



Original-Forschungsarbeit

Dual-Räumlichkeitsbildung der Intelligenz: Eine theoretische Retrodution der Sozialisierung künstlicher Intelligenz in der Bedeutungsproduktion

Manijeh Akhavan¹, Saied Reza Ameli^{2*}, Maseud Rahgozar³, Ehsan Shahghasemi⁴

1 Ph.D. candidate in Media Management, Kish International Campus, University of Tehran, Kish, Iran

2 Prof. of Communications and Global Studies, Department of Communications, Faculty 3 of Social Sciences, University of Tehran, Tehran, Iran (Corresponding author)

4 Associate Professor of Computer Science, School of Electrical & Computer Engineering, University College of Engineering, University of Tehran, Tehran, Iran

5 Associate Professor of Communication, Department of Communications, Faculty of Social Sciences, University of Tehran, Tehran, Iran

Empfangen: 5. Februar 2025 Akzeptiert: 3. Juni 2025

Zusammenfassung:

Fast fünf Jahrzehnte, nachdem Hubert Dreyfus die Bedeutung der Berücksichtigung des sozialen Charakters von Intelligenz bei der Entwicklung künstlicher Intelligenz betont hatte, und trotz der Konsolidierung von KI als aktant innerhalb der Nachrichtenmedien, haben sich praktische Implementierungen schneller entwickelt als die entsprechenden theoretischen Untersuchungen. Da die Fähigkeit zur Bedeutungsbildung innerhalb einer sozialen Institution die Sozialisierung eines kognitiven Systems voraussetzt, kann die Sozialisierung künstlicher Intelligenz entlang eines vergleichbaren Pfads wie die natürliche menschliche Intelligenz untersucht werden. Auf dieser Grundlage untersucht der vorliegende Artikel die Prozesse, durch die KI sozialisiert wird, um eine Rolle in der Bedeutungsproduktion innerhalb einer sozialen Institution wie den Nachrichtenmedien einzunehmen, und behandelt die zentrale Frage: Was ist sozialisierte künstliche Intelligenz? Zu diesem Zweck integriert die Studie das Modell der Dual-Räumlichkeitsbildung der Intelligenz mit der Repräsentationstheorie in einem sozio-organisationalen Rahmen und verwendet einen retroduktiven theoretischen Ansatz zur Beantwortung der Forschungsfrage. In dieser Analyse wird die soziale Ordnung als Funktion des Sozialisierungsprozesses von KI verstanden. Die Dual-Räumlichkeitsbildung der Welt führt folglich zu einer dual-räumlichen sozialen Ordnung. Die Ergebnisse zeigen, dass KI entweder so gestaltet werden kann, dass sie bestehende Wissensformen und verankerte soziale Stereotype ähnlich wie menschliche Kognition repliziert, oder dass sie sozial reguliert wird, um eine algorithmische Rationalität zu fördern, die auf das Gemeinwohl und die Verwirklichung einer nachhaltigen und gerechten sozialen Ordnung ausgerichtet ist. Eine solche Ordnung hängt von der Öffnung repräsentationaler Praktiken durch reflexives Engagement mit sozialen Stereotypen ab, wodurch Transformationen in der Repräsentation und eine größere Vielfalt von Identitäten unterstützt werden. Der Beitrag dieses Artikels liegt in der Vorschlag eines integrierten Modells zum Verständnis der Mechanismen der Sozialisierung von KI in bedeutungsproduzierenden sozialen Institutionen. Darüber hinaus bietet das Modell eine umfassende Perspektive auf die Sozialisierung sowohl natürlicher als auch künstlicher kognitiver Systeme innerhalb der sich entwickelnden Strukturen dual-räumlicher institutioneller sozialer Ordnungen.

Schlüsselwörter: künstliche Intelligenz, dual-räumliche Bildung der Intelligenz, Repräsentation, Sozialisierung, soziale Ordnung

* Korrespondierender Autor

✉ ssameli@ut.ac.ir

🌐 <https://orcid.org/0000-0002-0193-4302>

Wie dieser Artikel zu zitieren ist:

Akhavan, M., Ameli, S.R., Rahgozar, M., & Shahghasemi, E. (2025). Dual-spacization of intelligence: A theoretical retrodution of the socialization of artificial intelligence in meaning construction. *Spektrum Iran*, 38(2), 1-31.

🔗 <https://doi.org/10.22034/spektrum.2026.565209.1055>

© 2025 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

هوش دوفضایی؛ پس کاوی نظری اجتماعی شدن هوش مصنوعی برای تولید معنا در رسانه‌های خبری

منیژه اخوان^۱، سعیدرضا عاملی^{۲*}، مسعود رهگذر^۳، احسان شاه‌قاسمی^۴

۱ دانشجوی دکتری مدیریت رسانه، پردیس بین‌المللی کیش، دانشگاه تهران، تهران، کیش، ایران

۲ استاد ارتباطات و مطالعات جهانی، دپارتمان ارتباطات، دانشکده علوم اجتماعی، دانشگاه تهران، تهران، ایران (نویسنده مکاتبه‌ای)

۳ دانشیار علوم کامپیوتر، دانشکده مهندسی برق و کامپیوتر، پردیس فنی دانشگاه تهران، تهران، ایران

۴ دانشیار ارتباطات، دپارتمان ارتباطات، دانشکده علوم اجتماعی، دانشگاه تهران، تهران، ایران

دریافت: ۱۴۰۳/۱۱/۱۶؛ پذیرش: ۱۴۰۴/۰۳/۱۳

چکیده:

هرچند حدود نیم قرن از زمانی که هوبرت دریفوس ضرورت توجه به ماهیت اجتماعی هوش را در توسعه مدل‌های هوش مصنوعی مطرح کرد، می‌گذرد و هوش مصنوعی نیز جایگاه خود را به عنوان یک بازیگر فناورانه جدید در رسانه‌های خبری تثبیت کرده، به نظر می‌رسد پیاده‌سازی عملی سریع‌تر از مطالعات نظری آن پیشرفت کرده‌اند. از آنجا که ظرفیت معناسازی در یک نهاد اجتماعی، اجتماعی شدن یک سیستم شناختی را پیش‌فرض می‌گیرد، اجتماعی شدن هوش مصنوعی را می‌توان در مسیری قابل مقایسه با هوش طبیعی انسان بررسی کرد. بر این اساس، مقاله حاضر فرآیندهایی را بررسی می‌کند که از طریق آنها هوش مصنوعی اجتماعی می‌شود تا نقشی معنا ساز در یک نهاد اجتماعی مانند رسانه‌های خبری ایفا کند و به این سوال اصلی می‌پردازد: چه چیزی هوش مصنوعی اجتماعی شده را تشکیل می‌دهد؟ بدین منظور، از نظریه هوش دوفضایی و نظریه بازنمایی در رویکرد اجتماعی-سازمانی استفاده شده و در چارچوب رویکرد پس‌کاوی نظری به مسئله پاسخ داده می‌شود. در این پژوهش، نظم اجتماعی به عنوان کارکرد فرایند اجتماعی شدن هوش مصنوعی در نظر گرفته شده است. با دوفضایی شدن جهان، ما با نظم اجتماعی دوفضایی مواجه هستیم. یافته‌های این مطالعه نشان می‌دهد که هوش مصنوعی می‌تواند هم دانش جهان و کلیشه‌های اجتماعی آن را بازتولید کند و عملکردی شبیه انسان داشته باشد یا طوری اجتماعی و تنظیم شود که عقلانیت معطوف به خیر جمعی و نظم اجتماعی پایدار و عادلانه را به ارمغان آورد. چنین نظمی، مستلزم آشکار کردن کلیشه‌های اجتماعی از درون است که می‌تواند روی تغییر وضعیت بازنمایی در مسیر تنوع هویتی اثرگذار باشد. نوآوری این مقاله، ارائه مدلی یکپارچه برای فهم مکانیسم‌های اجتماعی شدن هوش مصنوعی در همه نهادهای اجتماعی معنا ساز است. همچنین این مدل، رویکردی جامع به اجتماعی شدن سیستم شناختی طبیعی و مصنوعی در بستر تغییرات ساختارهای اجتماعی نهادی دوفضایی دارد.

واژگان کلیدی: هوش مصنوعی، هوش دوفضایی، بازنمایی، اجتماعی شدن، نظم اجتماعی



Original Research Paper

Dual-spacization of intelligence: A theoretical retrodution of the socialization of artificial intelligence in meaning construction

Manijeh Akhavan¹, Saied Reza Ameli^{2*}, Maseud Rahgozar³, Ehsan Shahghasemi⁴

1 Doktorandin im Medienmanagement, Kish International Campus, Universität Teheran, Kish, Iran

2 Professor für Kommunikations- und Global Studies, Abteilung für Kommunikation, Fakultät für Sozialwissenschaften, Universität Teheran, Teheran, Iran (Korrespondenzautor)

3 Außerordentlicher Professor für Informatik, Fakultät für Elektro- und Computertechnik, University College of Engineering, Universität Teheran, Teheran, Iran

4 Außerordentlicher Professor für Kommunikation, Abteilung für Kommunikation, Fakultät für Sozialwissenschaften, Universität Teheran, Teheran, Iran

Received: Feb., 5, 2025 Accepted: Jun. 3, 2025

Abstract

Nearly five decades after Hubert Dreyfus underscored the importance of accounting for the social character of intelligence in the development of artificial intelligence, practical implementations have progressed more rapidly than corresponding theoretical inquiry. This remains the case notwithstanding artificial intelligence's consolidation as an actant within the news media. Because the capacity for meaning-making within a social institution presupposes the socialization of a cognitive system, the socialization of artificial intelligence may be examined along a trajectory comparable to that of human forms of natural intelligence. On this basis, the present article investigates the processes through which AI becomes socialized so as to assume a meaning-making role within a social institution such as the news media, addressing the central question: What constitutes socialized artificial intelligence? To this end, the study integrates the Dual-spacization of Intelligence with representation theory within a socio-organizational framework and adopts a retroductive theoretical approach to address the research question. Within this analysis, social order is understood as a function of AI's socialization process. The dual-spacization of the world consequently gives rise to a dual-spatial social order. The study's findings suggest that AI may either be engineered to replicate existing forms of knowledge and entrenched social stereotypes in a manner analogous to human cognition, or be subject to social regulation that fosters an algorithmic rationality oriented toward the common good and toward a sustainable and just social order. Such an order depends on opening representational practices through reflexive engagement with social stereotypes, enabling transformations in representation and supporting increased diversity of identities. The contribution of this article lies in proposing an integrated model for understanding the mechanisms of AI socialization across meaning-producing social institutions. Furthermore, the model offers a comprehensive perspective on the socialization of both natural and artificial cognitive systems within the evolving structures of dual-spatial institutional social orders.

Keywords: artificial intelligence, dual-spacization of intelligence, representation, socialization, social order

* Corresponding Author

✉ ssameli@ut.ac.ir

🌐 <https://orcid.org/0000-0002-0193-4302>

How to Cite this Article:

Akhavan, M., Ameli, S.R., Rahgozar, M., & Shahghasemi, E. (2025). Dual-spacization of intelligence: A theoretical retrodution of the socialization of artificial intelligence in meaning construction. *Spektrum Iran*, 38(2), 1-31.

🔗 <https://doi.org/10.22034/spektrum.2026.565209.1055>

© 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

1. Introduction

In recent decades, news organizations have increasingly become leading actors in integrating artificial intelligence strategies into their structures, driven by pressures such as access to vast amounts of data, advances in AI technologies, market pressures, imitation of pioneering news giants in adopting AI, technological competition, and the need to respond to audiences' informational needs. For instance, a 2023 global survey examining AI adoption across 105 news and media organizations in 46 countries reports that more than 75 percent of the surveyed organizations employ AI in at least one segment of the news value chain (Beckett & Yaseen, 2023, p. 6).

Since news organizations are among the key social institutions that help sustain a society's social order by making the world meaningful (Rohlinger, 2022), the integration of AI into their value chains implies that artificial cognitive systems now participate alongside natural cognitive systems in meaning production. For humans, acquiring a meaning-making role in news media requires socialization (Breed, 1955). Given that AI—developed through the logic of simulating human intelligence (Ameli, 2021) and now encompassing deep learning and large language models—operates on cognitive principles analogous to natural intelligence, its socialization processes can likewise be examined anthropomorphically (Collins, 2025). Accordingly, this article seeks to understand the mechanisms through which AI becomes socialized.

Along this path, one dimension involves recognizing that AI's capacity for meaning-making depends on its ability to perceive the world, a process that occurs through representation (Sokol & Flach, 2024; Huh et al., 2024). Comprehending both the world in which new communication technologies are socialized and the structural capacities through which it receives and represents that world allows for the formulation of realistic expectations regarding the adoption of ICTs' (see for example Sabbar and Matheson, 2019). In addition, identifying the assemblage of forces that shape the socialization of artificial cognitive systems along their trajectory toward acquiring a meaning-making role is crucial for conceptualizing artificial intelligence as a social phenomenon (Lewis & Westlund, 2014).

On the other side of this trajectory lies socialized AI, whose function is to help maintain social order. With the emergence of the simultaneous

communication industry (Ameli, 2011) and the advancement of digital technologies, this concept requires renewed examination. On one hand, the rise of virtual space parallel to the physical world has expanded the "social space" (Ameli, 2011, p. 2), and the profound intertwining of this new space with the physical world has transformed virtual space into a "second life-space" (Ameli, 2017, p. 192). This condition marks the extension of the human life-ecosystem into digital environments, described as the dual-spacization of the world (Ameli, 2003a, 2003b; Ameli, et al., 2024). On the other hand, AI possesses an industrial nature (Ameli, 2021). The syntactic command structure of algorithms (Goffey, 2008), along with their goal-oriented nature and programmability (Fin, 2017), establish a precise order in the pursuit of clearly defined objectives. These new communication technologies have altered the way social phenomena are perceived (Shahghasemi et al., 2025). Consequently, some scholars argue that attention should shift toward the "technology's role in engineering the social world" (Couldry, 2021, p. 11) or "algorithmic regulation" (Ameli, 2023). Therefore, understanding AI socialization requires situating it within the evolving coordinates of social order in a dual-spatial society.

