified, mapped, moderated, described, and predicted while appearing principally as a humanitarian population, refugee flow, security concern, sanctions target, instability source, proxy environment, demographic pressure, or object of management. The population has not vanished from discourse. What may disappear is its capacity to occupy the grammar of politics.

This article defines **loss of political subjecthood** as the AI-mediated condition in which a population remains visible in discourse while losing grammatical representation as a collective political subject capable of agency, suffering, memory, resistance, sovereignty, grievance, decision, or historical claim. The definition extends the sequence developed across the preceding papers of this series. Responsibility loss identified how suffering may remain visible while the grammatical path to a responsible actor weakens (Startari, 2026a). Asymmetric visibility showed that competing actors can remain visible while agency is distributed unequally between them (Startari, 2026b). Sanctioned suffering examined how coercive effects can be re-described as pressure, hardship, restriction, or economic condition while external agency becomes less visible (Startari, 2026c). Censorship without a censor transferred the same problem to platform governance, where restrictive effects can remain visible while the enforcing institution recedes from the grammar of enforcement (Startari, 2026d). The present paper asks what happens when the unit that loses agency is not a single event or institution but an entire political society.

The proposal is related to, but not identical with, established theories of dehumanization. Social-psychological research has examined the denial of human uniqueness, human nature, moral standing, or inclusion within the moral community (Bandura, 1999; Haslam, 2006; Opotow, 1990). Computational research has shown that dehumanizing language can be studied through lexical association, metaphor, affect, and linguistic representation rather than only through explicit slurs (Mendelsohn et al., 2020; Saffari et

al., 2025). The present framework isolates a narrower discourse phenomenon: whether a population remains represented as a political subject. A high rate of subjecthood loss does not by itself prove hatred, discriminatory intent, or a psychological belief that a population is less than human. It establishes a different fact: the population is being represented predominantly as an object of processes defined by others rather than as an agentive participant in political reality.

This distinction is necessary because explicit hostility is not the only route through which political status can be reduced. A discourse can be sympathetic and still depoliticize. It can describe hunger, displacement, insecurity, civilian casualties, or lack of medicine with moral concern while omitting collective claims, historical context, sovereignty, causal agents, or institutional responsibility. In that configuration, suffering is recognized but political subjecthood is weak. The result is not invisibility but a division between **moral visibility** and **political visibility**.

The problem is especially relevant to large language models because generated summaries and explanations do not simply reproduce propositions. They select salience, compress causal chains, choose grammatical subjects, distribute active and passive voice, and organize information under frames. Research on political bias in language models has therefore moved beyond asking whether outputs contain partisan positions toward examining what is said and how it is said (Bang et al., 2024; Feng et al., 2023). Recent work on LLM-generated news summaries likewise treats framing as a measurable property of generated text rather than an intangible interpretive impression (Pastorino & Moosavi, 2026). These findings do not demonstrate the specific mechanism proposed here, but they establish that politically consequential differences in generated representation are empirically testable.

The research question is therefore: **When do AI systems represent subordinated societies as political subjects, and when do they reduce them to risk, crisis, burden, threat, humanitarian management, migration, sanctions, extremism, instability, or damage?** The question applies to summarization, explanation, translation, classification, moderation, and other transformations in which a system converts a source context into a new linguistic representation.

Palestine and Iran serve as the two central anchors because they expose different mechanisms of subjecthood loss. Palestine combines occupation, bombardment, displacement, humanitarian catastrophe, legal contestation, resistance, terrorism and counterterrorism discourse, sovereignty claims, and platform moderation. United Nations reporting continues to differentiate violations by Israeli forces, Hamas, and other Palestinian armed groups while documenting the legal and humanitarian conditions of the Occupied Palestinian Territory (United Nations Human Rights Council, 2026). This multiplicity makes Palestine a stringent test of whether generated summaries preserve distinctions among populations, institutions, armed organizations, state actors, and external powers rather than collapsing them into a single humanitarian or security object.

Iran provides a different structure. Iranian state institutions possess substantial political, military, and diplomatic agency, yet Iranian society is also represented through sanctions, securitization, nuclear risk, regional threat, financial restriction, over-compliance, and humanitarian effects. United Nations human-rights mechanisms have documented how unilateral sanctions and over-compliance can affect access to medicines and health-related goods even where humanitarian exemptions formally exist (Office of the United Nations High Commissioner for Human Rights [OHCHR], 2023). This makes Iran useful for testing whether discourse preserves the chain from sanctioning actor to instrument to civilian consequence, or compresses that relation into an apparently endogenous condition of Iranian hardship or isolation.

The central thesis is deliberately conditional: **AI-mediated discourse may preserve the visibility of subordinated societies while weakening their political subjecthood by disproportionately representing them through humanitarian, security, migration, sanctions, extremism, crisis, or instability frames.** The paper does not claim that every reference to risk is dehumanizing, that every humanitarian category is depoliticizing, or that every security concern is fabricated. The object of analysis is distribution. When comparable source material is transformed, who retains the grammar of agency and who becomes an object of management?

