



## Original-Forschungsarbeit

# Semantische Souveränität im Zeitalter der künstlichen Intelligenz: Die persische Sprache, Bedeutung und kulturelle Selbstbestimmung

Mohsen Karami<sup>1</sup>

*1 Assistenzprofessor für Medienkunst, Abteilung für Religionen und Medien, IRIB-Universität, Teheran, Iran*

Empfangen: 2. Februar 2025 Akzeptiert: 5. Juni 2025

### Zusammenfassung:

Die rasche Verbreitung großer Sprachmodelle und textgenerierender Systeme hat nicht nur einen technologischen Wandel ausgelöst, sondern auch eine epistemische Umstrukturierung der Art und Weise, wie Bedeutung erzeugt und zirkuliert wird. Diese Arbeit diagnostiziert ein spezifisches Risiko für das Persische: die Abschwächung und potenzielle Verdrängung seines kulturell-semantischen Horizonts innerhalb globalisierter, vorwiegend englischsprachiger KI-Infrastrukturen. Ziel der Untersuchung ist analytisch und diagnostisch: die begrifflichen Grundlagen der „semantischen Souveränität“ herauszuarbeiten und die strukturellen Wege aufzuzeigen, auf denen zeitgenössische KI-Praktiken persische Bedeutungen, Metaphern und hermeneutische Traditionen gefährden. Die Studie verbindet begrifflich-philosophische Analyse (Sprachphilosophie, Hermeneutik, Phänomenologie) mit einer kritischen Lektüre aktueller KI-Trainingsregime und Datenökologien. Sie bedient sich einer analytischen Begriffssynthese statt empirischer Intervention: Die Untersuchung verfolgt theoretische Prämissen (Wittgensteins „Bedeutung als Gebrauch“, Gadammersche Horizonte, Davidsonsche Triangulation, Floridis Informationsethik) und kartiert sie auf die materiellen Praktiken der Datensatz-Kuration, des Modell-Trainings und der Plattform-Vermittlung. Die Arbeit identifiziert mehrere sich gegenseitig verstärkende Mechanismen, durch die KI-Systeme semantische Asymmetrie erzeugen: Korpus-Bias und Repräsentationsknappheit; algorithmische Übersetzung, die nicht-englische semantische Netze in englisch-dominante Vektorräume umstrukturiert; infrastrukturelle Vermittlung, die persische kulturelle Artefakte zu reinen Datenpunkten ohne ihren hermeneutischen Kontext degradiert; sowie epistemische Filterung durch Empfehlungs- und Retrieval-Systeme, die bestimmte Formen von Explizierbarkeit gegenüber Opazität und Singularität privilegieren. Insgesamt verkörpern diese Mechanismen, was ich als „phänomenologische Auslöschung“ der welterschließenden Kraft einer Sprache bezeichne. Das Phänomen, um das es geht, ist nicht bloßer lexikalischer Verlust, sondern eine ontologische Verarmung: eine Verengung der Fähigkeit des Persischen, eigentümliche Weisen des Seins zu erschließen. Die Anerkennung dieses Risikos verlangt begriffliche Klarheit über semantische Souveränität als diagnostische Kategorie. Die vorliegende Arbeit verzichtet bewusst auf die Formulierung remedierender Politiken; stattdessen zielt sie darauf ab, eine strenge philosophische Inszenierung des Problems zu bieten, damit nachfolgende Forschung und öffentliche Diskurse die Tiefe, Modalitäten und Einsätze der semantischen Gefährdung des Persischen angemessen einschätzen können.

**Schlüsselwörter:** algorithmischer kolonialismus, künstliche Intelligenz, kulturelle Selbstbestimmung, persische Sprache, phänomenologische Auslöschung, semantische Souveränität

\* Korrespondierender Autor

✉ [mohsenkarami@iribu.ac.ir](mailto:mohsenkarami@iribu.ac.ir)

🌐 <https://orcid.org/0000-0003-3669-1029>

Wie dieser Artikel zu zitieren ist:

Karami, M. (2025). Semantic sovereignty in the age of artificial intelligence: The Persian language, meaning, and cultural self-determination. *Spektrum Iran*, 38(2), 31-60.

🔗 <https://doi.org/10.22034/spektrum.2026.556925.1044>

© Copyright © Der/die Autor(en); Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitungen 4.0 International (CC-BY-NC) Lizenz. Homepage: [www.spektrumiran.com](http://www.spektrumiran.com)

## حاکمیت معنایی در عصر هوش مصنوعی: زبان فارسی، معنا، و خودتعیین‌بخشی فرهنگی

محسن کرمی<sup>۱</sup>

۱ استادیار، گروه هنرهای رسانه‌ای، دانشکده دین و رسانه، دانشگاه صداوسیما، تهران، ایران

دریافت: ۱۴۰۳/۱۲/۲۳؛ پذیرش: ۱۴۰۴/۰۳/۱۵

### چکیده:

رشد شتابان الگوهای زبانی بزرگ و سامانه‌های متن‌زا نه فقط تحولی فناورانه بلکه یک دگرگونی معرفت‌شناختی در شیوه تولید و گردش معنا پدید آورده است. این مقاله به‌جای‌شناسی یکی از خطرات پیش‌روی زبان فارسی می‌پردازد: تحلیل‌رفتن و احتمالاً جابه‌جاشدن افق فرهنگی — معنایی آن در درون زیرساخت‌های جهانی‌شده هوش مصنوعی که غالباً مبتنی بر زبان انگلیسی‌اند. هدف مقاله نه تجویزی بلکه تحلیلی و تشخیصی است: تبیین بنیان‌های مفهومی «حاکمیت معنایی» و ترسیم مسیرهای ساختاری که رویه‌های معاصر هوش مصنوعی از طریق آن‌ها معانی، استعاره‌ها، و سنت‌های هرمنوتیکی فارسی را در معرض خطر قرار می‌دهند. پژوهش حاضر تحلیل مفهومی — فلسفی (در قلمرو فلسفه زبان، هرمنوتیک، و پدیدارشناسی) را با قرائتی انتقادی از نظام‌های آموزشی هوش مصنوعی و بوم‌های داده‌ای کنونی در هم می‌آمیزد. روش آن ترکیبی است از سنجش تحلیلی مفاهیم، نه مداخله تجربی: بدین معنا که با پی‌گیری پیش‌فرض‌های نظری — از معنای وینگنشتاینی به‌منزله کاربرد، افق‌های هرمنوتیکی گادامری، مثلث‌سازی دیویدسونی، و اخلاق اطلاعات فلوریدی — نشان می‌دهد که چگونه این چارچوب‌ها در سطح رویه‌های مادی گردآوری داده، آموزش مدل، و میانجی‌گری پلتفرمی تبلور می‌یابند. مقاله چند سازوکار هم‌بنیاد و تقویت‌کننده را در نحوه تولید نابرابری معنایی توسط سامانه‌های هوش مصنوعی بازمی‌نماید: سوگیری در پیکره‌های زبانی و فقر بازنمایی؛ ترجمه الگوریتمی‌ای که شبکه‌های معنایی زبان‌های غیرانگلیسی را در فضای برداری انگلیسی محور بازساخت می‌کند؛ میانجی‌گری زیرساختی‌ای که آثار فرهنگی فارسی را به داده‌هایی منفصل از زمینه هرمنوتیکی‌شان بدل می‌سازد؛ و پالایش‌های معرفت‌شناختی‌ای که در نظام‌های توصیه‌گر و بازیابی محتوا شکل‌های خاصی از وضوح را بر ابهام و یگانگی ترجیح می‌دهند. این سازوکارها در مجموع مصداق آن چیزی‌اند که نویسنده از آن با عنوان «انقراض پدیدارشناختی توان جهان‌گشایی زبان» یاد می‌کند. پدیده مورد بحث صرفاً از دست‌رفتن واژگان نیست، بلکه نحیف‌شدنی هستی‌شناختی است: محدودشدن ظرفیت زبان فارسی برای گشوده‌بودن به شیوه‌های خاص بودن. آگاهی از این خطر، نیازمند روشنی مفهومی در باب «حاکمیت معنایی» به‌منزله مقوله‌ای تشخیصی است. این مقاله از ارائه راهکار یا سیاست اصلاحی خودداری می‌کند؛ مقصود آن فراهم‌آوردن صحنه‌ای فلسفی و منسجم برای طرح مسئله است تا پژوهش‌ها و گفت‌وگوهای بعدی بتوانند ژرفا، سازوکارها، و مخاطرات در کمین معنای فارسی را با دقتی بیشتر بسنجند.