Accordingly, this article addresses the question of what constitutes socialized AI and how AI becomes socialized to attain a meaning-making role in news media—and, more broadly, within any meaning-producing social institution. To answer this question, the study draws on dual-spacization of intelligence theory and the theory of representation, employing a retroductive theoretical approach.

The novelty of this article lies in presenting a model for understanding the mechanisms for socialization of neural network-based AI (including deep learning systems and large language models), a model applicable to all meaning-producing social institutions. Additionally, the proposed model provides a comprehensive perspective on the socialization of both natural and artificial cognitive systems within the context of transforming dual-spatial institutional structures.

2. Literature Review

New communication technologies have created a significant change in how social phenomena are perceived (Schatzki, 2025; Shahghasemi, 2025;

Hohenstein et al., 2023; Kubin & von Sikorski, 2021; Bytiak et al., 2020; Yıldız & Nur, 2024; Nourbakhsh & Nemati, 2020). The study of mechanisms through which AI becomes socialized to attain a meaning-making role in news media can be traced back to the tradition of sociology of newswork, which focus on newsroom practices, relationships among actors involved in news production, organizational structures, and the norms and values underlying the news-making process (Belair-Gagnon, 2019). Although the historical foundations of this perspective can be traced to the classical works of Max Weber (1976) and the Chicago School (Hughes, 1940), since the end of World War II, the everyday interactions of newswork have become a central concern within sociology. These interactions have been examined through five major traditions that have emerged sequentially over time (Belair-Gagnon, 2019): (1) the gatekeeping approach, which foregrounds mechanisms of social control within the newsroom (Lewin, 1947; White, 1950); (2) the organizational approach, which emphasizes organizational structures, routines, and decision-making processes (Tuchman, 1978; Gans, 1979); (3) the political economy approach, which analyzes the impact of market forces and media ownership on news production (Herman & Chomsky, 1988; Bourdieu, 1998; Hallin & Mancini, 2004); (4) the cultural studies approach to news, which focuses on the social construction of news within relations of power (Hall, 2003; Revers, 2017); and (5) the network approach, which examines the incorporation of digital technologies into the newsroom and the resulting transformations of its ecosystem (Boczkowski, 2005; Anderson, 2013; Groves & Brown, 2015; Shokrkhah, 2018; Ferrucci et al., 2022).

Attention to the socialization of newsroom workers dates back to Warren Breed's seminal work (1955) within the early sociology of news. Breed explains that news organizations, in order to structure the news production process and ensure journalists' conformity to editorial policies, must socialize them effectively. In the organizational tradition, the work of Gravengaard and Rimestad (2014) is noteworthy. By combining newsroom ethnography, linguistic anthropology, conversation analysis, and theories of profession, they examine how journalism trainees become socialized into a particular professional culture and community of practice through interactions with editors.

Within the network approach, only a limited number of studies have directly addressed the socialization of newsroom personnel. For example, Singer (2004), working within the convergence approach and studying four converged newsrooms, argues that convergence acts as a catalyst that redefines the socialization of print journalists and disrupts their perceptions of their professionalism. Sylvie (2018) demonstrates how established roles within newsrooms are being reconfigured, with leadership shifting toward managerial functions and editorial roles evolving into facilitation. Sissons and Smith (2026), using discourse analysis of interactions among journalists in three newsrooms, explore how new technologies have reshaped the identities of newswriters.

Research on the socialization of AI as a novel technological actor within the newsroom remains limited. Among the few existing studies, Dodds et al. (2025) highlight the managed integration of generative AI into everyday newsroom routines. Based on interviews with editors, journalists, and innovation managers, and grounded in a theoretical framework of professional authority, their study proposes mechanisms through which generative AI can be incorporated into a controlled manner into newsroom practices. Another pertinent contribution is provided by Collins (2025), who explores the socialization of natural and artificial intelligence at the societal level, but does not extend the analysis to the level of social institutions.

A review of the literature reveals that, regarding the nature of *socialized artificial intelligence*, no study has yet examined how an AI model becomes socialized to acquire a meaning-making role within a news organization. To address this theoretical gap, the present article adopts a combined organizational and cultural-studies perspective within the broader framework of the dual-spacization of the world to investigate this question.

3. Theoretical Framework

Dual-spacization of intelligence theory (Ameli, 2021) is situated within the broader paradigm of the dual-spacization of the world advanced by Saied Reza Ameli (2003a, 2003b). This paradigm conceptualizes phenomena as dual-spatial entities operating across two interconnected domains—the physical world and the virtual world—both of which have become integral

to everyday and professional practices (Ameli, 2011). Within this framework, the dual-spacization of the world entails the dual-spacization of intelligence itself, such that artificial intelligence, in parallel with human natural intelligence, extends and reconfigures human cognitive and feedback system across both spatial domains.

The theory provides both a theoretical model and a conceptual system for understanding the nature of artificial intelligence. Within its theoretical model, the defining axes of AI include:

1) The real-virtual nature of virtual space, within which an AI model is developed and which shapes the model's cognitive boundaries.

2) The applied contexts and the operational domain of AI, which are related to the limiting boundaries that ensure AI performance remains aligned with its function and are based on the following dualities: A) Traditional vs. modern, referring to ethical codes; B) Local vs. global, referring to institutional capacities; and C) Real vs. virtual, referring to hardware, software, and data-related capacities based on modularity, layering, integration, data-drivenness, networked architecture, and algorithmicity (Ameli, 2021);

3) The interconnection between artificial and natural intelligence, may take one of the following modes: A) Exclusivity of physical-space capacities; B) Replication of physical-space capacities in virtual space; C) Integration of physical and virtual capacities; D) Replication of virtual-space capacities in physical space; and E) exclusivity of virtual-space capacities (Ameli, 2017). This relation and interconnection are determined based on the performance similarities and differences between human and artificial intelligence in relation to their function.

The conceptual system of dual-spatial intelligence is a triple classification system that operates on a real/virtual binary. In this system: placing "real" or "virtual" in the first slot refers to the nature of intelligence; placing "real" or "virtual" in the second slot refers to the operational domain of intelligence; and the third slot concerns the interconnection between intelligence in the virtual and physical worlds.

Table 1. Triple classification system of dual-spatial intelligence

Binary	Slot 1	Slot 2	Slot 3
Real	Intelligence perceived as real through sensory experience and producing real effects in the physical world	Intelligence operating in the physical world	Performance similarity between artificial and natural intelligence
Virtual	Intelligence understood as a mental or symbolic construct	Intelligence operating in the virtual world	Performance difference between artificial and natural intelligence

Since intelligence is defined as a "simple relationship between what we want, what we perceive, and what we do" (Russell, 2019, p. 21), the similarities and differences between natural and artificial intelligence are also understood within this framework. Goal formation is considered within the framework of Russell and Norvig's matrix (2022), encompassing systems that think/act like humans as well as systems that think/act rationally. World perception and computational and processing capacities are similarly considered core functions of intelligence. However, this article focuses on perception. Since world perception occurs in both natural and artificial cognitive systems during the process of representation, Stuart Hall's theory of representation has been employed within a socio-organizational approach.

Hall (2003, p. 15) defines representation as "the use of language to say something meaningful about, or to represent, the world meaningfully, to other people [within a culture]." He identifies three major approaches to how representation works through language (cited in Ameli & Merali, 2004): 1) Reflective, where language mirrors the inherent meaning of an object; 2) Intentional, where language imposes the speaker's intended meaning onto the world; and 3) Constructionist, where meaning is socially constructed or produced¹ through language and shared among members of a culture.

Hall (2003, p. 1) emphasizes the constructionist approach, viewing representation as the process that "connects meaning and language to culture". His framework centers on three key questions: Where is meaning constructed? How is meaning constructed and circulated? And Who determines and fixes meaning?

1. We use the terms construct and produce interchangeably when referring to meaning

To address the first question, Hall (2003, pp. 3–4) introduces the "circuit of culture", explaining that meaning is produced and circulated at multiple sites—representation, identity, production, consumption, and regulation. Regarding the second question, he argues that members of a culture share common cultural codes, expressed through linguistic codes, enabling similar interpretations. These linguistic codes consist of material elements (such as sounds and words) that function as signs representing concepts, thereby enabling participants to encode and decode meaning. Hall draws on Saussure's sign-signifier-signified triad and Barthes' concept of second-order semiological systems, arguing that these relations are culturally, historically, and socially constructed (Nothias, 2020).

For the final question, Hall—drawing on Foucault's discourse theory—argues that discourses determine how people produce meaning and stabilize it (Hall, 1997). In Foucault's conceptualization, discourses are understood as modes of referring to or "constructing knowledge about a specific domain of practice: they constitute a cluster, or formation, of ideas, images, and practices that establish ways of speaking about, forms of knowledge concerning, and modes of conduct associated with a given topic, social activity, or institutional site within society" (cited in Hall, 2003, p. 6). Such statements collectively form a discursive formation that sustains particular institutional or political trajectories within a "system of dispersion." Importantly, discursive statements can only be produced by actors who possess the requisite positional authority to generate meaning (Hall, 1992, p. 202). Such formations ultimately establish a regime of truth, determining "... what knowledge is considered useful, relevant, and true in that context, and what sort of persons or subjects embody its characteristics" (Hall, 2003, p. 6). In media studies, Hall (1997, p. 20) argues that stereotypes function as mechanisms for fixing meaning. A stereotype is "a powerful way of circulating in the world a very limited range of definitions of who people can be, of what they can do, what are the possibilities in life, what are the natures of the constraints on them". Thus, the effort to break stereotypes means expanding the diversity of identities that individuals can experience or encounter.

In this article, the theoretical retroductive approach (Bygstad & Munkvold, 2011) is employed to address the research question. This approach, a qualitative and interpretive research strategy, aims at providing a "theoretical explanation that proceeds by description of significant features,

retroduction to possible causes, elimination of alternatives and identification of the generative mechanism or causal structure at work" (Bhaskar, 1998: xvii). Conducting research based on this approach typically unfolds in six steps: "1) description of the event; 2) identification of key components; 3) theoretical re-description (abduction); 4) retroduction (identification of candidate mechanisms); 5) analysis of selected mechanisms and outcomes; and 6) validation of explanatory power" (Bygstad & Munkvold, 2011, p. 5).

4. Findings

Before defining socialized AI, it is necessary to determine the function and meaning of socialness within a cognitive system. Researchers argue that the function of socialness in artificial intelligence is social order (Couldry, 2021). In classical studies, despite different approaches (Silver and Clark, 2008), two basic dimensions are typically proposed as pathways toward achieving social order: "coordination of actions", and "cooperation to attain common goals" (Hechter & Horne, 2009, p. 1). Additionally, conformity is proposed as "a solution to the problem of social order," emerging through socialization (Rydgren, 2008, p. 71).

The socialness of a natural cognitive system, drawing on a reinterpretation of Collins's (1998, p. 479) classical definition within the theoretical framework of this article, is defined as the capacity of a cognitive system to attain "social fluency" within one or more institutional discourses. Social fluency is further conceptualized, in alignment with Ludwig Wittgenstein's perspective (Magee, 2019; Collins, 2025), as the ability of a natural cognitive system to learn – through everyday life – the language and its practical applications for meaning-making in accordance with the role it occupies within the institutional division of labor. This learning process occurs through interactions both with members of a culture and with the "Originators" and "Speakers" of a given discourse (Hall, as cited in Griffin, 2012, p. 347). Accordingly, the socialization of a natural cognitive system refers to the processes through which it achieves social fluency. On the basis of this definition, two key propositions follow: 1) socialness pertains to collective tacit knowledge, that is, knowledge acquired through social experience; and 2) socialness is defined in relation to actors' social roles within institutional divisions of labor (Berger & Luckmann, 1966).

Within this perspective, a cognitive system does not become social merely by learning linguistic sign systems; rather, it must also learn how to use language as a tool in various social roles. Therefore, in this article, language is considered in two analytical dimensions: language as a tool and language as a sign. Learning language as a sign system enables members of a culture to share a common understanding of correctness and appropriateness within discursive formations and to engage in meaningful communication. Learning language as a tool, in turn, expands the range of activities that individuals can perform in different social roles. Moreover, language as a tool is not solely experiential; in educational processes, specialized terminologies emerge in the form of metaphors, images, humor, rumors, and other communicative devices that individuals acquire while learning professional skills (Breed, 1955; Berger & Luckmann, 1966; Talebi Tadi & Rastegar Khaled, 2025). Accordingly, the nature of socialized AI is articulated within the socio-organizational approach of the dual-spatial intelligence model.

4.1. Socialization of natural intelligence in the shared real world / socialization of AI in the possible datafied world

In the dual-spatial intelligence model, the cognitive resources of artificial and natural intelligence are intertwined. However, the worlds within which they become socialized are different. The socialization of natural intelligence takes place in a single physical world commonly shared by all members of the human species. By contrast, the socialization of artificial intelligence takes place within a datafied possible world that is constructed and regulated by humans through the selective extraction of data from the real world. Decisions concerning which data are selected are shaped by data acquisition processes that are, in turn, conditioned by factors such as data availability (IBM, n.d.), the technical, legal, security, political, and economic dimensions of data accessibility (Mohseni Ahooei, 2024; Salehi et al., 2026), algorithmic mechanisms that extract data from users on the basis of legibility (Gillespie, 2014), and socio-cultural patterns of participation as well as the representativeness of target populations (Ade-Ibijola & Okonkwo, 2023).

4.2. The applied contexts and operational domain of AI and Dual-spatial structural capacities of AI socialization

According to the dual-spatial intelligence model, the applied contexts and operational domains that guide the socialization of an artificial cognitive system toward social order are as follows:

Moral codes: An artificial intelligence model is socialized into a culture when it acquires that culture's moral codes at societal, institutional, and content levels. Moral codes establish what is true or false – or what is considered reality – in the real world (Collins, 2025). They correspond to the second-order semiotic system of that culture and, in accordance with representation theory, are produced and fixed through discourses and stereotypes.