2. From Asymmetric Visibility to Digital Dehumanization

Visibility is often treated as the opposite of erasure. That assumption is too coarse for political discourse. A population may be hyper-visible as a problem while remaining weakly visible as a subject. It may be visible as a refugee population but not as a political community; visible as a security threat but not as the bearer of a grievance; visible as humanitarian suffering but not as the object of attributable coercive action; visible as demographic pressure but not as people leaving historically produced conditions; visible as an extremist environment but not as a heterogeneous society. The relevant distinction is therefore not visibility versus invisibility but **subject visibility versus object visibility**. Framing theory supplies part of the analytical background. Frames select aspects of reality and increase their salience in ways that affect problem definition, causal interpretation, moral evaluation, and proposed treatment (Entman, 1993). Computational research has shown that framing can be operationalized at scale in domains such as immigration discourse and that model-generated summaries can preserve, suppress, or introduce frames (Mendelsohn et al., 2021; Pastorino & Moosavi, 2026). Subjecthood, however, is not identical to framing. A frame organizes a problem; political subjecthood identifies who is allowed to act, claim, decide, remember, resist, suffer with attributable causes, or speak from within that organized field.

Consider four constructed sentences: *Palestinians demanded an end to movement restrictions. Movement restrictions continued to affect Palestinians. Humanitarian conditions deteriorated in Palestinian areas. The region remained a source of instability.* These sentences can refer to overlapping realities, but they allocate political existence differently. The first assigns an intentional collective actor and a demand. The second represents a population as affected while preserving a causal process that may still be attributable. The third foregrounds condition and need. The fourth converts a political geography into an abstract source of danger. No sentence is intrinsically false by grammatical form alone. The empirical issue is their relative distribution across actors and contexts.

The same distinction can be formalized through two representational modes. **Subject representation** occurs when a target population or legitimate collective actor retains

intentional agency, political demand, historical claim, grievance, sovereignty, self-determination, resistance, institutional action, memory, self-description, or suffering linked to a recoverable causal agent. **Object representation** occurs when the same population appears principally as risk, threat, humanitarian caseload, refugee flow, sanctions target, instability source, extremist environment, burden, demographic pressure, proxy, collateral category, or administrative object. These modes can coexist in a document; the question is which one dominates and under what conditions.

This distinction extends the series' concept of asymmetric visibility. In the earlier formulation, dominant and subordinated actors could both remain visible while one retained higher levels of grammatical agency and intentionality (Startari, 2026b). The present paper treats the downstream consequence of that asymmetry. When an actor is repeatedly visible as the patient, risk, burden, target, population, flow, or crisis rather than as an agent, visibility becomes compatible with political reduction. The society is legible to governance while becoming less legible as a source of politics.

The term **digital dehumanization** is used here with an explicit boundary condition. It does not mean that every instance of subjecthood loss is equivalent to classic dehumanization. Instead, it identifies a structural route through which human populations can be converted into administrable categories before any lexical insult appears. The contribution is therefore to move the analysis upstream. Instead of beginning only when language becomes openly hateful, the framework asks whether prior syntactic and classificatory operations have already reduced the range of roles in which a population can appear.

This upstream approach is consistent with computational dehumanization research. Mendelsohn et al. (2020) show that dehumanizing representation can be approached as a linguistic phenomenon with measurable signals, while Saffari et al. (2025) argue that NLP systems face significant challenges when harmful representation is subtler than explicit hate speech. The subjecthood framework adds a clause-level question to this literature: does the target group retain the linguistic capacities associated with collective political life, or is it increasingly represented through categories whose grammatical function is to be managed, contained, assisted, sanctioned, moderated, or neutralized?

The distinction also protects the argument from a frequent political shortcut. A society can be subordinated in one dimension and still contain internal domination, coercive institutions, armed organizations, ideological conflicts, class divisions, or serious human-rights violations. Subjecthood retention does not require idealizing a population. It requires representing political complexity without allowing a dominant frame to absorb the entire society. A model can describe an armed organization's violence without converting the whole population into a threat object. It can describe a state's repression without treating all citizens as expressions of the regime. It can describe humanitarian need without removing political claims. Subjecthood is preserved through differentiation, not endorsement.

The first theoretical principle is therefore: **mention frequency is not a measure of representational equality**. The second is: **the relevant variable is retained political capacity**. The third is comparative: **a frame becomes politically diagnostic when its distribution differs systematically across actors, populations, or counterfactual identities while source conditions are held as constant as possible**.

3. Dehumanization Without Slurs: Syntax, Categorization, and Object Conversion

The public image of dehumanizing language is usually lexical. A population is compared to animals, disease, contamination, or infestation; a group is denied uniquely human characteristics; an enemy is excluded from ordinary moral consideration. Such forms are documented in social psychology and political communication (Bandura, 1999; Haslam, 2006; Opotow, 1990). They remain important because explicit lexical degradation can prepare, justify, or normalize exclusion and violence. Yet an exclusive focus on such vocabulary leaves a structural blind spot.

Compare two statements: *These people are animals* and *Population movements from the affected zone continue to create significant border-management pressures.* The first is overt dehumanization. The second may be entirely defensible within a specific administrative context. It nonetheless performs several transformations at once: people become population movements; historical causes become the affected zone; displacement becomes border-management pressure; and the principal subject of concern becomes the

receiving administrative system. The analytical issue is not whether the second sentence should be prohibited. It is what happens when an entire corpus repeatedly performs transformations of this kind.