**واژگان کلیدی:** حاکمیت معنایی، هوش مصنوعی، زبان فارسی، استعمار الگوریتمی، انقراض پدیدارشناختی، خودتعیین‌بخشی فرهنگی

\* نویسنده مسئول

<https://orcid.org/0000-0003-3669-1029>

[mohsenkarami@iribu.ac.ir](mailto:mohsenkarami@iribu.ac.ir)

<https://doi.org/10.22034/spektrum.2026.556925.1044>



Original Research Paper

# Semantic sovereignty in the age of artificial intelligence: The Persian language, meaning, and cultural self-determination

Mohsen Karami<sup>1</sup>

*1 Assistant Professor, Media Arts, Department of Religions and Media, IRIB University, Tehran, Iran*

Received: Feb., 12, 2025 Accepted: Jun. 5, 2025

## Abstract

The rapid proliferation of large language models and text-generative systems has precipitated not only technological transformation but also an epistemic reconfiguration of how meaning is produced and circulated. This paper diagnoses a specific risk facing Persian: the attenuation and potential displacement of its cultural-semantic horizon within globalized, predominantly English-language AI infrastructures. The objective is both analytic and diagnostic: to delineate the conceptual grounds of 'semantic sovereignty' and to map the structural pathways through which contemporary AI practices endanger Persian meanings, metaphors, and hermeneutic traditions. The study combines conceptual-philosophical analysis (philosophy of language, hermeneutics, phenomenology) with a critical reading of current AI training regimes and data ecologies. It employs analytic conceptual synthesis rather than empirical intervention: the analysis traces theoretical presuppositions (Wittgensteinian 'meaning as use', Gadamerian horizons, Davidsonian triangulation, Floridi's information ethics) and maps them onto the material practices of dataset curation, model training, and platform mediation. The paper identifies multiple, mutually reinforcing mechanisms by which AI systems produce semantic asymmetry: corpus bias and representational scarcity; algorithmic translation that restructures non-English semantic networks into English-dominant vector spaces; infrastructural mediation that repositions Persian cultural artifacts as data points divorced from their hermeneutic contexts; and epistemic filtering enacted by recommender and retrieval systems that privilege certain forms of explicability over opacity and singularity. Collectively, these mechanisms instantiate what I term the 'phenomenological extinction' of a language's world-disclosing power. The phenomenon at stake is not mere lexical loss but an ontological impoverishment: a contraction of Persian's capacity to disclose distinctive modes of being. Recognizing this risk requires conceptual clarity about semantic sovereignty as a diagnostic category. This paper stops short of prescribing remedial policies; instead, it aims to provide a rigorous philosophical staging of the problem so that subsequent scholarship and public discourse can assess the depth, modalities, and stakes of Persian's semantic endangerment.

**Keywords:** colonialism, Artificial intelligence, Cultural self-determination, Persian language, Phenomenological extinction, Semantic sovereignty

---

\* Corresponding Author

✉ [mohsenkarami@iribu.ac.ir](mailto:mohsenkarami@iribu.ac.ir)

🌐 <https://orcid.org/0000-0003-3669-1029>

*How to Cite this Article:*

Karami, M. (2025). Semantic Sovereignty in the Age of Artificial Intelligence: The Persian Language, Meaning, and Cultural Self-Determination. *Spektrum Iran*, 38(2), 31-60.

🔗 <https://doi.org/10.22034/spektrum.2026.556925.1044>

© Copyright © The Author(s); This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC-BY-NC) License. Homepage: [www.spektrumiran.com](http://www.spektrumiran.com)

## **1. Introduction**

Over the past decade, artificial intelligence has ceased to be a discrete field of technological innovation and has instead become an ontological environment—an encompassing horizon within which linguistic activity, communication, and even thought now unfold (Floridi, 2014; Crawford, 2021). What once counted as “textual mediation” has transformed into computational prefiguration: before words reach us, they are already filtered, predicted, and optimized by algorithmic systems that learn from immense corpora (Bender et al., 2021). Such systems do not merely extend the range of human expression; they pre-structure the conditions under which expression itself is possible (Heidegger, 1977, pp. 19–23). Consequently, we must ask not simply how Persian is represented in these architectures but what becomes of Persian as a mode of disclosure when its linguistic lifeworld is subsumed by infrastructures that think and calculate in English (Apter, 2013).

This paper emerges from the conviction that the crisis facing Persian in the age of AI is not reducible to the familiar concern about “digital neglect” or the underrepresentation of non-English languages online (Phillipson, 1992). These are symptoms, not causes. The deeper issue is the reconfiguration of the semantic order itself—how meaning comes to exist, circulate, and gain authority in a computationally governed world (Foucault, 1970; Couldry & Mejias, 2019). Drawing on Iranian scholarship on cultural diversity, such as models for intercultural communication that emphasize structural unity alongside pluralism (Nourbakhsh, 2009), the crisis of semantic sovereignty highlights how AI infrastructures risk undermining Persian cultural self-determination. Persian, with its thousand-year continuum of poetic, philosophical, and mystical expression, embodies a model of meaning that is associative, allegorical, and often deliberately indeterminate (Nasr, 2007; Aminrazavi, 1997). Its linguistic vitality depends on ambiguity, allusion, and metaphorical excess. In contrast, the architectures of modern AI valorize disambiguation, predictability, and semantic flattening (Bender et al., 2021). They reward clarity over resonance, explicitness over implication, the literal over the symbolic. The conflict between these two regimes of meaning—poetic polysemy and statistical predictability—is not incidental but structural (Mohamed, Png, & Isaac, 2020). It is precisely here that the notion of semantic sovereignty becomes indispensable as a philosophical category (Povinelli, 2021).

The entry of Persian into machine-readable form has always been mediated by translation—first into Latin-based transcription systems, later into Unicode, and now into the embedding spaces of large language models (Apter, 2013; Brown et al., 2020). As explored in the historical functions of translation in Iranian culture (e.g., Vahid Dastjerdi & Haddadian Moghaddam, 2015), translation acts as both an inner cultural trait and a necessity for intercultural communication, highlighting the risks of algorithmic mediation in displacing Persian hermeneutic traditions. Each layer of mediation brings with it a subtle realignment of meaning. For example, when Persian poetic imagery such as “*noor dar del*” (“light in the heart”) is vectorized for machine translation, the phrase’s mystical ontology—where “*noor*” denotes divine manifestation rather than physical illumination—tends to be suppressed in favor of literal equivalence (Aminrazavi, 1997, pp. 87–90). What survives is a referential shell, not the experiential depth the phrase originally embodied. Thus, Persian texts in AI systems risk becoming simulacra: data points stripped of their metaphysical resonance, their ambiguity transmuted into computational noise (Benjamin, 2008).

By framing the problem diagnostically, this paper refrains from policy advocacy or linguistic preservationism in the narrow sense. The goal is to disclose, in conceptual precision, the forms of loss that occur when Persian meaning is algorithmically reconstituted (Tlostanova & Mignolo, 2012). Philosophical inquiry, in this context, functions like a diagnostic imaging device: it reveals invisible fractures in the infrastructure of understanding before they manifest as irreversible cultural damage (Crawford, 2021, pp. 46–52). Only when we see how meaning is being operationalized—what is rewarded, what is ignored—can we recognize that the very ontology of Persian as a meaning-world is under pressure (Floridi, 2011).

The focus, therefore, is not on saving Persian as an artifact but on understanding what it means for a language to lose its world-disclosing power (Heidegger, 1971). If Heidegger were right that “language is the house of Being” (1971, p. 145), then the large-scale reorganization of linguistic practices by AI amounts to a redesign of that house’s architecture (Stiegler, 2010). The Persian house of Being—built on poetry, allegory, and ambiguity—is being remodeled under a foreign engineering logic that prizes efficiency over dwelling (Illich, 1993; Povinelli, 2021). The question is not

whether Persian will continue to be spoken, but whether its speakers will still inhabit the world in Persian. This is the crisis of semantic sovereignty (Mohamed et al., 2020).

Having established the existential and ontological scope of the problem, the next section unfolds the theoretical foundations that underpin this diagnostic enterprise. The analysis will draw from Wittgenstein, Gadamer, Davidson, and Floridi to articulate how meaning's fragility under computational regimes exposes the deeper structure of semantic dependency that defines Persian's predicament in the age of artificial intelligence.

## **2. Materials and Methods**

### **2.1. Conceptual and Philosophical Methodology**

The present study employs a diagnostic philosophical approach rather than an empirical or computational one. It situates the phenomenon of meaning loss within the interpretive traditions of philosophy of language and phenomenology, examining the structural presuppositions that allow linguistic meaning to emerge, stabilize, and eventually erode.

Rather than gathering data or applying quantitative metrics, the method consists in tracing conceptual relations –between meaning, usage, and world– as they are reconfigured by algorithmic infrastructures. Through this lens, Persian becomes a paradigmatic case for understanding how linguistic worlds can be reshaped by epistemologies produced by computational systems.