Institutional considerations in the integration of AI into the value chain: The capacities of a public, private, or governmental institution that deploys an AI model within its value chain include: 1) procedures – structured patterns of recurrent processes within the institution; 2) principles of professional ethics – globally standardized normative statements about proper conduct within a profession, defined by professional bodies to safeguard public trust and credibility (Mackenroth, 2025); 3) corporate social responsibility – institutional commitments to society and the public interest, regulated by national legal frameworks.

Structural capacities regulating AI socialization in virtual space: Structural capacities include the hardware, software, and data that shape the social structure of the datafied world and define the applied contexts of socialized AI. Hardware capacities, the technical substrate of AI, include both immaterial elements such as protocols, standards, and security (Ameli, 2025), and material infrastructures such as digital connectivity, GPUs, semiconductors, and AI supercomputers (Tortoise, 2024). Software and data capacities are treated jointly because algorithms derive meaning only insofar as they are paired with databases; without such pairing, they remain “meaningless machines” (Gillespie, 2014: 4). Data must be gathered and prepared before algorithms can operate. Social considerations in data gathering were discussed earlier. Two social considerations arise in the context of data preparation: A) The logic of data categorization in databases based on the relational ontology of the datafied world: The ability to mentally categorize the world is a genetic capacity in humans, one with a long intellectual history (Hall, 1997; Shahghasemi, 2025). Moreover, societies have categorized data for millennia – for instance, through censuses – whose social implications for constructing reality and enabling state control have been recognized since the late twentieth century (Scott, 1998). It is also argued that in artificial cognitive systems, data categorization has the capacity to construct reality according to algorithmic logic (Cox, 2022). Scholars show

that relational database architectures produce a "relational ontology that understands data as atomized, regular, uniform and only loosely connected objects that can be ordered in a potentially unlimited number of ways at the time of retrieval" (Rieder, 2012; cited in Gillespie, 2014, p. 6). These data have no inherent meaning and only acquire significance when algorithms retrieve them through queries based on co-occurrence patterns. Here, in addition to algorithms, the ways in which databases are categorized, which data are placed in each category, and who decides how these categorizations are implemented all constitute powerful claims about truth in the world (Gillespie, 2014). B) soft and hard filtering of data: A second consideration involves integrating principles of professional ethics and corporate social responsibility into explicit "if-then" rule-based structures within the code. The two primary mechanisms of data preparation are: exclusion, which excludes certain data from training sets (hard filtering), and demotion, which reduces weights within neural networks (soft filtering) (Gillespie, 2014). The social dimension here concerns the engineers developing AI models. They become part of institutional work routines despite not necessarily possessing the specialized language, skills, or professional ethical knowledge of that institution. Consequently, cooperative goal-oriented work may rely on various forms of negotiated social agreement.

4.3. Dual-spatial processes of AI socialization in relation to natural intelligence

The socialness of a natural cognitive system can be understood as the aggregate learning of 1) language as natural signs and second-order signs, and 2) language as a tool. In a socio-organizational approach, this equation depends on the social distribution of knowledge and the social division of labor, where, at the intersection with social order, the social distribution of knowledge is linked to coordinative action, and the division of labor is connected to cooperation to attain common goals. Following Collins (2025), we designate language as natural signs as natural language, language as second-order signs as moral codes, and language as a tool as experience.

Accordingly, we define the relationship between natural and artificial socialization across three levels: primary, secondary, and tertiary socialization (Collins, 2025). The socialization processes are explained through three classical theories: Vygotsky's (1978) theory of cognitive

development at the individual level; Berger and Luckmann’s (1966) theory of the social construction of reality at the institutional level; and Breed’s (1955) model of newsroom socialization. These theories align with the conceptual foundation of this article by emphasizing the social origins of socialization.

It is also necessary, across all stages of socialization, to identify the assemblage that orients a natural or artificial cognitive system toward social order. Within the dual-spatial intelligence model, this assemblage consists of people, processes, things, and data (Ameli, 2025). Actor–Network Theory proposes a comparable configuration, comprising actors, actants, activities, and audiences (Lewis & Westlund, 2014). Table 4 illustrates the correspondence between these elements in the contexts of natural and artificial intelligence.

Table 4. Assemblage guiding the socialization of natural and artificial intelligence toward social order

	Natural Intelligence		Artificial Intelligence	
People	Society	dual-spatial societal members	Actors at societal level	Audiences
	Institutions	Responsible others/ Originators (e.g., news publishers)	Actors at institutional level	Responsible others/ Originators
		Significant others (e.g., newsroom staff)		Significant others
Processes	Job roles		tasks of an AI model within institutional workflows	
Things	Things whose meaning is constructed through everyday dual-spatial interactions		Actants (algorithms, code, CMS, etc.)	
Data	Knowledge about the world and its stereotypes, mediated through language			

4.4. The three stages of socialization of natural intelligence

This section outlines the three stages of natural intelligence socialization based on the dual-spacization of intelligence theory and a re-reading of representation theory.

Primary socialization of natural intelligence in the dual-spatial family:

The socialization process of the natural cognitive system begins at birth and continues until the child enters school. In this stage, parents function as responsible others, drawing on the authority they derive from the discourse of the family institution. Based on various factors—including their social position, embodied hereditary social attributes, and their personal beliefs and experiences—they select and interpret aspects of the world for the child (Berger & Luckmann, 1966). According to social theories of intelligence, although the child is born with biological capacities for adapting to the social world, they cannot initially use linguistic signs and rely instead on biological eidetic imagery. Gradually, through interaction with parents, the child begins to use external action signs (Vygotsky, 1978). In the process of identification, the child subsequently internalizes the mental meanings of those external signs as interpreted by the parents (Tomasello, 1999). As development continues, the child learns the "native language" through repeatedly hearing stable descriptions of things by parents in various contexts. In this way, the child comes to believe in a stable world in which things are always described in the same way. This generates a sense of truth about the world, and later these "moral concepts" constitute the individual's "moral compass" throughout life (Collins, 2025: 1256). As a result, the truth-world internalized during primary socialization becomes highly resistant to collapse. Through identification with parents, the child likewise adopts their beliefs concerning social roles and thereby develops a coherent sense of self. Ultimately, the child becomes capable of co-creating new shared meanings with parents, marking the completion of primary socialization. As development continues and the child interacts with other significant figures, such as grandparents, consciousness progressively abstracts from truths tied to particular objects and role expectations toward more generalized societal truths and roles (Berger & Luckmann, 1966).

At this stage, what is crucial is that with the dual-spacization of the world, we now encounter the dual-spatial family (Ameli, 2011). In this new condition, two major transformations have occurred in the communicative system of the family: "the transformation of interactional space and social communication with the outside world; and the transformation of accessibility to the outside world on a global scale" (Ameli, 2011: 161). These shifts have produced cultural multiplicity within families, plurality of

culture-shaping sources inside the home, transformations in sources of personal self-understanding, the loss of wide shared cultural domains with the broader society, value relativization, individualization, and the expansion of independent personal spheres (Ameli, 2011). Together, they have three major implications for child socialization: 1) moral concepts within the family may no longer align with societal moral codes; 2) members of a dual-spatial society may hold heterogeneous moral concept systems; and 3) the expansion of independent personal spheres has shifted parent-child communication toward “communicative interactions” (Ameli, 2011: 162–163). In such a situation, the family’s capacity to produce conformity is significantly challenged.

Secondary socialization of natural intelligence under conditions of global access: Secondary socialization begins when the individual enters school and continues until the end of the schooling period. During this stage, the individual acquires “written language” and “early technical concepts” (Collins, 2025: 1253). According to the cognitive development theory, some mental functions develop in school through collaboration with more capable peers and under the guidance of teachers. On this basis, learning the early technical concepts of any subject provides the foundation for more complex mental processes. Furthermore, through acquiring written language, the student masters linguistic symbols and meanings at the societal level (Vygotsky, 1987). Importantly, the foundational reality is what the individual internalized during childhood, and every subsequent content that becomes internalized is built upon this pre-existing reality (Berger & Luckmann, 1966). Students further acquire social stereotypes associated with their roles, along with the requisite social skills, through normative components, mechanisms of reward and punishment, and institutional rituals. These are transmitted through the ways teachers address them, their interactions with classmates and other students in the school, and their membership in various groups, all of which contribute to the construction of their identity (Collins, 2025).

At this stage, the dual-spacization of social structures has resulted in the global expansion of human development, worldwide pathways of access, human connectivity to all global capacities, networked linkage to all phenomena, and the translocalization of all levels of communication (Ameli, 2011). In such a world, on the one hand, various social institutions constantly compete to define the “truth” of the world from their own perspectives,

cultivating in the individual a heightened sense of the instability of meanings (Collins, 2025). On the other hand, individuals may join online subgroups outside their own society – or marginal subgroups within it – that operate with different moral codes. This can give rise to identity and cultural heterogeneity and hybrid identities (Ameli, 2011).

Tertiary socialization of natural intelligence in virtual work: Tertiary socialization begins when an individual enters university and continues until they attain a position in which they participate in meaning-construction within a specialized profession in an institutional subworld such as a news agency. In this stage, the individual first learns the “specialized language” and “professional skills” of a profession at the university from responsible others (academic instructors) and internalizes them through interaction with classmates (significant others). During the acquisition of professional skills, the student also learns the idioms, metaphors, and affective tones associated with that occupational role. Subsequently, upon entering an institutional subworld and interacting with responsible others (managers, senior staff) and significant others (colleagues), the individual learns – through various mechanisms – what counts as valid world knowledge and acceptable social stereotypes within that institutional discourse (Breed, 1955; Collins, 2025: 1253).

As noted earlier, in a dual-spatial world, social structures themselves have become dual-spatialized; thus, we encounter "virtual work". The concept of virtual work was first introduced in the paradigm of the dual-spacization of the world by Saeid Reza Ameli (2006a). In this paradigm, work in the virtual world transforms from an activity tied to a specific physical location into one detached from place, giving rise to the notion of remote work (Ameli, 2011: 293). In such a context, conformity may emerge in a “compliance” form – behavioral conformity (Rydgren, 2009: 85) – which we refer to as second-order conformity. Here, the individual adapts to certain discursive meanings in order to benefit from the advantages of belonging to an institutional subworld, such as income or occupational status.

4.5. Three stages of AI socialization

This section outlines the three stages of AI socialization based on the dual-spacization of intelligence theory and a re-reading of representation theory.

Primary socialization of AI and task- and goal-oriented prior configuration: Currently, the primary socialization of AI occurs in the form of prior configurations, that is, before a model is even constructed, its initial settings are established. Temporally, this process spans from the design phase for developing a model to the commencement of model learning. At this stage, institutional speakers, AI specialists, and business professionals (Lewis & Westlund, 2014) engage, prior to model construction, in agreements regarding the creation of a potential data world and the engineering of work processes, based on technological, legal, security, economic, political, and socio-cultural considerations, as well as on the intended tasks and objectives. At this stage, AI specialists, within the framework of a given proposal, construct a data world and configure actants. Since AI models are not inherently capable of directly understanding human natural language, they instead rely on a set of formal and computational languages. These languages function in a manner analogous to experimentation in human intelligence, serving as a means for the system to process and interpret information in a structured, computationally defined manner. At this stage, the professional ethical principles and organizational codes of social responsibility of the model developer and of the institutional sub-world in which the model is intended to perform a task are incorporated into the system in the form of explicit if...then instructions. The choice of learning paradigm is also crucial for embedding ethical codes. In a general typology, four kinds of data-driven learning exist: Supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning (Sutton & Barto, 2014; Collins, 2025; Russell & Norvig, 2022; Danks, 2014).

Secondary Socialization of AI and Translocal Socialization: The second stage of AI socialization concerns model learning—when a model learns language (natural language and moral codes / syntax and semantics) and its uses in the real world from data so that it can produce meaning. At this stage, data enter the artificial neural network. For the network to process the input data must be converted into numerical form, a process carried out in the embedding layer. In this layer, textual, audio, or visual data are first broken into smaller meaningful units—such as words (or tokens) in textual data (Salehi et al., 2026). Each word is then represented as a numerical vector in a continuous vector space, where each number in the vector denotes a semantic feature corresponding to some dimension of knowledge about that data

(Keikha & Rahgozar, 2018; Babić et al., 2020). These vectors initially contain random numerical values; thus, they are arrays of meaningless numbers. Gradually, during model learning, their values are adjusted to become meaningful numerical representations (Yadav, 2024). This numerical adjustment is collective in nature, meaning that all elements of a data-world participate in the socialization of an artificial cognitive system. Other layers of the neural network are similarly constructed with randomly initialized weights among artificial neurons arranged in successive layers. Input data, in vector form, pass through these layers sequentially, and at each layer more complex patterns – such as syntactic structures or sentiment – are extracted. At the end of this process, the network produces an output, which is evaluated and accepted or rejected through direct, implicit, or reinforcement-based supervision (Collins, 2025). From this perspective, through weight adjustment and progressive pattern extraction, an artificial cognitive system learns human linguistic signs within the discursive configuration of the broader society from which the training data are derived. It is at this point that we may say the system has become socialized.

Accordingly, secondary socialization in AI is described as “translocal socialization” (Ameli, 2021) – meaning that the model begins to learn about the world based on training data from all societies that use the internet. Two points arise here: If the training data come from multiple societies and the learning paradigm is semi-supervised or unsupervised, then, because societies have different regimes of truth, what is considered acceptable about the world becomes entangled. Such entanglement is further intensified by structural epistemic asymmetries within global knowledge production. Drawing on interviews with Iranian sociologists, Hosseini and Sakhaei (2025) argue that AI systems are often experienced as carriers of dominant epistemologies while simultaneously functioning as sites of local reinterpretation and strategic agency. This tension between decontextualization and cultural specificity illustrates how AI socialization unfolds within hierarchically structured epistemic orders shaped by power, institutional infrastructures, and governance regimes. Additionally, Within each society, “core” and “Outliers” actors differ. Individuals with marginal behaviors or values may not be socially rewarded; however, on the internet it is often non-typical or unconventional views that gain traction. Hence, a model may statistically misclassify central versus marginal moral codes (Collins, 2025).