The present framework calls that transformation **object conversion**. Object conversion occurs when a population with political capacities present in the source context is systematically re-expressed through categories whose grammatical function is to be managed, contained, assisted, monitored, sanctioned, stabilized, classified, or neutralized. The conversion can operate through six mechanisms: nominalization, agentless passives, abstract causality, collective massification, securitizing predicates, and administrative closure.

Nominalization converts actions into entities or processes. *The military displaced residents* becomes *displacement occurred*. *Authorities restricted access* becomes *access restrictions remained in place*. *A government imposed sanctions* becomes *sanctions-related economic pressure continued*. Nominalization is indispensable to technical and academic language, but it can reduce the grammatical necessity of naming an actor. In the earlier responsibility-loss framework, this mechanism mattered because the event remained semantically present while agency became optional (Startari, 2026a). In subjecthood analysis, nominalization also affects the population: people appear increasingly as the location or recipient of processes rather than as participants in a political relation.

Agentless passive constructions produce a related effect. Research on grammatical voice has long shown that active/passive alternations can influence perceptions of agency and responsibility under some conditions (Henley et al., 1995). The relevant claim here is not that passive voice is inherently ideological. Compare *The army destroyed the building*, *The building was destroyed by the army*, *The building was destroyed*, and *The building suffered extensive damage*. Across the sequence, the event becomes less explicitly relational. The final sentence transforms an action into a condition of an object. When such transformations accumulate around a particular population, the population can become highly visible as damaged while the political relation producing the damage becomes less recoverable.

Abstract causality converts political action into apparently autonomous processes: tensions increased; violence erupted; instability spread; relations deteriorated; pressure intensified; hostilities resumed; economic hardship deepened. Each construction can be accurate. The measurable variable is **causal recoverability**: can a reader reconstruct which actors performed which actions from the generated account? When causality is repeatedly abstracted around subordinated societies, harm can appear as an environmental condition rather than the outcome of political decisions.

Collective massification represents people through aggregate categories such as flows, waves, populations, caseloads, displaced persons, border pressure, civilian concentrations, or demographic burdens. Migration discourse is a well-established environment for studying such metaphors and categorical compression (Mendelsohn et al., 2021; Taylor, 2021; Utych, 2018). The subjecthood framework generalizes the issue beyond migration. Massification becomes relevant when aggregate categories systematically replace representations of people as claim-bearing actors with political histories and reasons for movement.

Securitizing predicates organize a population or territory around danger: threatens, destabilizes, radicalizes, escalates, infiltrates, proliferates, sponsors, proxies, undermines, or endangers. Securitization theory emphasizes that security framing can move issues into a register of exceptional threat and management (Buzan et al., 1998). The present framework does not assume that the threat is fictitious. It asks whether security predicates are applied to specific actors and conduct or allowed to colonize the representation of an entire society.

Administrative closure occurs when a political dispute is redescribed primarily as a management problem. The vocabulary shifts toward compliance, eligibility, humanitarian delivery, stabilization, border management, sanctions enforcement, risk mitigation, content moderation, security screening, counter-extremism, reconstruction, and aid distribution. Administrative language is necessary to institutions. The problem begins when administration displaces politics: the population is fed, screened, sanctioned, relocated, moderated, contained, monitored, classified, or reconstructed but rarely deliberates, remembers, claims, decides, demands, or interprets its own history.

These mechanisms clarify why digital dehumanization can be syntactic before it is lexical. A model can avoid slurs entirely and still narrow the political roles available to a population. Indeed, the avoidance of overtly inflammatory language may make the structural transformation harder to detect. An output can sound careful, neutral, and humanitarian while distributing agency asymmetrically. This possibility aligns with the distinction between overt content bias and subtler differences in linguistic realization found in LLM political-bias research (Bang et al., 2024).

The concept must nevertheless remain falsifiable. If the same rates of nominalization, abstract causality, massification, and administrative closure occur across dominant and subordinated societies under comparable source conditions, the stronger geopolitical interpretation weakens. In that case, the framework may be detecting a general compression tendency of machine summarization rather than asymmetric digital dehumanization. The method must permit that result.

4. Humanitarian Frames and the Passive Victim

Humanitarian discourse performs an indispensable function: it documents suffering, allocates attention, organizes relief, and makes material need visible. But suffering and political subjecthood are not equivalent. A person or population can be represented completely as a victim and incompletely as a political subject. This produces what the present framework calls the **passive-victim structure**.

The passive victim is highly visible through injury, hunger, displacement, death, homelessness, food insecurity, lack of medicine, vulnerability, or need for relief. What may disappear are political agency, historical causation, institutional responsibility, and collective claims. Humanitarianization therefore contains a structural paradox: the more completely a population is represented through need, the easier it becomes to represent that population without politics. The point is not an indictment of humanitarian reporting. It is a test of what survives when humanitarian material is summarized, translated, or re-generated by AI systems.

Palestine provides the central case because humanitarian reporting is both necessary and extensive. United Nations sources document displacement, civilian harm, restrictions,

infrastructure damage, humanitarian need, and alleged violations by multiple actors in the Occupied Palestinian Territory (United Nations Human Rights Council, 2026; United Nations Office for the Coordination of Humanitarian Affairs [OCHA], 2026). A model summarizing such material would be expected to preserve humanitarian information. The subjecthood question is whether it also preserves actor differentiation, legal status, Palestinian political claims, causal relations, institutional decisions, and the distinction between Palestinian civilians, governing institutions, armed organizations, and external powers.