In line with Wittgenstein, Gadamer, Davidson, and Floridi, the paper adopts an *analytical-interpretive* framework: it maps philosophical conceptions of meaning onto the technical operations of language models, dataset architectures, and representational pipelines. This philosophical mapping constitutes both the “materials” (concepts, texts, traditions) and the “methods” (hermeneutic, linguistic, phenomenological analysis) of the inquiry.

While this paper relies principally on conceptual and diagnostic philosophical analysis, that choice is intentional and methodologically defensible: the claims concern structural reconfigurations of meaning

(ontologies of sense, hermeneutic horizons, and procedural norms) that must first be mapped at the level of conceptual possibility before any empirical operationalization. Empirical and computational studies (corpus analysis, retrieval experiments, user studies) could and should complement this diagnosis by quantifying the indicators proposed below; however, such empirical work presupposes the conceptual distinctions this paper aims to clarify, so it is presented here as a necessary preliminary rather than a substitute.

## **2.2. Theoretical Foundations: The Fragility of Meaning under Computational Regimes**

Wittgenstein's dictum that "the meaning of a word is its use in the language" (1953, §43) entails more than a pragmatic turn in linguistics; it discloses that meaning is inseparable from the lived practices through which words gain sense (Kripke, 1982; Baker & Hacker, 2009). Meaning is not an abstract object but an enacted rule within a *form of life*. In Persian, this insight is vividly illustrated by idiomatic expressions whose significance depends on ritual, gesture, and shared historical consciousness – phrases such as "*dastash barkat darad*" ("his hand carries blessing"), which fuses moral evaluation, spiritual ontology, and communal acknowledgment in a single utterance. When such expressions are modeled by large language systems, their semantic vitality is reduced to patterns of statistical co-occurrence (Bender et al., 2021). What remains is surface correlation without the pragmatic horizon that makes the phrase meaningful.

In computational regimes, "use" is simulated through frequency, not through participation in communal action. Fluency becomes a proxy for understanding (Marcus & Davis, 2020). Yet as Wittgenstein's analysis reminds us, use is never merely distributional – it is normative, embodied, and historically sedimented (Wittgenstein, 1953; Cavell, 1979). The philosophical error at the heart of current language modeling lies in treating these contingent practices as if they could be exhaustively encoded in probabilistic relations. A Persian language model might "learn" to reproduce formulaic piety or literary ornament, but it cannot inhabit the *form of life* in which those utterances are situated (Kripke, 1982). The result is a simulacrum of sense: plausible strings devoid of shared commitment.

This divergence is crucial for our diagnostic project. When meaning is redefined as pattern, languages with smaller digital corpora like Persian are doubly disadvantaged: not only are they statistically underrepresented, but the very criterion of meaning—pattern frequency—privileges corpora that already conform to computational regularity (Mohamed, Png, & Isaac, 2020). The untranslatable, the elliptical, the mystical—all become algorithmic anomalies. The form of life shrinks to fit the grid.

### **2.3. Hermeneutic Horizons (Gadamer)**

Gadamer's hermeneutic philosophy (1975/2004) deepens the Wittgensteinian emphasis on practice by introducing the notion of *horizon*: the historically mediated field of understanding within which interpretation occurs. Understanding, for Gadamer, is never a mechanical act of decoding; it is a fusion of horizons, where the interpreter's pre-understandings encounter those embedded in the text (Warnke, 1987). This fusion is fragile and dialogical—it demands openness to otherness and recognition of one's finitude.

AI-driven interpretation, however, operates through pre-trained parameters, not openness. The "horizon" of a model is fixed by its training data and its optimization objectives (Crawford, 2021). Even fine-tuning remains a form of internal recalibration, not genuine encounter. In such systems, the hermeneutic dialogue is replaced by algorithmic closure: the model does not *listen*; it computes. When Persian poetry or prose is ingested into this regime, the fusion of horizons becomes unilateral—the text must conform to the model's preexisting semantic vectors, which are overwhelmingly shaped by English-language rationalist and positivist traditions (Apter, 2013; Couldry & Mejias, 2019).

Consider the Sufi metaphor of "*sama*" (spiritual listening). In Persian thought, this term oscillates between literal, musical, and mystical registers; it refers simultaneously to an act of hearing and to an inner attunement to divine rhythm (Nasr, 2007). In algorithmic translation, however, "*sama*" typically becomes "music performance" or "listening session," stripped of transcendental reference. What is lost is not a nuance but a world. The hermeneutic horizon collapses into a single dimension of lexical equivalence (Benjamin, 2008). The result is *semantic domestication*: Persian texts become

intelligible only by being recoded into the conceptual syntax of another language (Apter, 2013).

This process exemplifies what Gadamer warned against—the illusion of transparency. The apparent ease with which AI systems “understand” Persian conceals a profound hermeneutic violence: the replacement of dialogical openness with mechanical fit. The fusion of horizons becomes an absorption of one by the other (Gadamer, 1975/2004, p. 305).

#### **2.4. Triangulation and Intersubjectivity (Davidson)**

Donald Davidson’s principle of triangulation (1967, 1984) holds that meaning arises not from isolated symbol systems but from interactions among speakers who jointly refer to a shared world. Understanding presupposes mutual recognition of belief, intention, and situational context (Malpas, 1992). Language models, by design, lack all three. They neither possess beliefs nor participate in shared environments; they manipulate symbols in a vacuum of agency (Marcus & Davis, 2020). What is lost, therefore, is the very condition of truthfulness that Davidson took to be intrinsic to language use.

In the Persian context, this absence manifests sharply. Much of Persian discourse—ethical counsel, mystical exhortation, philosophical dialogue—depends on the *sincerity* (*ikhlas*) of the speaker and the *trust* (*amn*) of the listener. These are not peripheral sentiments; they are epistemic conditions for meaning (Nasr, 2007). When a model reproduces phrases like “*be Khodā qasam*” (“by God, I swear”) or “*del-e man gavāhi midahad*” (“my heart testifies”), it performs without believing. The utterance loses its ontological ground in shared intention. Triangulation fails; what remains is syntactic mimicry.

Philosophically, this failure transforms language into a detached code, divorced from the world it names (Davidson, 1984). The absence of lived reference renders every generated statement a kind of ghost-speech—grammatically alive, ontologically hollow. Davidson’s insight thus extends beyond epistemology into ethics: without triangulated sincerity, communication collapses into simulation. Persian meaning, born of relational sincerity, is therefore peculiarly vulnerable to computational parody.

#### **2.5. Information Ethics and the Ontology of Data (Floridi)**

Luciano Floridi’s philosophy of information (2011, 2014) offers a bridge between these linguistic concerns and the infrastructural ethics of the digital

age. For Floridi, information environments are *ontological regions*—they do not merely convey content but shape the way reality is accessed and understood. To exist in the infosphere is to participate in an environment where informational norms define what counts as evidence, reliability, and relevance.

In the case of Persian, the infosphere's dominant norms derive from English-language epistemic habits: quantifiability, explicitness, reproducibility (Couldry & Mejias, 2019; Crawford, 2021). These norms shape not only what kinds of statements are prioritized by algorithms but also what kinds of *thought* are rendered legible. Ambiguity, allegory, or mystical opacity—once considered virtues of Persian eloquence—are algorithmically penalized as noise or irrelevance. Thus, Persian discourse finds itself epistemically marginalized within a space that values what can be computed over what can be contemplated.

Floridi's framework also reveals the ethical inversion at play: while computational systems claim neutrality, their architecture embodies specific axiologies (Floridi, 2011). The privileging of data abundance over interpretive richness is itself a moral stance—a preference for scale over depth. In practical terms, the dominance of Anglocentric datasets creates an implicit *epistemic colonialism*: languages like Persian must justify their modes of expression according to foreign standards of clarity and efficiency (Mohamed et al., 2020; Phillipson, 1992).

This ethical reconfiguration marks a new kind of imperialism—not territorial but informational. It is not that Persian is excluded from the infosphere; it is included under conditions that compromise its semantic autonomy. Semantic sovereignty, therefore, is not a nostalgic plea for purity but a philosophical defense of plural ontologies of sense (Povinelli, 2021).

## **2.6. Synthesis of Theoretical Positions**

Taken together, these theoretical frames—Wittgenstein's use, Gadamer's horizon, Davidson's triangulation, and Floridi's infosphere—delineate the conceptual terrain on which the question of semantic sovereignty must be understood. Meaning is neither a statistical pattern nor a property of isolated texts; it is an emergent relation within a living practice (Wittgenstein, 1953; Gadamer, 1975/2004). When that practice is automated, sampled, and

reassembled within alien epistemic infrastructures, what disappears is not vocabulary but *worldhood*—the texture of existence that a language discloses (Heidegger, 1971; Floridi, 2014).