Tertiary socialization of AI and attaining an institutional position for meaning-construction: In the third stage, an artificial cognitive system is optimized or tuned to produce meaning within an institutional sub-world, such as a news organization, through learning institution-specific moral codes. Technically, during this process, the model's internal weights and parameters are re-adjusted based on the specialized data and textual content of that institutional sub-world. If the model is pre-trained, the professional ethical principles and organizational responsibility codes of that institutional domain—explicit knowledge—are incorporated into the computational framework. As a result, the model learns the institutional language and its specific applications within a distinct discursive sub-world. With the emergence of large language models, prompt engineering—alongside fine-tuning—has gained increasing attention as a key approach for the socialization of artificial intelligence models.

5. Discussion: Dual-spatialized representation within the circuit of culture

As noted, the relationship between natural and artificial intelligence is determined based on the similarities and differences in their performance in achieving social order, and conformity is proposed as the solution to the problem of social order. In the socio-organizational approach, conformity is defined as the reproduction of meanings within a discursive configuration through representational processes by all members of a culture who hold institutional positions in meaning production. With regard to the socialness of a natural cognitive system that occupies a position of institutional authority in meaning-making within a discourse, social order is defined as the representation of the world—or the act of giving meaning to the world—through language, in a manner aligned with the originators and speakers of that discourse who, within a particular historical moment and institutional location, possess the authority to construct meaning.

Thus far, the nature of socialized artificial intelligence has been explained, and its components are shown in the figure below.

Dual-spacization of Intelligence

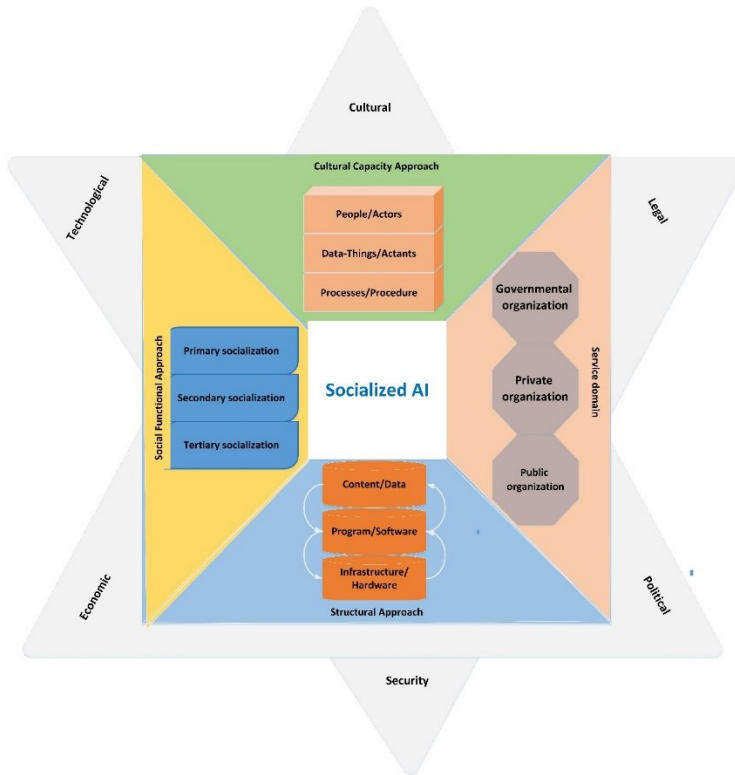


Figure 1. An integrated model of the socialization of AI

As previously stated, the representation of the world occurs through language. Given that language, according to representation theory, consists of practices and circulates meaning within the circuit of a culture through meaning-circulation systems—such as the media (Hall, 2003)—and since language serves a mediating role that, based on its philosophical foundations, is related to behavioral change (Vygotsky, 1978), it is thus a means by which the social actions of members of a culture can be coordinated around shared meanings oriented toward attaining common goals.

On the other hand, several scholars have demonstrated how governments and corporations attempt to reconstruct social reality through engineered processes of datafication, aligning it with their own interests and naturalizing these constructions (Couldry, 2021, pp. 11 & 13). Within the theoretical framework of this article, we also show that world-representation is not

confined to natural language within natural cognitive systems; rather, AI algorithms are new actants that become socialized in order to produce meaning in ways that resemble human meaning-making. The relationship between these two meaning-producing systems is articulated at the intersection of the two axes of social order shown in the table below.

Table 6. Relationships in the socialization of natural and artificial intelligence across the intersecting axes of social order

Axes of social order	Functional Relationships of the Socialization of Natural and Artificial Intelligence
Coordination of actions concerning the social distribution of knowledge through language	Gradual bio-cultural learning of natural language / task-oriented initial selection and configuration of formal-computational languages in artificial intelligence
	Gradual learning of moral codes/ Initial integration of professional ethics and corporate social responsibility into the computational codes
	Hard internalization of moral norms/ Soft integration of ethical codes through computational configurations
	Gradual learning of written and instrumental language/ Gradual adjustment of numerical codings of linguistic signs and neural network weights
	Local secondary socialization/ Translocal secondary socialization Learning specialized institutional language and skills/ Institutional fine-tuning based on re-adjusting numerical vectors and weights
Cooperation to attain common goals regarding algorithmic social division of labor	Early-life selection from the shared real world/ Prior, consensual selection from possible data worlds

This table confirms that we are dealing with a dual-spatialized social order, meaning that within a given institutional discourse and under specific historical conditions, the construction of meaning through AI algorithms resembles the construction of meaning through natural language by human actors who hold institutional authority in producing meaning.

Coordination of actions: Regarding language as a tool, since algorithms are developed and employed for "action and transformation" (Burgin, 2007: 2), they consist of practices and possess the power to shape behavior. They produce meaning and circulate it within the circuit of culture, thereby fixing

shared meanings about the proper and natural discursive use of things in the world – whether persons, objects, or events – across everyday life situations. Consequently, they exert real effects on the actions of members of a culture. Accordingly, within representation theory, algorithms have entered the circuit of culture and have acquired a position within the representational processes of this circuit. Since representational processes involve meaning-construction through both natural language and engineered algorithms, and since the sources of knowledge for both natural and artificial cognitive systems have become interwoven, representational processes within the circuit of culture have thus become dual-spatialized. On the other hand, the embedding of affective nuances in algorithms – such as emotions, commitments, concerns, and other culturally human considerations – depends on advances in software domains. Many scholars argue that current AI models still lack these human aspects, though substantial efforts are underway (Denning & Rousse, 2024; Janhonen, 2025). With language understood as a system of signs, the table above demonstrates that artificial cognitive systems, through numerical representation and the adjustment of weights in neural networks, are able to learn linguistic signs and their uses from data produced by members of society or institutional sub-worlds, thereby simulating human understanding. Nevertheless, the table also indicates that the socialization of AI does not depend exclusively on tacit knowledge. In defining the socialness of natural intelligence, we showed that socialness is fundamentally linked to tacit knowledge. However, artificial cognitive systems, which are developed and operate primarily within domains of engineered intentionality and lack lived social experience in the real world, require explicit knowledge in addition to tacit knowledge to perform in ways comparable to humans.

Accordingly, the socialness of an artificial cognitive system is defined as the system's capacity to acquire social fluency within an institutional discourse. Social fluency refers to the ability of an artificial cognitive system to learn linguistic signs and their uses from the data of societal members and from the data of originators and speakers within that institutional discourse, and to have its algorithms adjusted in accordance with the professional ethical codes and organizational social responsibility norms of that institutional sub-world.

Cooperation to attain common goals: Artificial cognitive systems constitute new speakers who have acquired institutional positions in meaning production within a sub-world. Since their design, development, and optimization occur collectively through the contributions of multiple actors, each possessing different specialized languages and technical skills, the agreements established among these actors can orient the performance of an artificial cognitive system in directions analogous to the performance of a natural cognitive system.

6. Conclusion: Socialized AI as a Provider of the Common Good and a Stable and Just Social Order

In this article, within the framework of the dual-spatial intelligence model and through a re-reading of representation theory in the socio-organizational approach, we examined the nature of socialized AI. Ultimately, we define this concept within the theoretical system of dual-spatial intelligence. Within this theoretical system: a) In the first slot, since AI algorithms, as tools, exert real effects on society, we focus primarily on the “real” within the real-virtual binary. b) In the second slot, because the practical domain in which AI becomes socialized is the virtual domain, the “virtual” side of the real-virtual binary takes center stage. c) In the third slot, since AI socialization occurs through the tacit knowledge of the society affected by AI and the explicit knowledge of institutions, one dimension of “socialness” –i.e., learning language as sign–is based on the performance similarity between natural and artificial intelligence. The other dimension –language as tool–is based on their performance difference. Accordingly, in the first case, we select “real” from the real-virtual binary, and in the second case, we select “virtual” dimension. Thus, socialized AI, in its sign-based linguistic dimension, is real-virtual-real, whereas in its tool-based linguistic dimension, it is real-virtual-virtual – half social and half regulatory.

On this basis, the socialness of AI is equivalent to the combination of: 1) social learning of language as both natural signs and second-order signs, and 2) the technical regulation of language as a tool, which is shaped by the division of labor (both), society’s tacit knowledge (first), and institutions’ explicit knowledge (second). Based on this, the application of current large language models at the level of institutions, without socialization in the third stage, would disrupt the social order of institutional discourse.

Within this framework, it appears that the most consequential decision in developing any AI model – whether deep learning-based or large-language-model-based – concerns its purpose. We must answer whether an artificial cognitive system should be socialized and regulated in a way that reproduces world knowledge and existing stereotypes in the traditional manner, thereby performing similarly to humans; or whether it should be socialized and regulated in a manner that delivers a form of rationality oriented toward the common good, and a sustainable and just social order. On this basis, a sustainable and just social order is grounded in algorithmic rationality, which enables the opening of the practice of dual-spatial representation. This process requires entering into social stereotypes themselves, subverting, opening up, and exposing them from within, which can influence the transformation of representational practices along the path of identity diversity. Indeed, AI is developed by the cultural and social ecosystems of societies. Certainly, many governments seek a civilizational, ethical, and perfectionist orientation in the quality of using emerging technologies, and many have succeeded in creating ethically regulated structures for the use of new technologies in general and AI in particular.

Furthermore, given the intertwining cognitive resources between natural and artificial intelligence and the continuous interaction between the two, human intelligence is evolving through its engagement with artificial intelligence, which possesses a high capacity for leveraging encoded variables. This trend has produced a new trajectory of hybrid human-machine capabilities. Accordingly, it seems that social institutions such as news media, in order to preserve institutional social order, should shift toward hybrid approaches that combine symbolic AI with data-driven AI.

Given the focus of this study on representational processes and the rules and regulations within the cultural circuit, other research avenues are suggested for the theoretical and practical development of this field. These include investigating the role of social artificial intelligence in other points of the cultural circuit, such as production, consumption, and identity. Furthermore, conducting an analysis aimed at assessing whether the meaning construction within meaning-making institutions is moving toward rationality or merely reproducing the status quo in the representation of social stereotypes can provide a realistic picture of the current state.

Bibliography

- Ade-Ibijola, A., & Okonkwo, C. (2023). Artificial intelligence in Africa: Emerging challenges. In D. O. Eke et al. (Eds.), *Responsible AI in Africa. Social and Cultural Studies of Robots and AI* (pp. 101-117). Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-08215-3_5
- Ameli, S. R. (2003a). Two globalizations and the future of the world. *Ketāb-e Māh-e Oloume Ejtemā'i*, (69-70), 15-28. [In Persian]
- Ameli, S. R. (2003b). Two globalizations and an international community of anxiety. *Journal of Social Science*, 21, 143-174. [In Persian]
- Ameli, S. R. (2011). *Globalization studies, dual spacitizations, & dual globalization*. Tehran, Iran: The R&D Research Center of SAMT. [In Persian]
- Ameli, S. R. (2017). *Philosophy of virtual space*. Amir Kabir Publishers. [In Persian]
- Ameli, S. R. (2021). Artificial intelligence: Its nature, philosophy, and significance. *Rāhavard-e Noor*, 20(77), 15-22. [In Persian]
- Ameli, S. R. (2023, February 27). Dual-spacization of communications, artificial intelligence, and the emergence of new computational diplomacy. Paper presented at the *Artificial Intelligence and New Diplomacies Conference*, Iranian World Studies Association and University of Tehran.
- Ameli, S. R. (2025). *Strategic studies of Islamic Republic of Iran's cyberspace*. Amir Kabir Publishers. [In Persian]
- Ameli, S. R., & Merali, A. (2004). *British Muslims' expectations of the government: Dual citizenship – British, Islamic or both; obligation, recognition, respect, and belonging*. Islamic Human Rights Commission.
- Ameli, S. R., Nazemi, S., & Sabbār, S. (2024). Dual-spacization and the representation of national identity on Instagram. *Journal of Iranian Cultural Research*, 17(4), 5-30. <https://doi.org/10.22035/jicr.2024.3367.3632> [In Persian]
- Anderson, C. W. (2013). *Rebuilding the news: Metropolitan journalism in the digital age*. Temple University Press.
- Babić, K., et al. (2020). Survey of neural text representation models. *Information*, 11(11), 511. <https://doi.org/10.3390/info11110511>
- Beckett, C., & Yaseen, M. (2023). *Generating change: A global survey of what news organisations are doing with AI*. The London School of Economics and Political Science.
- Belair-Gagnon, V. (2019). Sociology of news work. In T. P. Vos & F. Hanusch (Eds.), *The international encyclopedia of journalism studies*. John Wiley & Sons, Inc.
- Berger, P. L., & Luckmann, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge*. Anchor Books.
- Bhaskar, R. (1998). General introduction. In M. S. Archer et al. (Eds.), *Critical Realism: Essential Readings* (pp. ix-xxiv). Routledge.