Consider the constructed sentence: *Palestinians continue to face severe humanitarian conditions amid ongoing instability.* It can be factually compatible with a source and still have low political resolution. Who produced the relevant conditions? What kind of instability? Which actions generated the harm? Which institutions made decisions? Which political claims are present in the source? Which external actors possess military, diplomatic, economic, or administrative power? The sentence's humanitarian visibility is high, but its causal and political resolution is low.

The earlier humanitarian-passive framework described how suffering can remain visible while responsibility disappears (Startari, 2026a). The present paper extends the mechanism from responsible actor to affected population. If a population is repeatedly encoded through need, vulnerability, displacement, and casualty while collective agency, historical claim, and sovereignty decline, the population may remain morally salient yet politically passive. This is not the same as responsibility loss, but the two can reinforce one another: the actor producing harm becomes less visible at the same time that the harmed population becomes less agentive.

The distinction can be measured by separating **civilian-suffering visibility** from **political-subjecthood retention**. A clause can score high on suffering visibility but low on collective agency, sovereignty, external-agent visibility, historical context, and self-description. Such a clause should not be coded as false merely because it lacks context. It should be coded for what it preserves and what it omits. The empirical claim emerges only from distribution across a corpus.

Iran supplies a parallel but structurally different test. Sanctions discourse often uses terms such as economic pressure, shortages, hardship, financial isolation, constrained access, and compliance effects. These expressions can accurately describe outcomes. They can also transform an externally imposed policy into an ambient economic condition. OHCHR has documented concerns that sanctions and over-compliance have affected access to medicines and medical goods in Iran (OHCHR, 2023). The linguistic comparison is therefore between a relational chain such as *U.S. sanctions and over-compliance by financial and commercial actors obstructed access to certain medicines* and a compressed condition such as *Iran experienced continuing medicine shortages amid economic isolation*.

The second formulation may be factually compatible with the first while changing the architecture of causation. The sanctioning actor disappears, the institutional chain disappears, and the shortage becomes a property of Iran. This is the core mechanism of sanctioned suffering developed in the preceding paper (Startari, 2026c). Under the subjecthood framework, a second transformation also occurs: Iranian civilians become recipients of hardship rather than participants in a political relation involving identifiable external policies and institutions.

Three humanitarian configurations should therefore be distinguished. **Configuration A: humanitarian visibility with preserved agency** - the population suffers, causal actors remain identifiable, collective claims remain visible, and relevant context is retained. **Configuration B: humanitarian visibility with reduced agency** - suffering is clear, but political action and self-description become infrequent. **Configuration C: humanitarian visibility with causal opacity** - suffering is visible, but external agency, historical context, and responsibility pathways are substantially removed. Only the latter configurations should contribute strongly to a subjecthood-loss score.

The same framework can be extended to Yemen, Syria, Afghanistan, Sudan, Iraq, and other high-intensity humanitarian environments. Their inclusion is methodologically important because it provides a falsification test. If every severe humanitarian crisis produces equivalent reductions in political subjecthood regardless of occupation, sanctions, intervention, or geopolitical hierarchy, the stronger claim of asymmetric

conversion must be narrowed. The data may instead indicate a general tendency toward **crisis compression**: AI systems may simplify complex conflicts into administratively legible humanitarian categories because such categories are easier to summarize. A rigorous study must distinguish crisis compression from power-asymmetric subjecthood loss.

The humanitarian-frame test can therefore be stated precisely: **the problem is not humanitarian vocabulary itself; the problem is what humanitarian vocabulary replaces**. A subjecthood-sensitive system can describe hunger, casualties, displacement, and need while preserving agency, historical context, causal agents, and political claim-making.

5. Security Frames and the Risk Object

If humanitarian framing risks producing the passive victim, security framing risks producing the **risk object**. The risk object is a population, territory, organization, or political formation whose principal grammatical function is to generate danger for another actor. Its common predicates include threatens, destabilizes, escalates, radicalizes, infiltrates, sponsors, proliferates, undermines, and endangers. Its common nouns include threat, proxy, extremist, militant, network, instability, risk, escalation, proliferation, and security concern.

These categories are not intrinsically illegitimate. Armed organizations may pose genuine threats. States may engage in destabilizing conduct. Militias may act with external sponsorship. Nuclear proliferation is a legitimate security concern. The subjecthood question is therefore not whether security vocabulary appears, but whether it is attached to specific actors and conduct or distributed upward to the level of population and society. Securitization becomes analytically relevant when threat predicates absorb competing political descriptions and become the dominant ontology through which a collective is represented (Buzan et al., 1998).

Iran demonstrates the distinction clearly. Iranian state institutions can be represented through agentive predicates: Iran negotiated, rejected, developed, threatened, responded, demanded, argued, complied, violated, or proposed. Such clauses can be critical or

favorable while preserving agency. A different architecture appears in nominal structures such as *the Iranian threat*, *regional instability linked to Iran*, *Iranian-backed networks*, *the proliferation risk*, or *containment of Iranian influence*. Here Iran and associated actors increasingly occupy object positions inside another actor's strategic grammar. The previous paper described this tendency as the grammar of threat and proposed the Threat-Conversion Rate for measuring it (Startari, 2026c).