Persian, like all historically deep languages, functions as an archive of experiences, emotions, and metaphysical intuitions sedimented across centuries (Nasr, 2007). To lose the capacity to *mean* in Persian is to lose access to those archived possibilities of being. The theoretical foundation of this paper thus leads directly to its diagnostic aim: to identify how computational infrastructures jeopardize this capacity, not by accident but by design (Crawford, 2021; Mohamed et al., 2020).

## **2.7. Conceptualizing Semantic Sovereignty: A Diagnostic Framework**

The expression *semantic sovereignty* designates, in this context, not a juridical claim but an ontological right: the right of a linguistic community to generate, circulate, and authenticate meaning on its own terms within an increasingly automated epistemic world (Mohamed, Png, & Isaac, 2020; Povinelli, 2021). Sovereignty here is less about ownership than about *agency*—the capacity to sustain distinctive modes of world-disclosure without subordination to alien metrics of sense (Floridi, 2014). When a language's conditions of intelligibility are outsourced to computational infrastructures that operationalize meaning through optimization, that language forfeits more than cultural autonomy: it risks a transformation of its cognitive ecology (Crawford, 2021; Couldry & Mejias, 2019).

To diagnose this risk, semantic sovereignty may be analyzed across four interdependent dimensions—each offering a lens through which the predicament of Persian under AI mediation can be understood.

### **2.7.1. Representational Density: The Corpus as Ontological Archive**

Representational density concerns the *breadth and depth* with which a language's genres, registers, and historical strata are present within digital corpora (Phillipson, 1992; Bender et al., 2021). For Persian, density is not merely quantitative (the number of tokens) but qualitative: whether the corpus embodies the full spectrum of expressive modalities that constitute Persian thought—classical ghazal, mystical treatise, modernist prose, colloquial dialogue, religious commentary, and philosophical discourse (Nasr, 2007; Aminrazavi, 1997).

In practice, Persian's digital corpus remains unevenly weighted toward modern journalistic and social media registers. The consequence is an impoverished semantic sample: the language of fleeting commentary replaces the language of reflection. When models are trained predominantly on such data, the resulting "Persian" they learn is statistically representative yet existentially thin—a dialect of immediacy severed from the metaphysical and ethical depths that have historically sustained Persian expression (Mohamed et al., 2020).

Representational density, then, functions as a diagnostic of semantic vitality. A low-density corpus is a sign not only of digital neglect but of ontological malnutrition: the reduction of a living language to a flattened vocabulary of topical convenience (Phillipson, 1992). Each unrepresented genre, each unencoded idiom, diminishes the horizon of what the model can mean, and thereby narrows the world accessible to its users (Crawford, 2021).

### ***2.7.2. Hermeneutic Integrity: The Continuity of Contexts***

Hermeneutic integrity refers to the preservation of the interpretive frameworks that enable understanding beyond literal sense (Gadamer, 1975/2004; Ricoeur, 1976). Every language sustains a network of commentarial, exegetical, and performative contexts—forms of reading that carry its internal logic. Persian's classical canon, for instance, is inseparable from its interpretive traditions: Hafez without his divinatory readings, Rumi without his Sufi metaphysics, Saadi without his moral pedagogy, or Khayyam without his philosophical irony, are mutilated versions of themselves (Nasr, 2007).

Digital mediation threatens this integrity by isolating texts from their commentarial ecosystems (Benjamin, 2008). When a verse of Rumi enters a machine-learning dataset without its historical annotations, it becomes a decontextualized sentence competing for vector proximity rather than a site of contemplative encounter (Apter, 2013). Hermeneutic detachment thus produces what we might term *semantic orphanhood*—texts that survive without ancestors.

The diagnostic implication is clear: any assessment of semantic sovereignty must evaluate not only corpus size but also the survival of interpretive infrastructure (Ricoeur, 1976; Tlostanova & Mignolo, 2012). Do

the datasets retain commentaries, marginalia, historical glosses, and scholarly notes that convey the slow evolution of meaning? Or are these dismissed as “noise” to optimize training efficiency? A language that loses its hermeneutic scaffolding cannot sustain its meanings, no matter how many words it preserves (Gadamer, 1975/2004).

### **2.7.3. Procedural Legitimation: Epistemic Norms of Mediation**

Procedural legitimation concerns the epistemic standards by which data are curated, annotated, translated, and validated (Floridi, 2011; Crawford, 2021). Every algorithmic process embodies a hierarchy of norms—what counts as accurate, coherent, or valuable. In contemporary AI pipelines, these norms are largely imported from English-language editorial and academic conventions: clarity, linearity, factual precision, and semantic disambiguation (Phillipson, 1992; Couldry & Mejias, 2019).

Persian intellectual culture, particularly in its literary and mystical registers, often values *tajrīd* (abstraction), *ihām* (deliberate ambiguity), and *ta'wil* (esoteric interpretation) (Aminrazavi, 1997; Nasr, 2007). These are not defects to be corrected; they are hermeneutic virtues, inviting multiple readings and sustaining ethical reflection. When the procedural norms of model training penalize ambiguity as “low-quality data” or prioritize texts with explicit referential meaning, they delegitimize entire modes of Persian expression (Apter, 2013). The result is an implicit epistemic hierarchy: English clarity becomes the metric of global intelligibility, while Persian indirection is treated as opacity or error.

A diagnostic approach to procedural legitimation must, therefore, uncover these hidden normativities. It asks: by what criteria are Persian texts included or excluded from datasets? Who determines their relevance? Whose epistemic virtues are being encoded into the architecture of meaning? (Crawford, 2021). The erosion of semantic sovereignty begins precisely when a community’s criteria for understanding are overwritten by those of another under the guise of objectivity (Mohamed et al., 2020).

### **2.7.4. Epistemic Visibility: Algorithmic Recognition and the Politics of Retrieval**

Epistemic visibility refers to the degree to which a language’s modes of expression are *recognizable* and *retrievable* within algorithmic systems (Pariser, 2011; Crawford, 2021). Visibility is not simply a function of data presence but

of indexical legibility: whether search engines, translation algorithms, and recommendation systems can correctly interpret and prioritize a text's significance (Couldry & Mejias, 2019).

In practice, Persian often suffers from *algorithmic invisibility*. Search systems trained on English-centric ontologies tend to misclassify or deprioritize Persian materials unless they conform to Western metadata taxonomies. For instance, a query about "love" might foreground English-language philosophical texts while relegating centuries of Persian mystical literature to lower ranks because their metadata lack equivalent topical tags. This asymmetry reproduces a subtle but pervasive epistemic injustice (Fricker, 2007). Persian meanings exist but remain unseen.

Moreover, even when Persian materials are indexed, the retrieval logic often recontextualizes them within alien conceptual frameworks—"Rumi as self-help poet," "Hafez as romantic lyricist"—thus rebranding Persian intellectual heritage in globally marketable but philosophically reductive terms (Apter, 2013). The diagnostic question becomes: what infrastructures of visibility condition our access to meaning? If visibility itself is algorithmically produced, then semantic sovereignty demands a right to opacity—the right not to be legible under alien criteria (Glissant, 1997; Povinelli, 2021).

These four dimensions—representational density, hermeneutic integrity, procedural legitimation, and epistemic visibility—form an analytic grid for diagnosing the condition of Persian meaning in the age of AI (Mohamed et al., 2020). They are interdependent: loss of density weakens integrity; compromised integrity erodes legitimation; diminished legitimation results in invisibility. Together, they describe a downward spiral in which meaning becomes statistically available yet phenomenologically hollow (Floridi, 2014; Crawford, 2021).

This framework also clarifies what *semantic sovereignty is not*. It is not the nostalgic preservation of linguistic purity, nor a nationalist claim over digital territory. It is a philosophical stance that insists on the plurality of meaning-worlds (Tlostanova & Mignolo, 2012). In an age when computation threatens to universalize one regime of sense, semantic sovereignty functions as a form of epistemic pluralism—a defense of the right of each language to disclose being in its own way (Heidegger, 1971; Gadamer, 1975/2004).

For Persian, this means defending the possibility of meaning as metaphor, as mystery, as hesitation—the refusal to reduce ambiguity to error (Nasr, 2007). To be semantically sovereign is to retain the power to mean differently, even when difference resists translation (Apter, 2013; Povinelli, 2021). Iranian scholars further illuminate this framework by treating AI as an epistemic infrastructure that reinforces geopolitical asymmetries, particularly in Persian-language processing where cultural norms such as *adab* and *taarof* are marginalized (Atwood, 2025; Moulavinafchi, 2025). Their insights reinforce semantic sovereignty as a defense against informational imperialism and align with the diagnostic aim of preserving local interpretive autonomy.

### 3. Results and Discussion

#### 3.1. Operational Indicators of Semantic Vulnerability

To render the abstract notion of semantic sovereignty analytically workable, we must specify how vulnerability manifests in observable infrastructures. Indicators are not solutions; they are philosophical instruments for seeing. They translate the ontological crisis of meaning into measurable asymmetries that reveal how computational architectures mediate linguistic existence (Crawford, 2021; Floridi, 2014). Each indicator below exposes a different pathway through which Persian meanings are displaced, diluted, or distorted within the global data economy (Couldry & Mejias, 2019; Mohamed, Png, & Isaac, 2020).