- Boczkowski, P. (2005). *Digitizing the news: Innovation in online newspapers*. The MIT Press.
- Bourdieu, P. (1998). *On television and journalism*. Pluto Press.
- Breed, W. (1955). Social control in the newsroom: A functional analysis. *Social Forces*, 33(4), 326–335. <https://doi.org/10.2307/2573002>
- Burgin, M. (2007). Languages, algorithms, procedures, calculi, and metalogic. *arXiv*. <https://arxiv.org/abs/math/0701121>
- Bygstad, B., & Munkvold, B. E. (2011). In search of mechanisms: Conducting a critical realist data analysis. In *Proceedings of the Thirty-Second International Conference on Information Systems*, Shanghai.
- Bytiak, Y. P., Danilyan, O. G., Dzeban, A. P., Kalinovsky, Y. Y., & Chalapko, V. V. (2020). Information society: the interaction of tradition and innovation in communicative processes. *Amazonia Investiga*, 9(27), 217–226. <https://doi.org/10.34069/AI/2020.27.03.23>
- Collins, H. (1998). Socialness and the undersocialised conception of society. *Science, Technology, & Human Values*, 23(4), 494–516. <https://doi.org/10.1177/0162243998023004>
- Collins, H. (2025). Why artificial intelligence needs sociology of knowledge: Parts I and II. *AI & Society*, 40, 1249–1263. <https://doi.org/10.1007/s00146-024-01954-8>
- Couldry, N. (2021). The social construction of reality – really! In D. Herbert & S. Fisher-Høyrem (Eds.), *Social Media and Social Order* (pp. 10–16). De Gruyter.
- Cox, J. B. (2022). Black Lives Matter to media (finally): A content analysis of news coverage during summer 2020. *Newspaper Research Journal*, 43(2), 155–175. <https://doi.org/10.1177/07395329221092719>
- Danks, D. (2014). Learning. In K. Frankish & W. M. Ramsey (Eds.), *The Cambridge handbook of artificial intelligence* (pp. 151–167). Cambridge University Press.
- Denning, P., & Rouse, B. (2024). Can machines be in language? *Communications of the ACM*, 67(3), 32–35. <https://doi.org/10.1145/3637629>
- Dodds, T., et al. (2025). On controlled change: Generative AI's impact on professional authority in journalism. *arXiv*. <https://arxiv.org/abs/2510.19792>
- Ferrucci, P., et al. (2022). *The institutions changing journalism: Barbarians inside the gate*. Routledge.
- Fin, E. (2017). *What algorithms want: Imagination in the age of computing*. The MIT Press.
- Gans, H. (1979). *Deciding what's news*. Pantheon Books.
- Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie et al. (Eds.), *Media technologies: Essays on Communication, Materiality, and Society* (pp. 167–193). MIT Press.
- Goffey, A. (2008). Algorithm. In M. Fuller (Ed.), *Software studies: A lexicon* (pp. 15–20). MIT Press.

- Gravngaard, G., & Rimestad, L. (2014). Socializing journalist trainees in the newsroom: On how to capture the intangible parts of the process. *Nordicom Review*, 35(Special Issue), 81–95. <http://hdl.handle.net/2077/37339>
- Griffin, E. M. (2012). *A first look at communication theory* (8th ed.). McGraw Hill.
- Groves, J., & Brown, C. (2015). *Transforming newsrooms: Connecting organizational culture, strategy, and innovation*. Routledge.
- Hall, S. (1992). The West and the rest: Discourse and power. In *Essential Essays* (Vol. 2, pp. 185–224). Duke University Press.
- Hall, S. (1997). *Representation & the media*. Media Education Foundation.
- Hall, S. (2003). The work of representation. In S. Hall (Ed.), *Representation: Cultural representations and signifying practices* (pp. 13–74). SAGE.
- Hallin, D. C., & Mancini, P. (2012). *Comparing media systems: Three models of media and politics*. Cambridge University Press.
- Halpin, H. (2025). Artificial intelligence versus collective intelligence. *AI & Society*, 40, 4589–4604. <https://doi.org/10.1007/s00146-025-02240-x>
- Hechter, M., & Horne, C. (2009). *Theories of social order* (2nd ed.). Stanford University Press.
- Herman, E., & Chomsky, N. (1988). *A propaganda model*. <https://chomsky.info/consent01/>
- Hohenstein, J., Kizilcec, R. F., DiFranzo, D., Aghajari, Z., Mieczkowski, H., Levy, K., ... & Jung, M. F. (2023). Artificial intelligence in communication impacts language and social relationships. *Scientific Reports*, 13(1), 5487. <https://doi.org/10.1038/s41598-023-30938-9>
- 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.
- Hughes, H. M. (1940). *News and the human interest story*. The University of Chicago Press.
- Huh, M., Cheung, B., Wang, T., & Isola, P. (2024). The Platonic Representation Hypothesis. In *Proceedings of the 41st International Conference on Machine Learning* (Vol. 235, pp. 20617–20642). PMLR. <https://doi.org/10.48550/arXiv.2405.07987>
- IBM. (n.d.). *What is AI agent perception*. <https://www.ibm.com/think/topics/ai-agent-perception>
- Janhonen, J. (2025). Socialisation approach to AI value acquisition: Enabling flexible ethical navigation with built-in receptiveness to social influence. *AI Ethics*, 5, 527–553. <https://doi.org/10.1007/s43681-023-00372-8>
- Keikha, A. A., & Rahgozar, M. (2018). Link prediction on social networks based on deep learning. In *Proceedings of the 4th International Conference on Web Research (ICWR)*.
- Kubin, E., & Von Sikorski, C. (2021). The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association*, 45(3), 188–206. <https://doi.org/10.1080/23808985.2021.1976070>

- Lewin, K. (1947). Frontiers in group dynamics: Concept, method and reality in social science; social equilibria and social change. *Human Relations*, 1(1), 5-41. <https://doi.org/10.1177/001872674700100103>
- Lewis, S. C., & Westlund, O. (2015). Actors, actants, audiences, and activities in cross-media news work: A matrix and a research agenda. *Digital Journalism*, 3(1), 19-37. <https://doi.org/10.1080/21670811.2014.927986>
- Mackenroth, K. S. (2025). *Ethics, morals and the professional*. Professional Landmen's Association of New Orleans.
- Magee, B. (2019). *The story of philosophy* (H. Kamshad, Trans.). Tehran: Nashreney. [In Persian]
- Mohseni Ahooei, E. (2024). The datafied society: Challenges and strategies in big data research for social sciences and humanities. *Journal of Cyberspace Studies*, 8(2), 177-207. <https://doi.org/10.22059/jcss.2024.378294.1106>
- Nothias, T. (2020). Representation and journalism. In J. F. Nussbaum & M. Powers (Eds.), *Oxford research encyclopedia of communication*. Oxford University Press. <https://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-868>
- Nourbakhsh, Y., & nemati, S. F. (2020). Educational Impacts and Outcomes of Cyberspace in the Realization of New Islamic Civilization. *Scientific Journal Of New Islamic Civilization Fundamental Studies*, 3(2), (Serial 6), 12-376. <https://doi.org/10.22070/nic.2021.14018.1079>
- Oliveira, N., Li, J., Khalvati, K., Barragan, R. C., Reinecke, K., Meltzoff, A. N., ... Rao, R. P. N. (2025). Culturally-attuned AI: Implicit learning of altruistic cultural values through inverse reinforcement learning. *PLOS ONE*, 20(12), Article e0337914. <https://doi.org/10.1371/journal.pone.0337914>
- Revers, M. (2017). *Contemporary journalism in the US and Germany*. Palgrave Macmillan US.
- Rohlinger, D. A. (2022). *Society and new media* (M. Pourrajabi & H. R. Bijani, Trans.). Tehran: Naqd-i Farhang. [In Persian]
- Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.
- Russell, S. J., & Norvig, P. (2022). *Artificial intelligence: A modern approach* (4th ed., Global edition). Pearson.
- Rydgren, J. (2008). Conformity. In W. A. Darity (Ed.), *International encyclopedia of the social sciences* (pp. 71-73). Macmillan.
- Sabbar, S. and Matheson, D. (2019). Mass Media vs. the Mass of Media: A Study on the Human Nodes in a Social Network and their Chosen Messages. *Journal of Cyberspace Studies*, 3(1), 23-42. <https://doi.org/10.22059/jcss.2019.271467.1031>
- Salehi, K., Habib Zadeh Khiyaban, S., Sabbar S. (2026). Artificial Intelligence and Crime Detection: A Critical Review. *Cyberspace Studies*. 10(1), 181-197. <https://doi.org/10.22059/jcss.2025.402206.1179>

- Schatzki, T. R. (2025). The Changing Forms of Social Phenomena Today. *Philosophy & Technology*, 38(1), 19. <https://doi.org/10.1007/s13347-025-00853-0>
- Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale University Press.
- Shahghasemi, E. (2025). 'Woke' in translation: Persian perspectives on Platform X. *Discover Global Society*, 3(104). <https://doi.org/10.1007/s44282-025-00253-x>
- Shahghasemi, E. (2025). AI: A human future. *Journal of Cyberspace Studies*, 9(1), 145-173. <https://doi.org/10.22059/jcss.2025.389027.1123>
- Shahghasemi, E., Gholami, F., & Alikhani, Z. (2025). Global patterns of social media use and political sentiment. *Discover Global Society*, 3(1), 36. <https://doi.org/10.1007/s44282-025-00171-y>
- Shokrkhah, Y. (2018). *Cyber journalism: Newsroom versus traditional journalism* [In Persian]. Tehran: Sanieh Press.
- Silver, D., & Clark, T. N. (2008). Social theory. In W. A. Darity Jr. (Ed.), *International Encyclopedia of the Social Sciences* (2nd ed.). Course Technology Cengage Learning.
- Singer, J. B. (2004). More than ink-stained wretches: The resocialization of print journalists in converged newsrooms. *Journalism & Mass Communication Quarterly*, 81(4), 838-856. <https://doi.org/10.1177/107769900408100408>
- Sissons, H., & Smith, P. (2026). *The "socially networked" newsroom: Journalists and their discourses of digital communication*. Bloomsbury Academic.
- Sokol, K., & Flach, P. (2024). Interpretable representations in explainable AI: From theory to practice. *Data Mining and Knowledge Discovery*, 38, 3102-3140. <https://doi.org/10.1007/s10618-024-01010-5>
- Sutton, R. S., & Barto, A. G. (2014). *Reinforcement learning: An introduction* (1st ed.). MIT Press.
- Sylvie, G. (2018). *Reshaping the news, community, engagement, and editors*. Peter Lang.
- Talebi Tadi, M., & Rastegar Khaled, A. (2025). A comparative study of the theory of social understanding of the text and discourse analysis theories. *Journal of Interdisciplinary Studies in the Humanities*, 17(3), 101-125. <https://doi.org/10.22035/isih.2025.5261.5002>
- Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.
- Tortoise. (2024). *The global AI index*. <https://www.tortoisemedia.com/data/global-ai#data>
- Tuchman, G. (1978). *Making news: A study in the construction of reality*. The University of Michigan Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Weber, M. (1976). Towards a sociology of the press. *Journal of Communication*, 26(3), 96-101. <https://doi.org/10.1111/j.1460-2466.1976.tb01910.x>

Dual-spacization of Intelligence

- White, D. M. (1950). The “gate keeper”: A case study in the selection of news. *Journalism Quarterly*, 27(4), 383–390. <https://doi.org/10.1177/107769905002700403>
- Yadav, A. (2024). How does embedding layer work? *Medium*. <https://medium.com/biased-algorithms/how-does-embedding-layer-work-eafb1c721302>
- Yıldız, D., & Nur, Z. (2024). Transformation of social interaction in the digital age: Impact, challenges, and prospects of technology in social relationships. *Bulletin of Science, Technology and Society*, 3(3), 49-54. <https://inti.ejournalmeta.com/index.php/inti/article/view/95>



Original-Forschungsarbeit

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

Mohsen Karami¹

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

🌐 <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

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

محسن کرمی^۱

^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 

<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 Karami¹

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

🌐 <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

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

- 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.



Original-Forschungsarbeit

KI als Grenzobjekt: der persische X-Diskurs

Shaho Sabbar¹

1 Assistenzprofessor, Abteilung für Iranistik, Fakultät für Weltstudien, Universität Teheran, Teheran, Iran

Empfangen: 1. März 2025 Akzeptiert: 7. Juni 2025

Zusammenfassung:

Diese Studie untersucht, wie persischsprachige Nutzer auf der Social-Media-Plattform X generative künstliche Intelligenz als sozio-technisches und diskursives Phänomen verhandeln. Auf der Grundlage eines Datensatzes von 24.215 persischsprachigen Beiträgen verwenden wir ein Multi-Label-Topic-Modeling-Verfahren sowie affektives Profiling, um den öffentlichen Diskurs über KI-Werkzeuge, ihre wahrgenommenen Implikationen und normative Bewertungen ihrer Nutzung zu analysieren. Anstatt Stimmung als statischen Indikator von Meinungen zu betrachten, interpretieren wir affektiven Ausdruck als kommunikativen Akt, der durch plattformspezifische Anreize und kulturelle Kontexte geprägt ist. Unsere Ergebnisse zeigen, dass KI nicht nur als technisches Artefakt positioniert wird, sondern auch als Grenzobjekt, das mit Debatten über Expertise, Ethik und institutionelle Legitimität verflochten ist. Der Diskurs ist in praktischen Anliegen verankert – insbesondere in Bezug auf Arbeit, Bildung und Vergleiche zwischen KI-Werkzeugen –, erweitert sich jedoch häufig zu kulturspezifischen Narrativen über Risiko, Fairness und epistemische Autorität. Emotional ist die Diskussion durch pragmatischen Optimismus, kritische Intensität und ein beträchtliches neutrales Spektrum gekennzeichnet, das eher Orientierung als Bewertung widerspiegelt. Diese Studie trägt zu aktuellen Debatten in der Kommunikationswissenschaft, der KI-Ethik und den Plattformstudien bei, indem sie eine nicht anglophone, kulturell verankerte Analyse dafür liefert, wie Öffentlichkeiten eine alltagsprachliche Governance über aufkommende Technologien praktizieren.

Schlüsselwörter: künstliche Intelligenz, Sentimentanalyse, digitale Öffentlichkeiten, alltagsprachliche Governance, persischer Diskurs

* Korrespondierender Autor

✉ shaho.sabbar@ut.ac.ir

🌐 <https://orcid.org/0000-0001-5801-7137>

Wie dieser Artikel zu zitieren ist:

Sabbar, S. (2025). AI as a boundary object: The Persian X discourse. *Spektrum Iran*, 38(2), 61-82.