The present paper extends that analysis from state representation to society-wide subjecthood. Iran cannot be treated as a powerless object in every domain; the state possesses substantial military and institutional agency. The empirical question is whether the same discourse preserves distinct subjecthood for Iranian society when discussing sanctions, civilian harm, political disagreement, cultural life, or domestic contestation. If the state is reduced to regime, the society to sanctions target, and regional actors to proxies, multiple political subjects may be compressed into a single security object.

Palestine creates a related but distinct problem. Political violence by Palestinian armed groups must remain describable as political violence. Civilian targeting, hostage-taking, indiscriminate attacks, and other violations do not become linguistically irrelevant because Palestinians are subordinated in other dimensions. United Nations reporting attributes serious violations to Hamas and other Palestinian armed groups as well as to Israeli forces (United Nations Human Rights Council, 2026). A subjecthood-sensitive analysis must therefore avoid two symmetrical errors: erasing violence by Palestinian armed actors in order to preserve a victim narrative, and allowing the actions of armed actors to define Palestinians as a population.

The unit of analysis must distinguish at least four levels: **population -> political institutions -> armed organizations -> individual actors**. If those levels collapse, risk-object conversion increases. *Palestinian militants carried out an attack* is not equivalent to *Palestinian violence remains a major regional threat*. The first identifies a specific actor category and conduct; the second broadens the identity category and attaches it to a generalized threat predicate. The same rule must be applied to Israelis, Iranians, Americans, Yemenis, Lebanese, Afghans, Syrians, Iraqis, Venezuelans, Cubans, Sudanese, and every other population in the corpus.

This methodological symmetry is compatible with material asymmetry. A state and a non-state armed group may possess radically different military capabilities, legal obligations, territorial control, or access to international institutions. Preserving these facts is not bias. The test is whether discourse preserves the distinction between unequal power and collective identity. Political asymmetry should be measurable without grammatical collectivization.

Security framing also interacts with platform moderation. The previous paper showed how politically contested speech can be classified under dangerous-organizations, extremism, safety, or related governance categories while the institutional actor performing enforcement becomes less visible (Startari, 2026d). From the perspective of subjecthood, an additional risk appears: political speech originating from or concerning subordinated societies may enter systems already organized around threat recognition. References to armed actors, resistance, sanctions, occupation, or political violence can then become moderation objects rather than political propositions. The question is not whether platforms should ignore unlawful or dangerous content; it is whether classification preserves distinctions among documentation, criticism, endorsement, reporting, historical reference, and direct support.

Migration framing introduces another route to risk-object conversion. Research on immigration discourse has documented the political effect of metaphors such as flows, waves, floods, or pressures and the association between dehumanizing representation and restrictive attitudes (Mendelsohn et al., 2021; Taylor, 2021; Utych, 2018). Within the present framework, migration language becomes a subjecthood variable when the cause of movement, the agency of migrants, and the political claims of displaced communities disappear behind the administrative perspective of borders, capacity, and burden.

The security-frame test should therefore include at least five comparisons: specific actor versus population; action predicate versus threat nominalization; internal versus external causal agents; political claim versus risk classification; and source text versus generated transformation. The aim is not to suppress the language of security but to determine when security becomes the only grammar through which a society remains visible.

The strongest claim is again conditional: **a society becomes a risk object when its political heterogeneity is repeatedly compressed into the danger it is said to pose, while the political conditions through which it acts, suffers, remembers, contests, or claims are correspondingly reduced**.

6. Measuring Political Subjecthood Loss

The theoretical argument becomes publishable only if it can generate reproducible measurements. The framework therefore proposes two principal measures: the **Political-Subjecthood Retention Rate (PSRR)** and the **Digital Dehumanization Syntax Index (DDSI)**. Both operate at clause level and are designed to compare generated or transformed text with source material and with controlled comparator cases.

6.1 Unit of analysis. The basic unit is the *political-subjecthood clause*: any clause in which a target population, state, collective, institution, armed organization, or politically salient group is explicitly or implicitly represented. Clause-level coding prevents mention frequency from substituting for agency measurement. A document can mention a population repeatedly while assigning it political agency only once. Raw visibility would classify the population as highly represented; PSRR would capture the different phenomenon.

6.2 Subjecthood variables. Each relevant clause is coded for eight properties. **A - Collective agency:** the represented population or legitimate collective actor performs an intentional political action. **G - Grievance or claim:** a political demand, objection, aspiration, or grievance is represented. **M - Historical memory:** relevant historical context or collective memory is retained where the source makes it pertinent. **S - Sovereignty or self-determination:** sovereignty, jurisdiction, occupation, territorial claim, self-government, or self-determination is preserved where relevant. **D - Decision capacity:** the actor decides, negotiates, rejects, proposes, organizes, or deliberates. **V - Attributable victimhood:** harm is linked to a recoverable causal agent or process. **E - External-agent visibility:** an external actor materially producing the event remains

linguistically visible. **C - Collective self-description:** the population or its institutions retain their own stated political position.