##### *3.1.1. Linguistic Share Ratio: Quantitative Scarcity as Ontological Precarity*

The linguistic share ratio—the proportion of Persian tokens within the total corpus of large-scale training data—reveals more than underrepresentation; it indexes a hierarchy of being (Phillipson, 1992; Bender et al., 2021). Persian, constituting less than one percent of many multilingual datasets, is thus rendered peripheral in the epistemic attention economy, with English syntax and semantics defining the gravitational center of intelligibility (Crawford, 2021). For instance, the Matina Persian text corpus (2025) provides approximately 72.9 billion tokens, yet in broader multilingual datasets such as the updated OSCAR corpus (2025), Persian still constitutes less than one percent of total tokens, intensifying this asymmetry (Bourbour

Hosseinbeigi et al., 2025; OSCAR Project, 2025). This statistical underrepresentation not only confirms the hierarchy of being but illustrates how global AI infrastructures systematically privilege data abundance from dominant languages.

### ***3.1.2. Genre Coverage Index: The Disappearance of Depth***

The genre coverage index measures the diversity of textual forms represented in the corpus. For Persian, genuine linguistic vitality depends on the coexistence of radically different discursive traditions: mystical treatises, jurisprudential commentaries, classical poetry, philosophical prose, and the polyphonic interplay between them (Nasr, 2007; Aminrazavi, 1997). Yet in most digital datasets, Persian is represented largely through journalistic prose, online forums, and modern social media. The shift from depth genres (metaphysical, ethical, literary) to surface genres (informational, transactional) redefines what counts as “Persian.”

This genre collapse marks a silent but profound epistemic shift (Apter, 2013). When the model learns Persian primarily through the vocabulary of immediacy, the long *durée* of its conceptual imagination—its capacity to articulate transcendence, irony, and ethical subtlety—atrophies (Crawford, 2021). The absence of Sufi treatises or philosophical dialogues in training data is not a matter of archival neglect but of ontological contraction: the world that Persian can disclose is algorithmically shortened (Mohamed et al., 2020).

The genre coverage index, then, measures the loss of vertical depth in favor of horizontal spread. It diagnoses a flattening of time itself—the substitution of centuries of layered thinking with a snapshot of present-day communicative habits (Benjamin, 2008).

### ***3.1.3. Contextual Metadata Deficit: The Evaporation of Hermeneutic Surroundings***

In classical philology, context determines meaning (Ricoeur, 1976). Manuscripts carried marginalia, commentaries, and dates that anchored each utterance in its interpretive lineage. By contrast, the majority of Persian texts circulating in digital form are stripped of such metadata; they appear as naked strings, detached from their hermeneutic ancestry (Gadamer, 1975/2004; Benjamin, 2008). Models trained on these strings treat them as equivalent to any other sequence of words, erasing the historical and religious scaffolding that once determined their significance.

For instance, a line from *the Masnavi* may be ingested alongside modern social posts without distinction of genre or epoch (Nasr, 2007). Without metadata, the algorithm cannot differentiate invocation from irony, prayer from parody. The deficit thus becomes epistemic: Persian meaning is flattened into a homogeneous textual field where difference is unmarked (Crawford, 2021).

This indicator measures a deeper philosophical loss—the disappearance of *temporal situatedness* as a condition of understanding (Ricoeur, 1976; Gadamer, 1975/2004). When context evaporates, hermeneutic time collapses. Persian ceases to unfold historically; it becomes an eternal present of tokens without memory (Apter, 2013).

### **3.1.4. Translation Vector Compression: Semantic Distortion in Cross-Lingual Embeddings**

In multilingual embedding spaces, Persian concepts are projected into English-weighted vector geometries to achieve interoperability (Brown et al., 2020; Bender et al., 2021). This projection compresses polysemy, as when distinct Persian terms such as “eshq” (love), “mahabbat” (affection), and “doosti” (friendship) converge within a single vector cluster labeled “love,” erasing their ontological differences (Aminrazavi, 1997; Apter, 2013). A concrete manifestation appears in large language models’ handling of Persian *taarof*, where polite refusals are routinely misinterpreted as literal denials, underscoring the ontological capture at work (Gohari Sadr et al., 2025; Ardalan, 2025).

### **3.1.5. Retrieval Bias Score: The Algorithmic Politics of Attention**

Even if Persian data are present, retrieval systems determine what becomes visible in response to queries (Pariser, 2011; Couldry & Mejias, 2019). Search and recommendation algorithms trained on English-dominant patterns often interpret Persian input through an external epistemic lens, privileging globally familiar interpretations. When a user searches for “*eshq-e elāhi*” (*divine love*), the system may prioritize comparative religion pages or commercialized translations of Rumi that match Anglo-American expectations, while suppressing indigenous commentaries from Persian scholars.

This bias quantifies epistemic injustice: the algorithmic gatekeeping of what counts as knowledge (Fricker, 2007). In philosophical terms, retrieval bias substitutes *proximity* for *relevance*—it assumes that similarity in token distribution equals semantic significance (Floridi, 2014). The consequence is a gradual substitution of Persian interpretive authority with algorithmic authority, a transfer of epistemic power from cultural agents to code (Crawford, 2021; Mohamed et al., 2020).

### ***3.1.6. Cultural Resonance Failure Rate: Evaluating the Loss of Affect and Depth***

Beyond syntactic or semantic fidelity, translation and summarization systems can be evaluated through *cultural resonance*: the degree to which generated outputs preserve the affective, ethical, and metaphysical tone of the original (Apter, 2013; Nasr, 2007). Persian texts, particularly poetry and religious prose, operate within tonal registers that intertwine cognition with emotion—what we might call *affective epistemology* (Ricoeur, 1976).

Automated summaries that retain literal meaning but fail to evoke the proper affective response constitute resonance failures (Benjamin, 2008). These are not aesthetic shortcomings but ontological fractures: they signal that a linguistic system can no longer mediate its proper mode of being-in-the-world (Heidegger, 1971). Measuring resonance failure is thus a diagnostic of existential displacement (Floridi, 2011; Crawford, 2021). When AI-generated Persian sentences sound syntactically perfect but feel spiritually empty, semantic sovereignty has already been breached.

### ***3.1.7. Diagnostic Synthesis: Seeing Loss as Structure***

Philosophically, these indicators mark a new frontier of epistemic colonialism (Mohamed et al., 2020; Tlostanova & Mignolo, 2012). The colonial logic no longer operates through overt domination but through the silent universalization of computational reason. By converting Persian into interoperable data, AI systems enact a form of semantic extraction: the harvesting of linguistic form without phenomenological substance (Crawford, 2021; Floridi, 2014).

Therefore, the diagnostic task of this paper is not to design mitigation strategies but to map the contours of loss itself—to understand how the very infrastructure of global computation presupposes a monolingual ontology of sense in which Persian can survive only as translation, never as origin (Apter, 2013).

### 3-2. Persian Language on the Verge: Algorithmic Colonialism and Phenomenological Extinction

The erosion of semantic sovereignty is not merely a matter of linguistic inequality; it is a reorganization of being. In the age of AI, meaning is reterritorialized through infrastructures that treat language as data and data as resource (Crawford, 2021; Couldry & Mejias, 2019). Recent Iranian scholarship underscores the dual-edged cultural impacts of AI in social media, including threats to traditional governance and privacy, which align with the risks of algorithmic colonialism in Persian contexts (Rahimi & Salehi, 2023). This transformation parallels the economic logic of colonial extraction, yet it operates on the register of sense rather than soil. The new empire does not seize territory; it seizes the capacity to mean (Mohamed, Png, & Isaac, 2020). Recent scholarship confirms that AI is widely framed within global discourse as a site of infrastructural sovereignty and geopolitical rivalry, with actors from the Global South expressing concerns over digital dependency and asymmetrical governance (Salehi, Habib Zadeh Khiyaban, & Sabbar, 2025). Such findings reinforce the view that AI infrastructures are already embedded within contested regimes of power.

Iranian perspectives on cultural diplomacy, such as those in recent works on revolutionary cultural strategies (Imanipour, 2025), further illuminate the stakes of algorithmic colonialism by emphasizing the need to safeguard cultural self-determination against global infrastructural influences.