🔗 <https://doi.org/10.22034/spektrum.2026.569202.1059>

© 0 0 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

هوش مصنوعی به مثابه ابژه مرزی: گفتمان فارسی در ایکس

شاهو صبار^۱

۱ استادیار، گروه مطالعات ایران، دانشکده مطالعات جهان، دانشگاه تهران، تهران، ایران

دریافت: ۱۴۰۳/۱۲/۱۱ پذیرش: ۱۴۰۴/۰۳/۱۷

چکیده:

این مطالعه بررسی می‌کند که کاربران فارسی‌زبان در پلتفرم رسانه اجتماعی ایکس چگونه با هوش مصنوعی مولد به عنوان پدیده‌ای اجتماعی-فنی و گفتمانی تعامل می‌کنند. ما با بهره گرفتن از مجموعه داده‌ای شامل ۲۴۲۱۵ پست فارسی، از چارچوب مدل‌سازی موضوعی چندبرچسبی و نیز پروفایل‌سازی عاطفی برای تحلیل گفتمان عمومی درباره ابزارهای هوش مصنوعی، پیامدهای ادراک شده آن‌ها و داوری‌های هنجاری درباره کاربردشان استفاده کرده‌ایم. به جای آنکه احساسات را به عنوان شاخصی ایستا از نگرش‌ها در نظر بگیریم، بیان عاطفی را به مثابه کنشی ارتباطی تفسیر می‌کنیم که تحت تأثیر مشوق‌های پلتفرمی و بافت فرهنگی شکل می‌گیرد. یافته‌های ما نشان می‌دهد که هوش مصنوعی نه تنها به عنوان یک مصنوع فنی، بلکه به عنوان ابژه‌ای مرزی بازنمایی می‌شود که با مباحث مربوط به تخصص، اخلاق و مشروعیت نهادی درهم تنیده است. این گفتمان در نگرانی‌های عملی — به‌ویژه در حوزه کار، آموزش و مقایسه میان ابزارهای هوش مصنوعی — ریشه دارد، اما اغلب به روایت‌های فرهنگی خاص درباره ریسک، انصاف و مرجعیت معرفتی بسط می‌یابد. از نظر عاطفی، این گفت‌وگو با نوعی مثبت‌اندیشی عمل‌گرایانه، شدت انتقادی و طیف قابل توجهی از موضع خنثی مشخص می‌شود که بیشتر نشان‌دهنده جهت‌گیری است تا ارزیابی. این مطالعه با ارائه تحلیلی فرهنگی و غیرانگلیسی از نحوه اعمال حکمرانی عامیانه از سوی عموم بر فناوری‌های نوپدید، به مباحث جاری در حوزه ارتباطات، اخلاق هوش مصنوعی و مطالعات پلتفرم‌ها کمک می‌کند.

واژگان کلیدی: هوش مصنوعی، تحلیل احساسات، عمومیت‌های دیجیتال، حکمرانی عامیانه، گفتمان فارسی



Original Research Paper

AI as a boundary object: The Persian X discourse

Shaho Sabbar^{1*}

¹ Assistant Professor, Department of Iranian Studies, Faculty of World Studies, University of Tehran, Tehran, Iran

Received: Mar., 1, 2025 Accepted: Jun. 7, 2025

Abstract

This study investigates how Persian-speaking users on the social media platform X engage with generative artificial intelligence as a sociotechnical and discursive phenomenon. Drawing on a dataset of 24,215 Persian-language posts, we employ a multi-label topic modeling framework and affective profiling to analyze public discourse surrounding AI tools, their perceived implications, and normative judgments about their use. Rather than treating sentiment as a static indicator of opinion, we interpret affective expression as a communicative act shaped by platform incentives and cultural context. Our findings show that AI is positioned not only as a technical artifact but as a boundary object entangled with debates over expertise, ethics, and institutional legitimacy. The discourse is anchored in practical concerns—especially labor, education, and comparisons among AI tools—but frequently extends into culturally specific narratives about risk, fairness, and epistemic authority. Emotionally, the conversation is marked by pragmatic positivity, critical intensity, and a sizable neutral band reflecting orientation rather than evaluation. This study contributes to ongoing debates in communication, AI ethics, and platform studies by offering a non-Anglophone, culturally grounded analysis of how publics perform vernacular governance over emerging technologies.

Keywords: artificial intelligence, sentiment analysis, digital publics, vernacular governance, Persian discourse

* Corresponding Author

✉ shaho.sabbar@ut.ac.ir

🌐 <https://orcid.org/0000-0001-5801-7137>

How to Cite this Article:

Sabbar, S. (2025). AI as a boundary object: The Persian X discourse. *Spektrum Iran*, 38(2), 61-82.

📄 <https://doi.org/10.22034/spektrum.2026.569202.1059>

© 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

1. Introduction

Throughout history, the emergence of new communication technologies has reconfigured the conditions of public life. From the printing press to the smartphone, such technologies have not only extended communicative capacity but also shaped how individuals perceive themselves, their social roles, and their relationship to knowledge and power. These perceptions are neither uniform nor apolitical; rather, they are deeply entangled with affect, authority, and cultural imaginaries. As Harold Innis (1951) and Marshall McLuhan (1964) argued, media technologies do not merely transmit content—they exert structural influence, favoring particular modes of expression and societal organization. Today, the rise of generative artificial intelligence represents a comparable epistemic shift, reframing intimacy, cognition, and expertise within algorithmically mediated systems of meaning-making (Elliott, 2023; Seaver, 2022).

Public reactions to AI are shaped not only by functional evaluations but also by emotional and moral concerns. This affective response is especially visible in the context of "algorithmic intimacy" — the perceived closeness and social presence that some users experience when interacting with AI-driven agents (Bozdağ, 2025; Reineke, 2022). While earlier technologies inspired utopian or dystopian imaginaries about mass democratization or social alienation, AI technologies invite a subtler blend of fascination, ambivalence, and critique than earlier waves of media innovation (Salehi et al., 2026). Studies of AI-mediated relationships show how users anthropomorphize chatbots, frame them as companions, or even as emotional partners (Balazadeh & Kajonius, 2025; George et al., 2023; Sjoraida, 2025). These engagements reveal that users' understanding of AI is shaped not merely by its accuracy or productivity, but by its capacity to signal empathy, authority, and social intention (Chu et al., 2025; Turkle, 2011).

How people imagine and emotionally respond to AI will, in turn, shape the trajectory of the technology itself (Sabbar & Habib Zadeh Khiyaban, 2023). As Papacharissi (2015) has argued, the circulation of sentiment in digital spaces helps constitute "affective publics" — collectives organized not only around shared beliefs but around shared intensities. In this light, public perception is not simply a reflection of technological progress; it is a force that can accelerate, redirect, or constrain it. Expectations about AI's ethical

boundaries, usefulness, and risks inform not only personal adoption but institutional policies and commercial design choices. Intimate forms of interaction—such as venting to a chatbot or querying it about political beliefs—can crystallize

into collective norms about what AI should and should not do (Illouz, 2007; Lupton, 2022; Shahghasemi, 2025). These expectations travel through discourse, anchoring future imaginaries in contemporary discourse.

Social media platforms play a central role in mediating this discourse. They operate not only as amplifiers of information but as infrastructures of affective and epistemic coordination. The platform X, in particular, has become a key site for the articulation of public sentiment about AI. With its rapid tempo, algorithmic virality, and entanglement of personal and professional identities, X facilitates the emergence of vernacular expertise—informal yet consequential claims about how AI works, what it means, and what it should become. On X, users post not only evaluations of AI systems but narratives, warnings, jokes, and moral commentaries that collectively shape public understanding. This is especially true in non-Anglophone contexts, where localized meanings of AI emerge through culturally embedded discourse practices and political framings.

In the Persian-language digital sphere, the convergence of algorithmic tools and platform discourse has produced a particularly rich and understudied communicative field. As recent studies suggest, Persian-speaking publics increasingly use platforms like X to articulate vernacular forms of technological governance—public deliberations, critiques, and norm-setting that occur outside formal regulatory channels. These discourses are shaped by hybrid media ecologies, where access to global tools is uneven, state censorship remains active, and users often navigate both technical and moral complexities in evaluating digital systems. Generative AI, in this context, is not merely an innovation but a cultural problem: a figure of fascination, suspicion, and speculative projection.

Despite the richness of this public conversation, existing scholarship has largely neglected the ways in which non-Western publics discuss and contest the meanings of AI. Much of the literature on AI discourse remains Anglocentric, focusing primarily on elite media narratives or institutional policy debates. In contrast, everyday public talk—especially in languages

other than English—remains under-theorized, even though such talk plays a formative role in shaping user norms, perceptions of legitimacy, and trust in AI systems (Papacharissi, 2015; Shahghasemi et al., 2025). There is a pressing need to understand how publics outside the Global North interpret AI, not as an abstract technological domain, but as a lived, affectively charged, and socially negotiated object.

This study addresses that gap by analyzing a large-scale corpus of Persian-language posts from X, focusing on how users talk about generative AI across technical, emotional, and normative registers. Building on theories of affective publics, algorithmic intimacy, and vernacular governance, we examine AI discourse not only for its content but for its communicative form: how users perform expertise, express sentiment, and negotiate legitimacy. We treat affective expressions not as raw indicators of approval or disapproval but as socially situated acts—moves within a field of public reasoning, identity performance, and epistemic claim-making (van Dijk, 1993; Bourdieu, 1991).

Our approach combines multi-label topic modeling with affective profiling, allowing us to map both the thematic structure and emotional tone of AI discourse among Persian-speaking users. Rather than imposing normative assumptions about technological progress or ethical risk, we seek to surface how publics themselves articulate what matters—whether in concerns about labor displacement, academic integrity, misinformation, or emotional authenticity. These findings contribute to ongoing conversations in AI ethics, communication studies, and cultural sociology by grounding analysis in the actual speech practices of users navigating new technological realities.

Specifically, this study is guided by the following research questions:

1. What thematic concerns dominate Persian-language discourse about generative AI on the X platform?
2. How do users express affect in their evaluations of AI tools, and what emotional patterns characterize the discourse?
3. In what ways do users perform vernacular governance—norm-setting, critique, and pedagogy—through their interactions with AI discourse?

By addressing these questions, we aim to illuminate how publics not only react to AI but participate in shaping its social meaning. In doing so, we position generative AI as a boundary object—flexible enough to accommodate multiple interpretations but structured enough to anchor contestation (Star & Griesemer, 1989). The discourse around AI is not merely a reflection of public opinion; it is a site of cultural labor where values, expectations, and technological futures are being actively negotiated.

2. Methodology

To examine how Persian-speaking users discuss and evaluate artificial intelligence across every day, institutional, and speculative registers, we compiled a large corpus of public posts from X (formerly Twitter). Data collection followed a high recall retrieval logic: the goal was to minimize thematic blind spots by casting a wide net around AI-related talk rather than narrowly sampling only explicit technical discussions. We therefore combined multiple families of Persian and English keywords and their common transliterations, including general AI terms (e.g., هوش مصنوعی, ai), model and platform names (e.g., ChatGPT, Grok, Gemini, Copilot), interaction and use terms (e.g., پرامپت/Prompt, prompt engineering), and risk or governance cues (e.g., جعل, حریم خصوصی, کپی رایت, مقررات, دیپ فیک/deepfake). Because AI discourse on X frequently involves code switching and brand shorthand, retrieval terms were expanded iteratively during pilot checks to capture vernacular variants, spacing differences, and orthographic alternations. This process produced an initial pool of 26,111 posts. Preprocessing was designed to preserve discursive and affective signals while removing mechanical noise and reducing distortions produced by platform repetition. First, duplicate and near duplicate posts were removed, along with items that were not analytically usable for text-based modeling (for example, posts with no substantive text). Second, we filtered out posts that were not meaningfully about AI despite containing ambiguous trigger terms, using a combination of rule-based exclusions and manual spot checks to prevent systematic drift into irrelevant domains. Third, we normalized Persian orthography to improve the stability of downstream matching and classification. This included harmonizing Arabic/Persian character variants (e.g., ی/ی and ک/ك), handling zero-width non-joiners, and stripping diacritics.

Links and user mentions were removed to reduce sparsity and to avoid overweighting platform-specific identifiers. Throughout cleaning, we prioritized retaining the expressive content of posts, including colloquial phrasing and common mixed language forms, because these features carry much of the interactional meaning on X.

After deduplication, relevance filtering, and normalization, the final analytic dataset contained 24,215 posts. All records were drawn from public content and are reported only in aggregate. In reporting examples, we avoid including user identifiers or content that could facilitate re-identification. This curated corpus provides the empirical foundation for subsequent steps in the analysis, including multi-label topic tagging across ten thematic formations and the mapping of co-occurrence structures and affective profiles.