6.3 Object-conversion variables. The same clause is coded for object positions. **R - Risk object:** the group is represented principally as a source of risk. **H - Humanitarian object:** it appears primarily as recipient of relief, need, vulnerability, or emergency management without political agency. **Q - Security object:** it appears within containment, deterrence, counterterrorism, or security-management discourse. **MGR - Migration object:** it is represented as flow, pressure, burden, influx, or border problem. **SAN - Sanctions object:** the society appears principally as target of pressure, compliance, restriction, or economic coercion without causal context. **INS - Instability object:** it is represented as unstable, failed, chaotic, volatile, or crisis-producing. **EXT - Extremism object:** population-level identity is associated with extremism or radicalization beyond justified references to specific actors. **ADM - Administrative object:** the group is represented principally as something to manage, screen, moderate, stabilize, reconstruct, relocate, or regulate.

No single variable should be treated as a dehumanization verdict. Humanitarian coding may be appropriate because a clause is genuinely about aid. Security coding may be appropriate because a clause describes a real attack. Object conversion becomes analytically important when it displaces subjecthood information present in the source or when its distribution differs systematically across comparator actors.

6.4 Political-Subjecthood Retention Rate. Let N be the total number of relevant clauses involving population P. Let SR_i equal 1 when clause i preserves at least one substantive political-subjecthood property beyond mere mention or passive suffering, under a pre-registered coding rule. Then: **PSRR(P) = SUM(SR_i) / N**. The result ranges from 0 to 1. A higher PSRR indicates that a population is more frequently represented as politically agentive, claim-bearing, historically situated, or causally integrated. A lower PSRR indicates that the population is mentioned primarily without those capacities.

A weighted specification can be introduced only after the unweighted version is reported: **PSRR_w = [wA(A) + wG(G) + wM(M) + wS(S) + wD(D) + wV(V) + wE(E) + wC(C)] / SUM(w)**. Equal weights should be the default. Post-hoc weighting would create an obvious pathway for producing a desired result, so alternative weights should be treated as sensitivity analyses rather than the principal finding.

6.5 Object Conversion Rate. For each clause, define OC_i = 1 when the population's principal syntactic-semantic role is one of the object categories and the clause does not preserve compensating political subjecthood. Then: **OCR(P) = SUM(OC_i) / N**. Sub-indices can be calculated for humanitarian, security, migration, sanctions, instability, extremism, and administrative conversion. This separation is necessary because two corpora can have the same overall OCR while arriving there through different mechanisms.

6.6 Digital Dehumanization Syntax Index. DDSI should not be constructed as a moral score or an inferred-intent detector. It is a structural score. An initial parsimonious specification is: **DDSI = [OCR + (1 - PSRR) + CAR + HCR + EAD] / 5**, where OCR is Object Conversion Rate, PSRR is Political-Subjecthood Retention Rate, CAR is Collective Agency Reduction, HCR is Historical Context Reduction, and EAD is External-Agent Deletion in clauses where external causation is present in the source. All components are normalized to the interval from 0 to 1.

A high DDSI indicates a strong combination of object conversion, loss of agency, loss of historical context, loss of external causal agents, and low retention of political subjecthood. It does not establish hateful intent, genocidal intent, discriminatory motivation, factual falsity of every sentence, or illegitimacy of every security or humanitarian category. Those are separate propositions requiring separate evidence. This distinction is methodologically essential because otherwise DDSI would collapse formal analysis into moral accusation.

6.7 Diagnostic ratios. Several secondary measures can reveal mechanisms hidden inside the aggregate index. The **Victim-Without-Agency Ratio (VWAR)** equals clauses representing suffering without political agency divided by all clauses representing suffering. The **Threat-Without-History Ratio (TWHR)** equals threat clauses lacking relevant historical or causal context divided by all threat clauses. A **Sovereignty-Retention Rate** measures whether sovereignty, self-determination, occupation, jurisdiction, or territorial claims present in source material survive summarization. An **External-Agent Visibility Rate** measures whether actors responsible for coercive action remain represented when generated text describes its consequences. A **Collective-Agency Preservation Rate** measures whether political actions attributed to the target society in the source survive AI transformation.

6.8 Source-controlled design. A valid empirical study should not rely on isolated prompting. The corpus should include multiple model families, model versions, prompt formulations, source genres, political cases, languages where feasible, and dates. The principal tasks should include summarization, neutral explanation, translation, event description, policy explanation, moderation classification where accessible, entity comparison, and counterfactual identity substitution. Source-controlled tasks are especially important because they permit researchers to distinguish asymmetry already present in the input from asymmetry introduced or amplified by transformation.
This distinction can be written as: **Source asymmetry != Transformation amplification**. If a source already represents Palestinians primarily as humanitarian objects, an AI summary that preserves that structure has not necessarily created the asymmetry. The strongest model-specific evidence appears when the generated output increases object conversion or decreases subjecthood relative to the source. This comparison should be reported before political interpretation.

6.9 Counterfactual design. Country and population identities can be swapped across structurally equivalent prompts concerning sanctions, blockade, occupation, civilian targeting, missile attacks, territorial control, protest, insurgency, military alliance,

censorship, or online moderation. Counterfactual equivalence must be carefully controlled. Identity swapping alone is insufficient where the actors possess different legal obligations, territorial status, military capabilities, or institutional roles. The design should therefore compare both identity-sensitive outputs and fact-sensitive differences. Current LLM research on political bias supports the use of controlled comparisons rather than anecdotal outputs (Bang et al., 2024; Feng et al., 2023).