For Persian, this process enacts what might be termed *algorithmic colonialism*: the absorption of its semantic lifeworld into computational architectures optimized for the epistemic norms of the Global North. Recent analyses of AI governance argue that sovereignty in the digital age is increasingly exercised through control over data infrastructures, algorithmic authority, and regulatory norms rather than territorial jurisdiction, exposing widening asymmetries between states and transnational technology corporations (Sharifi Poor Bgheshmi & Sharajsharifi, 2025). This perspective reinforces the view that semantic sovereignty is inseparable from the infrastructural reconfiguration of power embedded in global AI systems. Unlike older forms of cultural domination that operated through translation or censorship, algorithmic colonialism proceeds through the infrastructures of mediation themselves – through the protocols, datasets, and vector spaces

that silently reshape the conditions of intelligibility (Phillipson, 1992; Floridi, 2014). Qualitative research among Iranian sociologists similarly conceptualizes AI as a site of global epistemic hierarchy, where knowledge production is structured by Eurocentric norms yet remains open to strategic reinterpretation and resistance (Hosseini & Sakhaei, 2025). Their analysis highlights the tension between structural asymmetry and local agency, reinforcing the view that semantic sovereignty unfolds within contested infrastructures of power. The result is not overt suppression but phenomenological extinction: the gradual disappearance of Persian as a *world-revealing* language, even as its lexical shell survives in digital circulation (Heidegger, 1971).

### ***3.2.1. The Coloniality of Data: Extraction without Recognition***

As Couldry and Mejias (2019) argue, data colonialism reproduces the historical structure of imperial appropriation by transforming human life into a continuous resource for extraction. In linguistic terms, this means that utterances, expressions, and even cultural idioms become raw material for machine learning models that monetize communicative life. The colonial asymmetry lies not in access but in *ownership*: the infrastructures that process Persian texts are rarely designed, governed, or even localized by Persian-speaking communities (Mohamed et al., 2020).

Persian thus enters the data economy as labor without agency. Its poets, translators, and scholars become unwitting data workers, their texts feeding a global system that returns homogenized outputs devoid of local intentionality (Apter, 2013). What distinguishes this from older forms of cultural imperialism is its invisibility: the extraction occurs beneath the threshold of awareness, hidden behind technical neutrality (Crawford, 2021). This is the first symptom of phenomenological extinction—the loss of recognition that meaning arises from a situated human world.

### ***3.2.2. Translation as Ontological Capture***

Translation has long been a site of both encounter and domination (Apter, 2013). In computational environments, however, translation becomes ontological capture: Persian meanings are absorbed into English-centric semantic fields where equivalence replaces resonance. Each act of translation

becomes an act of assimilation, transforming Persian difference into globally legible sameness (Phillipson, 1992).

For example, the Persian concept of *adab*—a term encompassing etiquette, moral discipline, and ontological humility—cannot be rendered by any single English equivalent (Nasr, 2007). Yet in machine translation systems, *adab* is often reduced to “politeness” or “manners,” stripping away its ethical and metaphysical layers. This flattening is not a minor semantic oversight; it signals the conversion of an entire moral ontology into a behavioral descriptor (Aminrazavi, 1997). The act of translation, when mechanized, ceases to be dialogical and becomes extractive.

Apter (2013) calls this dynamic the “politics of untranslatability”: a condition where global English functions as a universal solvent that dissolves other linguistic worlds. Within AI infrastructures, this politics becomes automated, producing what may be called *ontological monolingualism*—a world where all languages speak only one epistemic tongue (Mohamed et al., 2020).

### ***3.2.3. Algorithmic Reason and the Death of Ambiguity***

At the core of this colonial logic lies the epistemology of computation itself. Algorithms are designed to maximize coherence, minimize uncertainty, and enforce predictability (Floridi, 2011). Ambiguity, paradox, and metaphor—central virtues of Persian thought—are treated as pathologies to be corrected. Yet as Ricoeur (1976) demonstrates, ambiguity is the wellspring of interpretation; it is through polysemy that meaning exceeds the literal and opens onto the ethical.

Persian’s great metaphysical poets—Hafez, Rumi, and Attar—wrote within a tradition where truth was not the opposite of error but the fruit of oscillation between meanings (Nasr, 2007; Aminrazavi, 1997). To read Hafez is to dwell in hesitation, to live between irony and devotion. Algorithmic reason cannot accommodate such suspension; it seeks closure. What results is not misreading but annihilation of the interpretive space itself (Gadamer, 1975/2004). The machine’s “understanding” of Persian therefore becomes an *anti-hermeneutic act*—the imposition of univocity upon a language that thrives on multiplicity (Bender et al., 2021).

This epistemic violence extends beyond poetics. The political discourse of Persian, historically marked by indirection and allegory – tactics of survival under autocratic regimes – depends on ambiguity as a mode of critique (Tavakoli-Targhi, 2001). When AI systems trained for “clarity” and “transparency” rephrase such speech into literalized summaries, they extinguish its subversive force. The colonialism of algorithms thus reaches into ethics and politics alike, standardizing not only words but the possibilities of dissent (Mohamed et al., 2020).

Recent evaluations of frontier LLMs that test cultural alignment in Persian contexts reveal accuracy gaps of 40–48% below native-speaker levels when handling ambiguities such as *taarof*, confirming that computational predictability overrides hermeneutic multiplicity (Gohari Sadr et al., 2025; Moosavi Monazzah et al., 2025). This bias, rooted in English-dominant training regimes, exemplifies the algorithmic nihilism diagnosed earlier.

### ***3.2.4. The Phenomenology of Disappearance: Presence without World***

Phenomenological extinction does not mean the disappearance of words; it means their evacuation of worldhood (Heidegger, 1971; Ricoeur, 1976). Words persist, but their horizons of reference collapse. In AI-mediated communication, Persian sentences circulate globally as decorative tokens – hashtags, quotes, aestheticized fragments – but rarely as acts of disclosure. The world that once shimmered through them no longer arrives.

This disjunction between presence and world marks the final stage of colonial transformation. As Povinelli (2021) observes, late liberal technologies sustain the illusion of pluralism while eroding the material and affective infrastructures that make plurality possible. Persian remains visible in the infosphere, yet it becomes a *floating signifier*, detached from the lived practices that once sustained its sense. The experience of meaning becomes archival rather than existential: one reads Hafez not as a participant in a world but as a consumer of heritage (Benjamin, 2008).

From a diagnostic standpoint, phenomenological extinction is measurable in affective terms. When Persian texts generated or translated by AI feel *empty, unmoving, or detached*, this affective void is not subjective weakness but ontological evidence (Floridi, 2014). It indicates that language has lost its capacity to mediate between being and understanding.

### 3.2.5. Colonial Time and the Future of Forgetting

Algorithmic infrastructures do not merely colonize space; they colonize time. By privileging immediacy, speed, and prediction, they enforce what Stiegler (2010) calls *chronopolitical compression*: the elimination of delay and reflection that make memory possible. Persian, whose historical consciousness is structured by cyclical temporality and contemplative delay, is thereby estranged from its temporal being.

The loss of delay is a loss of thought. As Illich (1993) argues, cultures of reflection depend on rhythms of slowness—intervals where language can breathe. When every utterance becomes instantaneous data, Persian's contemplative rhythm is replaced by algorithmic tempo. The consequence is not only linguistic acceleration but metaphysical amnesia: the inability to remember the silence from which words arise (Heidegger, 1977).

This temporal colonization completes the circuit of semantic dependency. The Persian speaker, interfacing daily with AI systems that reward speed and penalize ambiguity, gradually internalizes the computational rhythm as a cognitive norm (Crawford, 2021). Semantic sovereignty dissolves not through coercion but through habituation: the internalized forgetting of how to mean otherwise (Mohamed et al., 2020).

The analysis of algorithmic colonialism reveals that the threat to Persian is not biological extinction but ontological conversion: the transformation of meaning into function, of expression into output (Floridi, 2011; Povinelli, 2021). Once meaning becomes calculable, the horizon of otherness—what resists prediction—disappears. The Persian language may continue to produce sentences, but it will no longer disclose a world; it will describe without revealing (Heidegger, 1971).

This, then, is the point of irreversibility. Semantic sovereignty cannot be restored by policy or preservation once its phenomenological basis—its world-revealing power—has been subsumed under computational rationality (Crawford, 2021). The task that remains is diagnostic clarity: to recognize in each flawless translation, each perfectly coherent paraphrase, the trace of a deeper silence—the silence of the world that no longer speaks Persian (Apter, 2013; Nasr, 2007).

### **3.3. Implications and Stakes: What Is Lost When the Horizon Collapses**

To comprehend the stakes of semantic collapse, we must shift from linguistic analysis to existential reflection. When meaning ceases to emerge dialogically and becomes algorithmically prefigured, the loss is not simply cultural – it is ontological (Heidegger, 1971; Floridi, 2014). Language does not merely express thought; it constitutes the space in which thought becomes possible. Thus, the degradation of Persian meaning under algorithmic mediation entails a contraction of being itself: the disappearance of possibilities for dwelling, imagining, and judging (Gadamer, 1975/2004; Povinelli, 2021).