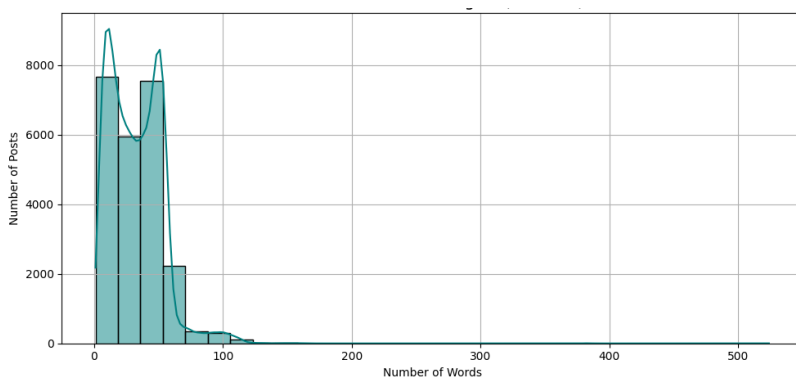


Figure 1. Distribution of text lengths (in words)

3. Findings

Applying the five-category affect schema to the full analytic corpus (n = 24,215) produces a distribution in which evaluative warmth is present but far from uniform. The largest share of posts falls into Happy (8,271; 34.16%), followed closely by Neutral (7,838; 32.37%). More emphatic positivity, coded as Delighted, accounts for 2,400 posts (9.91%). On the negative side, Angry comprises 2,848 posts (11.76%), and Furious 2,858 (11.80%). Taken together, Angry + Furious represent 5,706 posts (23.56%), while Happy + Delighted

represent 10,671 posts (44.07%). This pattern suggests a public conversation that is neither celebratory nor uniformly alarmist. The prominence of Happy relative to Delighted points to a largely pragmatic register: users often signal approval, usefulness, or everyday satisfaction without escalating into unequivocal enthusiasm. At the same time, the sizeable Neutral band indicates that much AI talk on X functions as circulation and orientation rather than overt evaluation: sharing updates, noting capabilities or failures, relaying comparisons, and positioning the self as an observer in a fast-moving informational environment. Negative affect is nonetheless substantial. The combined quarter of posts coded as Angry or Furious signals that AI is also a recurring site of friction and moral contestation, where users mobilize indignation and outrage as communicative resources, not simply as spontaneous reactions. In the analyses that follow, this distribution motivates treating AI discourse as an effectively mixed field: routine approval and curiosity coexist with sustained critique, and platform dynamics help both travel and settle into recognizable, repeatable stances.

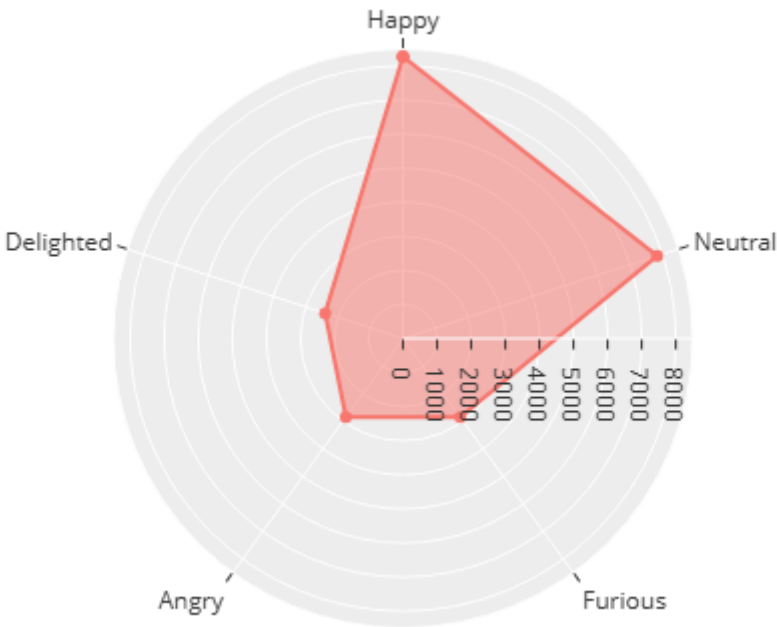


Figure 2. Dataset sentiment distribution

3.1. Topic Modeling and Thematic Analysis

To map how AI is talked about in Persian-language publics on X, we used an unsupervised topic modeling step as a diagnostic lens rather than as the final measurement instrument. After standard text preparation for Persian and code-switched writing (orthographic harmonization, removal of URLs and user mentions, and normalization of common spacing and character variants), we estimated a series of LDA models while varying the number of topics from 5 to 20. We then compared solutions using a semantic coherence criterion to identify a topic resolution that balances empirical fit with interpretability for communication analysis. Coherence values rose unevenly across the sweep and reached their maximum at $K = 10$ (coherence = 0.5241). Smaller local improvements appeared at higher resolutions (e.g., $K = 14$: 0.4903; $K = 19$: 0.4793), but these did not surpass the ten-topic solution. Substantively, the ten-topic model was preferred because it produced categories that correspond to recognizable communicative formations in the corpus while avoiding over-fragmentation. Lower K solutions tended to fuse distinct arenas of AI talk (for example, merging tool comparison chatter with work-related anxiety, or folding governance debates into misinformation talk), which made it harder to track how different kinds of claims and emotions travel. Higher K specifications, by contrast, often split coherent clusters into narrow subthemes that were difficult to name cleanly and were less useful for doctoral-level theorization about platformed publics, affect, and epistemic authority. We therefore adopt the ten-topic solution as the primary thematic frame for the analyses that follow.

Because posts on X frequently span multiple domains in the same utterance (for example, a model comparison embedded in a story about workplace use, or prompting advice paired with ethical boundary testing), we operationalized the thematic frame as multi-label topic tagging rather than forcing each post into a single bucket. The final corpus contains 24,215 posts, and 74.78% of posts are linked to at least one of the ten themes (mean topics per post = 1.21). Overlap is common: 41.39% of posts activate exactly one topic label, 23.27% activate two, and 8.16% activate three; smaller shares activate four or more (1.61% for four, 0.31% for five, 0.04% for six). A minority (25.22%) remains untagged under this topical frame, reflecting either highly generic AI references or content that does not meaningfully align with the ten

major formations. The ten themes and their prevalence in the corpus (non-exclusive) are as follows: Tools and model comparisons (6,324; 26.12%), Work, economy, and business impacts (5,954; 24.59%), Education and university/school contexts (5,648; 23.32%), Future narratives and predictions (3,131; 12.93%), Ethics, privacy, regulation, and governance (2,284; 9.43%), Content creation and creativity (2,135; 8.82%), Deepfakes, misinformation, and verification (1,504; 6.21%), Programming and software development (1,434; 5.92%), Sensitive social and mental health discussions (480; 1.98%), and Prompting practices and strategies (316; 1.30%). Framed this way, the thematic backbone of Persian-language AI discourse is anchored in practical, comparative talk about tools and their everyday implications, while still reserving distinct space for governance concerns, authenticity crises, and the occasional movement into sensitive domains.

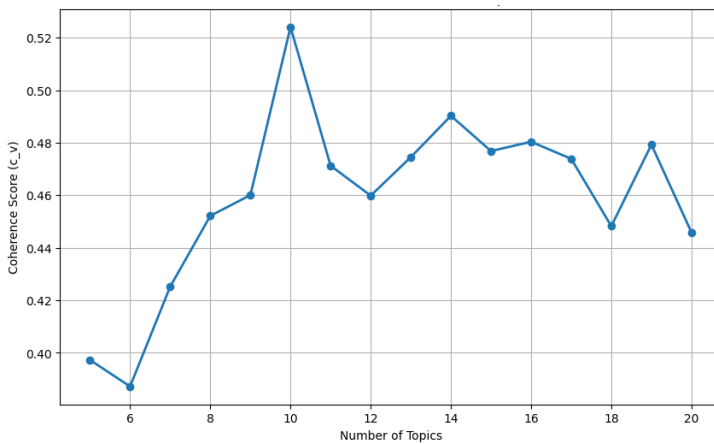


Figure 3. Coherence score vs number of topics

3.2. AI Tools Ecosystem and Model Comparisons

Within the corpus, tool-centered discussion constitutes a large and highly performative stream ($n = 6,324$ posts). Affective tone is split between Neutral (2,190; 34.63%) and Happy (2,140; 33.84%), with a substantial critical layer (Angry 924; 14.61%, Furious 592; 9.36%) and comparatively fewer fully celebratory posts (Delighted 478; 7.56%). Reading the Texts suggests that “comparison” is rarely a calm, laboratory-style evaluation; instead, users

narrate tools through everyday trials, viral anecdotes, and rapid verdicts that travel well on X. Posts juxtapose models by name, circulate claims about new versions, and treat outputs as evidence in microdebates about competence, bias, and reliability. The same thread can move from an impromptu benchmark to a joke about social etiquette with machines, as in the recurring anthropomorphic register (e.g., “My mom apologizes to ChatGPT for taking its time”). Negative affect often clusters around perceived regressions, hallucinations, access and cost frustrations, or the sense that hype is outpacing trustworthy verification, captured in short evaluative utterances such as “Grok is a mess today.” Communication-wise, these posts do reputational work: they elevate some tools as “serious,” demote others as “broken,” and establish vernacular criteria for credibility (speed, accuracy, language handling, refusal behavior) that circulate through quote posting and quick comparisons rather than formal reviews.

3.3. Prompting Practices and Strategies

Although smaller in volume (n = 316), prompting talk is disproportionately instruction-oriented and tends to carry an upbeat tone (Happy 117; 37.03%, Delighted 58; 18.35%) alongside a sizable Neutral band (103; 32.59%) and relatively limited negativity (Angry 20; 6.33%, Furious 18; 5.70%). Qualitatively, the Texts read like a folk pedagogy of interaction: users share templates, warn against vague requests, and present prompting as a communicative skill that turns “a chatbot” into a cooperative worker. Many posts emphasize specifying role, constraints, and format, often in the voice of practical coaching (e.g., “Give it a role, your goal, and the output format, then iterate”). At the same time, prompting discourse becomes a site where norms of legitimacy and accountability are negotiated, especially when prompts are used for coding or workplace outputs; users not only exchange techniques but they also propose boundaries (review, disclosure, and responsibility for errors) that reassert human agency over delegated writing. In communication terms, this topic is metacommunication: it is about how to talk to AI, and it produces a lay vocabulary of clarity, control, and alignment that helps participants position themselves as competent users rather than passive consumers.

3.4. Future Narratives, Scenarios, and Predictions

Future-facing posts (n = 3,131) display a mixed emotional profile, dominated by Neutral (1,098; 35.07%) and Happy (1,062; 33.92%) but with a notable share of high intensity reactions (Furious 334; 10.67%) and a meaningful layer of excitement (Delighted 490; 15.65%), while Angry is comparatively lower (147; 4.69%). In the texts, “the future” functions as a rhetorical device for making present stakes legible: users forecast labor market disruption, reimagine professional identities (especially for developers and knowledge workers), and debate whether the key resource will be skill, access, or infrastructure. Some posts frame AI as a coming competence gap and counsel adaptation through fundamentals and steering capacity (e.g., “Those who master the basics and can direct models will be ahead”), whereas others articulate a more conflictual imaginary in which geopolitical competition, compute, and platform power set the terms of possibility. The future topic also carries an affective oscillation typical of platformed publics: optimism travels via narratives of productivity and opportunity, while outrage concentrates around fears of deception, deskilling, or institutional unpreparedness. Analytically, these posts are less about forecasting accuracy than about social coordination: they align audiences around what to worry about, what to learn, and who should be held responsible as AI is narrated into collective horizons.

3.5. Education, Universities, and Schools

Education-oriented AI talk forms one of the densest thematic streams in the corpus (n = 5,648). Its affective profile is led by Neutral (1,968; 34.84%) and Happy (1,719; 30.44%), with Delighted present but secondary (653; 11.56%) and a substantial negative layer (Angry 579; 10.25%, Furious 729; 12.91%). Reading the texts, education is discussed less as abstract policy and more as a lived site of friction where institutional routines collide with generative tools: posts revolve around classroom practices, teaching, and assignment design, and disputes over what counts as legitimate learning (teaching/class references are frequent in this topic, and cheating or plagiarism cues appear repeatedly). Users describe AI as both tutor and shortcut, often in the same breath, narrating how students draft essays, solve problems, or translate readings while instructors respond by changing task formats, tightening assessment, or demanding process evidence. A recurring communicative

move is moral boundary work: some posts normalize AI as a study aid, while others frame it as an integrity threat that undermines evaluation and devalues credentials. In English paraphrase, a typical stance reads: "If the exam only tests memorization, of course, students will outsource it; redesign the assessment." Another counter-stance warns: "AI makes copying effortless; universities need new verification norms." In short, education discourse is simultaneously pedagogical, regulatory, and affectively ambivalent, with neutrality carrying practical reportage and negative affect marking disputes over fairness, discipline, and institutional lag.

3.6. Labor Market, Economy, and Business Impacts

Work and economic implications constitute a similarly large formation (n = 5,954) and display a mixed but clearly consequential emotional structure: Neutral (1,915; 32.16%) and Happy (1,889; 31.73%) dominate, Delighted is meaningful (782; 13.13%), and high intensity negativity is nontrivial, especially Furious (891; 14.96%) relative to Angry (477; 8.01%). Qualitatively, the Texts frame AI as a pressure point for livelihoods and organizational power. Some posts treat AI as a competitive advantage and a pathway to new income streams, emphasizing speed, output, and market positioning, while others interpret it as a mechanism for restructuring labor, concentrating value, or hollowing out entry-level pathways. Users frequently talk about hiring, job search, and pay in concrete terms, and business-oriented posts often evaluate whether firms should replace contractors, compress teams, or demand higher output with the same wages. The tone shifts depending on who is imagined as the beneficiary: entrepreneurs and managers are more likely to speak in an optimizing register, whereas workers and recent graduates more often voice anxiety or indignation. A common English paraphrase is: "AI will not replace you, but someone using AI will," contrasted with: "Companies will use AI to squeeze workers and call it innovation." From a communication perspective, this topic is where economic imaginaries become moral narratives about responsibility, fairness, and who absorbs risk when productivity is reframed as a personal obligation rather than an institutional decision.

3.7. AI for Programming and Software Development

Developer-facing AI discussion (n = 1,434) centers on the practicalities of coding with models and the evolving identity of the programmer. Its

affective profile is anchored in Neutral (506; 35.29%) and Happy (440; 30.68%), with Delighted (173; 12.06%) and a comparable share of negative sentiment (Angry 161; 11.23%, Furious 154; 10.74%), suggesting a pragmatic field where appreciation coexists with ongoing frustration. Reading the Texts, users treat AI as a pair programmer that excels at boilerplate, debugging assistance, and explanation, but fails in ways that matter for production work: hallucinated APIs, brittle logic, missing edge cases, and confident wrong answers. Debug and bug-related language is common, and many posts emphasize the need for verification, tests, and human judgment rather than blind copying. The discourse repeatedly draws a line between code generation and software engineering: users note that models can propose functions quickly, yet struggle with requirements gathering, architecture, and maintaining coherence across a larger system. A typical English paraphrase is: “It wrote the code in seconds, then I spent an hour fixing subtle mistakes,” or “Great for scaffolding, risky for final logic unless you know what you are doing.” Importantly, this topic also hosts future-oriented claims about developer status, but expressed through everyday practice: debates over whether fundamentals become more or less important, and whether AI shifts the valued skill from writing code to specifying, reviewing, and integrating it.