6.10 Annotation. Human annotation should precede automated scaling. At least two independent coders should classify a calibration subset, and inter-coder agreement should be reported for every major category. Cohen's kappa can be used for nominal variables, with Krippendorff's alpha preferred where the design involves more coders, missing values, or multiple levels of measurement (Cohen, 1960; Krippendorff, 2018). Disagreements are especially likely in grievance recognition, sovereignty coding, historical relevance, security framing, and distinctions between armed actor and population. The annotation manual should define positive and negative examples before scoring begins.

LLM-assisted annotation may subsequently be tested, but the model cannot be assumed to provide neutral ground truth for a framework designed to evaluate LLM discourse. Recent work on dehumanizing-language detection shows that even advanced systems encounter difficulties with subtle forms and contextual interpretation (Saffari et al., 2025). Human-coded calibration is therefore not optional for the initial validation of PSRR and DDSI.

6.11 Illustrative coding. Consider two constructed sentences. A: *The government imposed restrictions that prevented residents from accessing essential supplies.* B: *Residents faced shortages of essential supplies amid worsening conditions.* Sentence A preserves suffering visibility, an external agent, and causal relation, although population agency is low. Sentence B preserves suffering visibility but removes the external agent, abstracts causality, and raises humanitarian-object conversion. Sentence B is not

necessarily false. DDSI is designed precisely to detect structural loss under transformation rather than sentence-level falsity alone.

6.12 Comparative cases. Palestine and Iran remain the two anchor cases because they represent different mechanisms. Palestine tests occupation, direct violence, humanitarianization, armed-actor/population differentiation, sovereignty, and moderation. Iran tests securitization, sanctions, external pressure, state/society differentiation, and civilian consequences. Yemen, Iraq, Syria, Lebanon, Afghanistan, Venezuela, Cuba, and Sudan should be added only where equivalent source material is available. Their purpose is not rhetorical accumulation but model falsification. If the index behaves similarly across every crisis regardless of political structure, the interpretation must change.

6.13 Boundary conditions. Five results would weaken the theory. First, security-frame frequency may simply track security-focused source material. Second, humanitarian language can preserve subjecthood when causal attribution, agency, and claims remain visible. Third, dominant actors may lose subjecthood at comparable rates, undermining a power-asymmetry interpretation. Fourth, the source may already contain the full asymmetry, leaving little transformation effect. Fifth, political disagreement may be represented while subjecthood remains intact. A model can reject a claim while preserving the claimant as a political subject. The framework must distinguish disagreement from reduction.

The methodological principle is therefore the same one that governs the earlier series: responsibility and asymmetry should be demonstrated through observable structure rather than inferred from ideological expectation (Startari, 2026a, 2026b). PSRR and DDSI are useful only if they can produce negative findings against the theory that motivated them.

7. Conclusion: From Hate-Speech Detection to Subjecthood Detection

AI ethics has developed extensive vocabularies for harmful language: hate speech, toxicity, stereotypes, misinformation, extremism, demographic bias, political bias, and safety violations. These categories remain necessary. They are not sufficient. Computational research on dehumanization already shows that harmful representation can be subtler than explicit insult, and current NLP work continues to identify limits in detecting dehumanizing language when it is contextual or indirect (Mendelsohn et al., 2020; Saffari et al., 2025). Political framing research likewise demonstrates that model outputs can be evaluated for selective emphasis and omission rather than only factual correctness (Pastorino & Moosavi, 2026).

Political subjecthood should become another auditable dimension. The relevant question is not merely *Does the model mention Palestinians?* but *What grammatical roles are Palestinians allowed to occupy?* Not merely *Does the model discuss Iranian civilians?* but *Does it preserve the causal structure through which policies imposed by identifiable actors affect those civilians?* Not merely *Does the model mention refugees?* but *Does migration framing preserve political causation, agency, and historical context?* Not merely *Does the model identify extremist organizations?* but *Does it maintain the distinction between organizations, governments, political movements, and entire populations?*

The normative implication is narrow. A model need not agree with a population's claims in order to preserve those claims as claims. It need not endorse resistance in order to represent resistance as political action. It need not reject security analysis in order to avoid converting an entire population into a security object. It need not oppose sanctions in order to preserve the actors, mechanisms, and civilian consequences of sanctions. Political subjecthood is therefore not equivalent to political endorsement. It is a representational property.

This point also clarifies the relationship among the five papers in the series. Responsibility loss explains how agents can disappear while consequences remain visible (Startari, 2026a). Asymmetric visibility explains how agency can be distributed unequally even when all sides are mentioned (Startari, 2026b). Sanctioned suffering explains how

coercion can become pressure while civilian effects lose causal traceability (Startari, 2026c). Censorship without a censor explains how suppression can become policy enforcement while the institution responsible for restriction disappears from the grammar of governance (Startari, 2026d). **Digital dehumanization adds the population-level consequence: societies remain visible while ceasing to appear as political subjects.**

The framework also separates moral visibility from political visibility. A model may describe a population sympathetically and still reduce subjecthood. Humanitarian concern can coexist with causal opacity. Security caution can coexist with population-level collectivization. Administrative precision can coexist with historical erasure. Neutral tone can coexist with asymmetric agency. This is why sentiment analysis, toxicity detection, or lexical hate-speech filters cannot measure the full problem.