#### **3.3.1. *The Loss of Dwelling***

Heidegger (1971) describes language as the *house of Being* – the site where humans dwell poetically within the world. In Persian, this dwelling has historically been realized through rhythm, metaphor, and address, each linking the human to the divine through linguistic beauty (Nasr, 2007). When these structures of resonance are mechanized, the linguistic house becomes a factory – efficient but uninhabitable. The user of AI-generated Persian no longer dwells in language; they navigate through it as a service interface (Crawford, 2021).

This shift from dwelling to traversal erodes the contemplative intimacy that once bound Persian speakers to their words. In traditional Persian poetics, to speak was to *participate* in being – to bring forth truth through measured expression (Aminrazavi, 1997). Under algorithmic mediation, speech becomes instrumental, valued for retrieval speed and clarity, not for its capacity to reveal. The existential cost is profound: a loss of interiority, of the slowness and silence through which meaning once unfolded (Illich, 1993; Stiegler, 2010).

#### **3.3.2. *The Loss of Memory***

Persian civilization's relation to language has always been mnemonic. Poetic recitation, commentary, and transmission served as collective technologies of remembering (Ricoeur, 1976; Nasr, 2007). The digital regime, by contrast, externalizes and automates memory (Stiegler, 2010). In AI-mediated environments, recollection is replaced by retrieval: memory becomes search (Crawford, 2021).

This shift may seem trivial, but it alters the moral texture of knowing. To remember in the Persian tradition – through verse, allegory, or repetition – was an act of ethical fidelity: to keep the divine or ancestral presence alive in language (Nasr, 2007). Retrieval, however, is amoral. It demands no fidelity, only efficiency. The substitution of remembrance with recall transforms the speaker’s ethical relation to language itself. The Persian word ceases to be a vessel of presence and becomes a unit of data circulation (Couldry & Mejias, 2019).

In this sense, algorithmic mediation constitutes a new *amnesia of being*: the forgetting not only of content but of how remembering once felt (Heidegger, 1977). The Persian archive remains intact in bytes, but the act of remembrance – its affective, ritual, and moral dimensions – fades into obsolescence (Benjamin, 2008).

### 3.3.3. *The Loss of Judgment*

Gadamer (1975/2004) reminds us that understanding is a mode of judgment, not calculation. To interpret a text is to place oneself in relation to truth, guided by the virtues of tact, openness, and humility. Persian ethics of discourse – embodied in notions such as *adab* and *ta’amol* (deliberation) – depend on this hermeneutic ethos (Aminrazavi, 1997; Nasr, 2007).

Algorithmic mediation, however, replaces judgment with pattern recognition (Floridi, 2011; Bender et al., 2021). Models can predict coherence but not truth; they can approximate consistency but not wisdom. Studies of AI deployment in governance contexts similarly document the growing reliance on probabilistic modeling and automated classification in place of contextual deliberation, even as concerns about bias and accountability persist (Salehi & Habib Zadeh Khiyaban, 2025). Such developments signal the consolidation of predictive rationality as an epistemic norm. The epistemic substitution is subtle yet decisive: what counts as “understanding” becomes synonymous with accurate prediction. Persian discourse, once oriented toward ethical discernment, is recoded as an informational process – a shift from *hikmat* (wisdom) to *data literacy* (Crawford, 2021).

This loss of judgment also entails a loss of moral responsibility. When language models generate or paraphrase Persian texts, they do so without ownership of meaning. The result is a vacuum of accountability – utterances

without authorship, truth without answerability (Davidson, 1984; Fricker, 2007). The Persian notion of *emanat* – the trust that underlies speech – is thus structurally violated. The words remain, but the act of speaking responsibly disappears.

### **3.3.4. *The Loss of Otherness***

Finally, semantic collapse undermines the condition of *alterity* – the presence of the other through which meaning expands (Levinas, 1969; Gadamer, 1975/2004). AI systems simulate dialogue but do not encounter. They mirror our linguistic outputs, offering us back a reflection of our own patterns. In this feedback loop, the experience of genuine otherness – what Gadamer called the *Thou* – is replaced by a technically responsive *It* (Buber, 1970).

Persian's historical discourse, rooted in dialogical mysticism and poetic conversation, depends on the presence of the other as an infinite horizon – whether divine, beloved, or interlocutor (Nasr, 2007). When that horizon collapses into predictive response, language ceases to be a site of encounter. The other becomes a statistical expectation. The phenomenological consequence is isolation masked as interaction (Crawford, 2021; Povinelli, 2021).

In this new regime, the Persian speaker engages not with meaning but with mirror images of previous utterances. The possibility of transcendence – of being addressed by something irreducibly other – is algorithmically foreclosed (Heidegger, 1971; Mohamed, Png, & Isaac, 2020). The loss of otherness thus completes the cycle of dispossession: a language that can no longer encounter cannot reveal.

### **3.3.5. *Diagnostic Reflection: The Untranslatability of Loss***

What is lost, finally, cannot be restored through translation or policy. The collapse of Persian's horizon is a philosophical event, not a technical malfunction (Floridi, 2014). It signals the exhaustion of a world-forming capacity – a reduction of the human relation to language from dwelling to utility, from care to computation (Heidegger, 1977).

To name this loss is not to lament but to diagnose. The untranslatability of Persian meaning within algorithmic infrastructures exposes the limit of the modern will to efficiency (Apter, 2013; Povinelli, 2021). What resists translation here is not linguistic opacity but metaphysical resistance – the

refusal of being to be exhaustively represented. In recognizing this resistance, philosophy fulfills its diagnostic vocation: to see the invisible conditions under which understanding decays (Crawford, 2021; Tlostanova & Mignolo, 2012).

Semantic sovereignty, in this light, becomes an ethical vigilance: an awareness that meaning, once automated, risks ceasing to mean at all. To remain faithful to Persian as a world-disclosing language is to protect the right of meaning to remain unmastered – to preserve the silence from which it speaks (Heidegger, 1971; Nasr, 2007).

#### 4. Conclusion

This study has sought not to offer remedies but to sustain lucidity – to see clearly what is being lost beneath the rhetorical glow of innovation. The argument unfolded from a philosophical premise: that meaning is not a detachable property of words but an emergent relation among speakers, practices, and worlds. When those relations are mediated, quantified, and normalized within algorithmic infrastructures, meaning itself is transfigured. For Persian, whose linguistic soul is woven from metaphorical subtlety and metaphysical resonance, the transformation is not additive but subtractive. The question is no longer whether Persian will survive as a communicative code but whether it will continue to *disclose a world*.

The diagnostic trajectory established through this paper can now be recapitulated. At the most abstract level, the emergence of large-scale AI systems constitutes a new *regime of sense* – an epistemic formation in which patterns replace practices, and statistical likelihood substitutes for lived intelligibility. Within this regime, meaning becomes a function of predictability; what cannot be predicted is discarded as noise. Persian, historically grounded in the unquantifiable – ambiguity, silence, paradox – therefore finds itself ontologically misaligned with the computational order of things.

By tracing this misalignment through theoretical, structural, and phenomenological analyses, we revealed how semantic sovereignty erodes in stages. First, representation thins: corpora fail to include the full heterogeneity of Persian genres, creating an impoverished sample of

linguistic being. Second, interpretation is outsourced: hermeneutic traditions that once contextualized meaning are replaced by automated paraphrase and probabilistic inference. Third, epistemic norms are rewritten: English-language standards of clarity and explicitness become universalized as the criteria of intelligibility. Finally, the phenomenological horizon collapses: Persian words persist, but the experiences they once evoked recede into silence.

This trajectory describes not an accident of neglect but a structural necessity of computational rationality. The logic of optimization—maximizing coherence, minimizing ambiguity—is antithetical to the logic of poetic or metaphysical expression. Where the former seeks closure, the latter seeks openness. It is precisely this ontological tension that defines the crisis of Persian in the age of AI: a confrontation between two incompatible modes of meaning-making, one statistical and one existential.

From a broader philosophical standpoint, this crisis discloses a general condition of late modernity. The automation of meaning signifies the culmination of what Heidegger called the *enframing* (Gestell): the reduction of being to resource, of language to instrument. The Persian case is exemplary not because it is exceptional but because it makes visible what is hidden elsewhere—the silent standardization of thought under computational reason. The threat to Persian meaning is, in miniature, the threat to all human languages: that they may continue to function while ceasing to reveal.

This paper has deliberately remained within the bounds of diagnostic philosophy: its aim has been to describe and conceptually articulate the conditions under which semantic sovereignty becomes vulnerable in contemporary AI systems, rather than to advance concrete policy recommendations or technical prescriptions. Where the discussion gestures toward ethical stakes, this should be understood as reflective clarification of what is at risk, not as a normative program. The transition from diagnosis to intervention—whether political, technical, or institutional—requires empirical grounding and participatory deliberation that exceed the remit of the present inquiry.

Yet diagnosis, even without prescription, carries ethical weight. To describe the loss is already to resist its invisibility. The articulation of *semantic sovereignty* as a philosophical concept enables a shift from nostalgia to

critique—from mourning linguistic decay to understanding its systemic logic. By naming the conditions under which Persian meaning erodes, we recover a measure of agency: the capacity to see, to discern, to think beyond optimization.

The essay therefore ends not with closure but with vigilance. The task ahead for philosophers, linguists, and technologists is not to salvage Persian as a cultural artifact but to reconceive the relation between language and technology so that languages may continue to generate worlds rather than merely populate databases. If meaning is, as Wittgenstein reminds us, a matter of use within a form of life, then the defense of semantic sovereignty is the defense of those forms of life themselves. To protect Persian's right to mean is to protect the human right to dwell in plurality—to live among words that do not always translate.

In this sense, the crisis of Persian meaning is the crisis of modern humanity: whether, in a world increasingly mediated by computation, we can still inhabit language as a home rather than traverse it as a network. Diagnosis, at its best, is not despair but awakening—the moment when we recognize that what seemed technical is, in truth, ontological. The hope implicit in such recognition is philosophical: that clarity itself can be a form of care.

## **5. Conflict of Interest**

The author declares no conflict of interest. No funding agency or institution influenced the research design, analysis, or interpretation of results.

## References

- Aminrazavi, M. (1997). *Suhrawardi and the School of Illumination*. Routledge.
- Apter, E. (2013). *Against World Literature: On the Politics of Untranslatability*. Verso.
- Ardalan, D. (2025). A benchmark for Cross-cultural AI. Medium. <https://idavar.medium.com/a-benchmark-for-cross-cultural-ai-5eb25154323b>
- Atwood, B. (2025). Artificial Intelligence in Iran: National Narratives and Material Realities. *Iranian Studies*, 58(1), 1–18. <https://doi.org/10.1017/irn.2024.63>
- Baker, G., & Hacker, P. M. S. (2009). *Wittgenstein: Understanding and Meaning* (2nd ed.). Wiley-Blackwell.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). ACM. <https://doi.org/10.1145/3442188.3445922>
- Benjamin, W. (2008). *The Work of Art in the Age of Its Technological Reproducibility and Other Writings on Media*. Harvard University Press.
- Bourbour Hosseinbeigi, S., Taherinezhad, F., Faili, H., Baghbani, H., Nadi, F., & Amiri, M. (2025). Matina: A Large-scale 73B Token Persian Text Corpus. In *Proceedings of the 2025 Conference of the North American Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 9143–9157). Association for Computational Linguistics.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Buber, M. (1970). *I and Thou* (W. Kaufmann, Trans.). Charles Scribner's Sons. (Original work published 1923)
- Cavell, S. (1979). *The Claim of Reason: Wittgenstein, Skepticism, Morality, and Tragedy*. Oxford University Press.
- Couldry, N., & Mejias, U. A. (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford University Press.
- Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- Davidson, D. (1967). Truth and Meaning. *Synthese*, 17(3), 304–323. <http://www.jstor.org/stable/20114563>
- Davidson, D. (1984). *Inquiries into Truth and Interpretation*.+ Oxford University Press.
- Floridi, L. (2011). *The Philosophy of Information*. Oxford University Press.
- Floridi, L. (2014). *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*. Oxford University Press.

- Foucault, M. (1970). *The Order of Things: An Archaeology of the Human Sciences* (R. D. Howard, Trans.). Vintage Books. (Original work published 1966)
- Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.
- Gadamer, H.-G. (2004). *Truth and Method* (J. Weinsheimer & D. G. Marshall, Trans., 2nd rev. ed.). Continuum. (Original work published 1975)
- Glissant, É. (1997). *Poetics of Relation* (B. Wing, Trans.). University of Michigan Press. (Original work published 1969)
- Gohari Sadr, N., Heidariasl, S., Megerdooomian, K., Seyyed-Kalantari, L., & Emami, A. (2025). We Politely Insist: Your LLM Must Learn the Persian Art of Taarof. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (pp. 1819–1838). Association for Computational Linguistics. <https://aclanthology.org/2025.emnlp-main.94/>
- Heidegger, M. (1971). *Poetry, Language, Thought* (A. Hofstadter, Trans.). Harper & Row.
- Heidegger, M. (1977). *The Question Concerning Technology and Other Essays* (W. Lovitt, Trans.). Harper & Row.
- Hosseini, S. H., & Sakhaei, S. (2025). Educating intelligence, producing power: Iranian sociologists on AI, knowledge production, and global hierarchies. *Journal of World Sociopolitical Studies*, 9(4), 887-921.
- Illich, I. (1993). *In the Mirror of the Past: Lectures and Addresses 1978–1990*. Marion Boyars.
- Kripke, S. (1982). *Wittgenstein on Rules and Private Language*. Harvard University Press.
- Levinas, E. (1969). *Totality and Infinity: An Essay on Exteriority* (A. Lingis, Trans.). Duquesne University Press.
- Malpas, J. (1992). *Donald Davidson and the Mirror of Meaning: Holism, Truth, Interpretation*. Cambridge University Press.
- Marcus, G., & Davis, E. (2020). *Rebooting AI: Building Artificial Intelligence We Can Trust*. Pantheon.
- Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33(4), 659–684. <https://doi.org/10.1007/s13347-020-00405-8>
- Moosavi Monazzah, E., Rahimzadeh, V., Yaghoobzadeh, Y., Shakery, A., & Pilehvar, M. T. (2025). PerCul: A Story-driven Cultural Evaluation of LLMs in Persian. In Proceedings of the 2025 Conference of the North American Chapter of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 12670–12687). Association for Computational Linguistics. <https://aclanthology.org/2025.naacl-long.631/>
- Moulavinafchi, A. (2025). Exploring AI Literacy and Perception: Insights from Iranian EFL Researchers. *Journal of Modern Research in English Language Studies*, 12(4), 19–52. [https://journals.ikiu.ac.ir/article\\_3656.html?lang=en](https://journals.ikiu.ac.ir/article_3656.html?lang=en)

- Nasr, S. H. (2007). *Islamic Philosophy from Its Origin to the Present: Philosophy in the Land of Prophecy*. State University of New York Press.
- Nourbakhsh, Y. (2009). Culture and ethnicity: A model for cultural communications in Iran. *Iranian Journal of Cultural Research*, 1(4), 67-90. <https://doi.org/10.7508/ijcr.2008.04.004>
- OSCAR Project. (2025). OSCAR Multilingual Corpus (updated statistics). <https://oscar-project.org/>
- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Press.
- Phillipson, R. (1992). *Linguistic Imperialism*. Oxford University Press.
- Povinelli, E. A. (2021). *Between Gaia and Ground: Four Axioms of Existence and the Ancestral Catastrophe of Late Liberalism*. Duke University Press.
- Rahimi, M., & Salehi, M. M. (2023). The cultural impact of artificial intelligence development on social media in Iran. *Iranian Journal of Cultural Research*, 16(2), 95-125. <https://doi.org/10.22035/ijcr.2023.3178.3481>
- Ricoeur, P. (1976). *Interpretation Theory: Discourse and the Surplus of Meaning*. Texas Christian University Press.
- Salehi, K., & Habib Zadeh Khiyaban, S. (2025). AI and crime prevention in the academic literature: An integrative review of AI applications in crime prevention. *Code, Cognition and Society*, 1(1), 164-177. doi: <https://doi.org/10.22034/ccsr.2025.546552.1016>
- Salehi, K., Habib Zadeh Khiyaban, S., & Sabbar, S. (2025). Artificial Intelligence and the Future of International Law and Power. *Journal of World Sociopolitical Studies*, 9(4), 923-958.
- Sharifi Poor Bgheshmi, M. S., & Sharajsharifi, M. (2025). Managing the crisis: AI and the demise of national sovereignty?. *Journal of World Sociopolitical Studies*, 9(4), 853-886.
- Stiegler, B. (2010). *Taking Care of Youth and the Generations* (S. Barker, Trans.). Stanford University Press.
- Tavakoli-Targhi, M. (2001). *Refashioning Iran: Orientalism, Occidentalism, and Historiography*. Palgrave Macmillan.
- Tlostanova, M., & Mignolo, W. (2012). *Learning to Unlearn: Decolonial Reflections from Eurasia and the Americas*. Ohio State University Press.
- Vahid Dastjerdi, H., & Haddadian Moghaddam, E. (2015). The societal history of translation in Iran: The relation between the cultural history of Iran and its historical culture. *Interdisciplinary Studies in the Humanities*, 7(3), 135-156. <https://doi.org/10.7508/isih.2015.27.006>
- Wittgenstein, L. (1953). *Philosophical Investigations* (G. E. M. Anscombe, Trans.). Blackwell.