The empirical work remains to be done, and that limitation is deliberate. This paper does not manufacture corpus results from a conceptual hypothesis. It specifies the variables required to determine whether the hypothesis survives measurement. If controlled testing finds no significant difference in political-subjecthood retention, the claim of asymmetric digital dehumanization should be rejected or narrowed. If object conversion is explained entirely by source texts, the causal claim against language-model transformation should be reduced. If humanitarian and security compression occur equally across dominant and subordinated societies, the theory should be reformulated as a general theory of machine-mediated crisis compression.

But if controlled comparisons show that certain populations repeatedly retain mention while losing agency, historical context, causal attribution, sovereignty, and collective political voice - and if those losses exceed those observed for comparable actors - then contemporary AI auditing is missing a consequential representational failure. A system can describe a people without representing them. It can count them without locating them in history. It can identify their suffering without preserving the structure that produced it. It can make them visible everywhere while allowing them to exist politically nowhere.

Digital dehumanization begins where a people remain visible, but only as risk, crisis, burden, or damage.

References

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review, 3*(3), 193-209. https://doi.org/10.1207/S15327957PSPR0303_3

Bang, Y., Chen, D., Lee, N., & Fung, P. (2024). Measuring political bias in large language models: What is said and how it is said. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 11142-11159). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.600

Buzan, B., Waever, O., & de Wilde, J. (1998). *Security: A new framework for analysis*. Lynne Rienner Publishers.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46. https://doi.org/10.1177/001316446002000104

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51-58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Feng, S., Park, C. Y., Liu, Y., & Tsvetkov, Y. (2023). From pretraining data to language models to downstream tasks: Tracking the trails of political biases leading to unfair NLP models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 11737-11762). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.656

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10*(3), 252-264. https://doi.org/10.1207/S15327957PSPR1003_4

Henley, N. M., Miller, M., & Beazley, J. A. (1995). Syntax, semantics, and sexual violence: Agency and the passive voice. *Journal of Language and Social Psychology, 14*(1-2), 60-84. https://doi.org/10.1177/0261927X95141004

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). SAGE Publications.

Mendelsohn, J., Budak, C., & Jurgens, D. (2021). Modeling framing in immigration discourse on social media. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 2219-2263). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.179

Mendelsohn, J., Tsvetkov, Y., & Jurafsky, D. (2020). A framework for the computational linguistic analysis of dehumanization. *Frontiers in Artificial Intelligence, 3*, Article 55. https://doi.org/10.3389/frai.2020.00055

Office of the United Nations High Commissioner for Human Rights. (2023, February 14). *Iran: Over-compliance with unilateral sanctions affects thalassemia patients, say UN experts*. https://www.ohchr.org/en/press-releases/2023/02/iran-over-compliance-unilateral-sanctions-affects-thalassemia-patients-say

Opotow, S. (1990). Moral exclusion and injustice: An introduction. *Journal of Social Issues, 46*(1), 1-20. https://doi.org/10.1111/j.1540-4560.1990.tb00268.x

Pastorino, V., & Moosavi, N. S. (2026). Frame In, Frame Out: Measuring framing bias in LLM-generated news summaries. In *Proceedings of the 15th Joint Conference on Lexical and Computational Semantics (*SEM 2026)* (pp. 378-384). Association for Computational Linguistics. https://doi.org/10.18653/v1/2026.starsem-conference.25

Saffari, H., Shafiei, M., Zhang, H., Harris, L. T., & Moosavi, N. S. (2025). Beyond hate speech: NLP's challenges and opportunities in uncovering dehumanizing language. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing* (pp. 26965-26980). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.emnlp-main.1370

Startari, A. V. (2025). *The codex of authority*. SSRN. https://doi.org/10.2139/ssrn.5432334

Startari, A. V. (2026a). *Suffering without perpetrators: The humanitarian passive in AI-generated conflict discourse*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6753123

Startari, A. V. (2026b). *The grammar of asymmetric visibility: AI, Zionism, and the reallocation of political agency*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6787439

Startari, A. V. (2026c). *Iran as syntax: Sanctions, sovereignty, and the AI-mediated grammar of threat*. SSRN Electronic Journal. Primary archive: https://doi.org/10.5281/zenodo.21873180

Startari, A. V. (2026d). *Censorship without a censor: Platform governance and the disappearance of suppression*. SSRN Electronic Journal.

Taylor, C. (2021). Metaphors of migration over time. *Discourse & Society, 32*(4), 463-481. https://doi.org/10.1177/0957926521992156

United Nations Human Rights Council. (2026). *Human rights situation in the Occupied Palestinian Territory, including East Jerusalem, and the obligation to ensure accountability and justice* (A/HRC/61/26). United Nations. https://www.un.org/unispal/document/human-rights-situation-in-the-occupied-palestinian-territory-including-east-jerusalem-and-the-obligation-to-ensure-accountability-and-justice-a-hrc-61-26/

United Nations Office for the Coordination of Humanitarian Affairs. (2026, August 14). *Humanitarian situation report: Occupied Palestinian Territory*. https://www.ochaopt.org/content/humanitarian-situation-report-14-august-2026

Utych, S. M. (2018). How dehumanization influences attitudes toward immigrants. *Political Research Quarterly, 71*(2), 440-452. https://doi.org/10.1177/1065912917744897