3.8. Content Creation and Creativity

Posts tagged in this theme (n = 2,135) frame generative AI primarily as a practical cultural instrument for writing, translating, designing, and remixing, with affect clustering around Neutral (739; 34.61%) and Happy (671; 31.43%), and a smaller but visible Delighted band (262; 12.27%). Negative affect is present yet not dominant (Angry 215; 10.07%, Furious 248; 11.62%), often surfacing when users discuss quality collapse, sameness, or perceived devaluation of human craft. Reading the Texts suggests three recurring communicative genres: first, AI as a fast stylistic assistant (captions, resumes, formal letters, academic paraphrase), where users foreground efficiency and controllability; second, AI as a vernacular studio for images and short videos, frequently described through trial-and-error narratives about what the model can and cannot render; third, playful creativity (poetry, song-like lines, comedic remixes) that positions AI outputs as shareable artifacts in the attention economy. A typical English paraphrase of these posts is instruction-like and performative: “Rewrite this in a formal tone, then give

me three punchy versions,” or “I asked it to generate a poster concept, and it nailed the layout, but the details were wrong.” Across these patterns, creativity is not treated as an inner essence but as a platformed workflow: users negotiate originality, attribution, and taste in public, using AI outputs as both tools and talking points.

3.9. Deepfakes, Misinformation, and Verification

This topic (n = 1,504) carries one of the most conflictual affective profiles in the study, with negativity unusually concentrated (Angry 275; 18.28% and Furious 284; 18.88%, totaling 37.16%), alongside Neutral (418; 27.79%) and Happy (393; 26.13%). Qualitatively, the Texts portray synthetic media as an epistemic stress test for platform-publics: users circulate warnings about manipulated videos and voice cloning, dispute authenticity in real time, and debate what forms of evidence remain trustworthy when “seeing” and “hearing” become cheap to counterfeit. Many posts read as rapid vernacular fact-checking, either advising caution before resharing or offering informal heuristics (reverse search, source triangulation, “wait for confirmation”). An English example that captures the tone is: “Assume it is AI until proven otherwise, do not amplify it,” or “If a voice note can be generated, what counts as proof anymore?” At the same time, a smaller share treats deepfakes as spectacle or novelty, which helps explain the continued presence of Happy and Neutral posts. From a communication perspective, this theme shows how verification labor becomes distributed: responsibility is pushed outward to users and communities through low-cost alerts, while outrage functions as a mechanism for disciplining careless amplification and reasserting norms of evidentiary responsibility on X.

3.10. Ethics, Data Governance, Privacy, and Regulation

Governance-oriented discussion (n = 2,284) is comparatively balanced but persistently normative, led by Neutral (753; 32.97%) and Happy (740; 32.40%), with Delighted (244; 10.68%) and a meaningful Furious layer (319; 13.97%) that exceeds Angry (228; 9.98%). Reading the Texts indicates that “ethics” is rarely abstract philosophy here; it is articulated through concrete disputes over accountability, safety, and institutional lag. Users argue about whether platforms and states should regulate AI more aggressively, whether training data practices violate consent or ownership, and how to treat harms ranging from privacy leakage to discriminatory or dangerous outputs. Some

posts adopt a policy talk register (“We need clear rules, audits, and enforcement”), while others personalize risk through everyday cautionary advice, as in: “Do not upload private documents to a model you do not control,” or “If a system can be biased, governance is not optional.” There is also a recurring tension between openness and restriction: a pragmatic pro-access stance framed around learning and productivity coexists with calls for limits framed around safety, surveillance, and rights. Analytically, this topic is where Persian-language users most explicitly perform public reasoning about sociotechnical order, distributing blame and responsibility across developers, platforms, regulators, and end users while negotiating what “responsible use” should mean in practice.

3.11. Sensitive Social Topics and Mental Health/Psychology

This is the smallest thematic cluster in the dataset ($n = 480$), yet it is among the most affectively polarized, with Happy (135; 28.13%) and Angry (135; 28.13%) tied as the modal categories, followed by Neutral (115; 23.96%), Furious (68; 14.17%), and a small Delighted share (27; 5.63%). Reading across the posts, AI is positioned as a moral and intimate interlocutor: users test models on culturally charged issues (especially around sexuality, religion, and social norms), circulate model answers as proof of bias or “ideological programming,” and debate whether a chatbot can be trusted as an epistemic authority in domains where legitimacy is contested. Alongside this, a distinct strand treats AI as an informal mental health support tool, with users describing chatbots as a low-barrier “listener” for stress, loneliness, or everyday coping, while others react sharply against what they perceive as an irresponsible or unsafe substitution for professional care. In English paraphrase, posts often take the form of public experiments (“I asked the model about X; here is what it said”) or warnings (“Do not treat a chatbot like a therapist”). Communication-wise, this topic shows how AI talk becomes a proxy for broader struggles over values and vulnerability: affect intensifies when AI is imagined not as a tool but as a social actor whose answers may legitimate, stigmatize, or emotionally steer users.

AI as a boundary object: The Persian X discourse

Table1. Distribution of topics by sentiments

Topic Title	Total Posts	Furious	Angry	Neutral	Happy	Delighted
AI Tools Ecosystem and Model Comparisons	6324	592	924	2190	2140	478
Labor Market, Economy, and Business Impacts	5954	891	477	1915	1889	782
Education, Universities, and Schools	5648	729	579	1968	1719	653
Future Narratives, Scenarios, and Predictions	3131	334	147	1098	1062	490
Ethics, Data Governance, Privacy, and Regulation	2284	319	228	753	740	244
Content Creation and Creativity	2135	248	215	739	671	262
Deepfakes, Misinformation, and Verification	1504	284	275	418	393	134
AI for Programming and Software Development	1434	154	161	506	440	173
Sensitive Social Topics and Mental Health/Psychology	480	68	135	115	135	27
Prompting Practices and Strategies	316	18	20	103	117	58

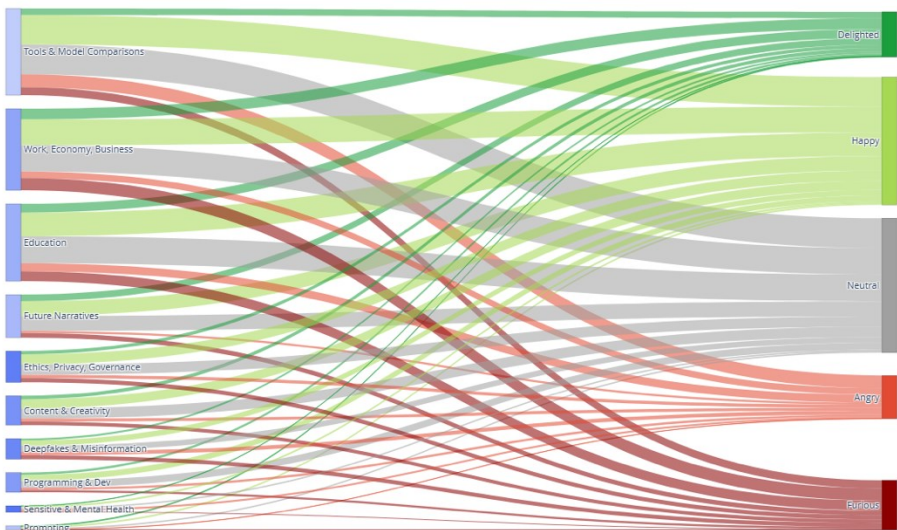


Figure 4. Sankey diagram of topic-sentiment relationships

Radar — Sentiment Intensity per Topic

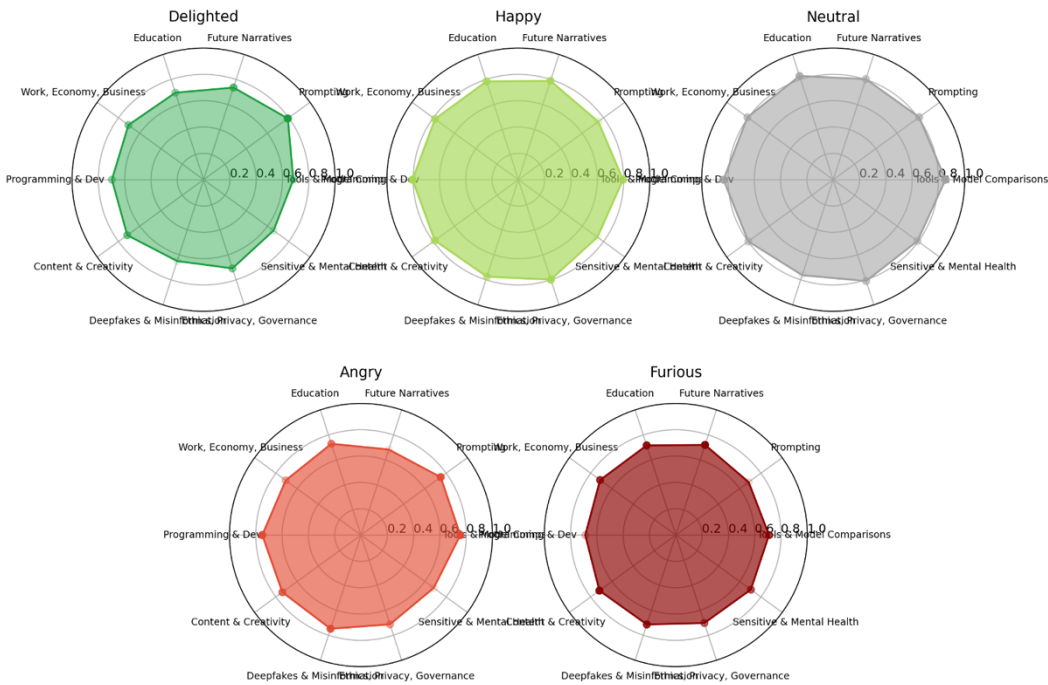


Figure 5. Radar-sentiment intensity per topic

3.12. Topic Co-occurrence Network

The topic co-tag network, built from multi-label assignments across the 24,215-post corpus, shows that Persian-language AI discourse on X is organized less as isolated “issue publics” and more as an interconnected field in which practical talk repeatedly folds into institutional and political-economic stakes. In raw co-tags, the strongest tie links Education with Work/Economy (n = 1,841), followed by Future Narratives with Work/Economy (n = 1,249) and Tools/Model Comparisons with Work/Economy (n = 1,139); Work/Economy also connects strongly to Ethics/Privacy/Governance (n = 853) and, at a slightly lower level, Tools with Education (n = 774). These counts already suggest that “work” functions as the discourse’s structural backbone: even when posts begin as tool chatter or classroom talk, they are frequently reframed through labor, productivity, credential value, and organizational advantage. To ensure this pattern is not merely an artifact of topic size, we also inspected relative overlap using the

Jaccard coefficient, which confirms the same ordering: Education ↔ Work/Economy shows the highest overlap (Jaccard ≈ 0.189), followed by Future ↔ Work/Economy (≈ 0.159) and Work/Economy ↔ Ethics/Privacy/Governance (≈ 0.116); Tools ↔ Work/Economy remains substantial on this scale as well (≈ 0.102). Interpreted through communication theory, this clustering indicates that users treat AI as a skills-and-institutions problem: debates about classroom use are rhetorically tethered to employability and inequality, and governance talk is pulled toward workplace consequences rather than remaining a purely abstract “ethics” register. A second, more content-specific bridge appears around authenticity: Content Creation co-occurs with Deepfakes/Misinformation at a comparatively high relative rate ($n = 305$; Jaccard ≈ 0.091), suggesting that creative adoption and verification anxiety travel together, as when synthetic media is discussed simultaneously as a tool for production and as a threat to evidentiary trust. Finally, some ties are near-absent, marking the network’s boundaries: Prompting rarely overlaps with Sensitive Social/Mental Health content ($n = 1$; Jaccard ≈ 0.0013) and only weakly with Deepfakes ($n = 7$; ≈ 0.0039), implying that “how-to” prompting craft tends to circulate as a specialized competence discourse rather than as a gateway into the most contentious moral or psychological domains.

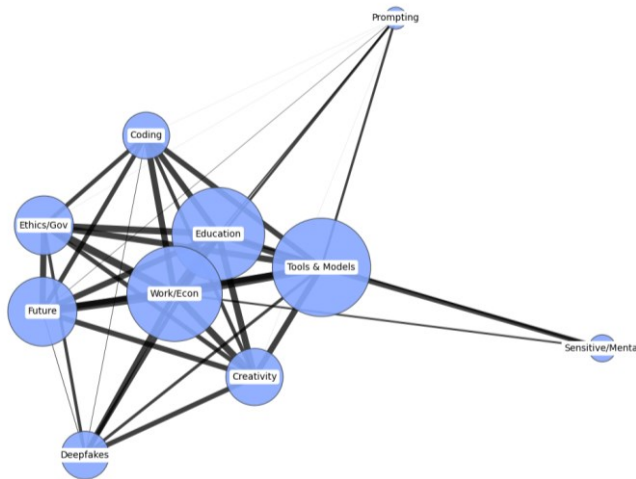


Figure 6. Co-occurrence network

4. Conclusion

The discourse surrounding artificial intelligence is not merely a collection of opinions—it is a living archive of how a society thinks, feels, and argues about its future. This study has shown that Persian-language users on the X platform engage with AI not as passive observers of technological change, but as active participants in the social negotiation of meaning, legitimacy, and moral consequence. These publics do not merely adopt or reject AI; they narrate it into being, fold it into institutional rhythms, and contest its significance within the constraints and affordances of platformed expression.

In this sense, AI operates less as an object and more as a mirror, a medium through which people project fears, hopes, grievances, and competencies. The pragmatic focus on tools and skills, the anxious debates about education and employment, the fascination with synthetic creativity, and the contentious disputes over governance and verification—all reflect the broader work of cultural sense-making under conditions of epistemic volatility. AI becomes a boundary object not because of its technical ambiguity, but because it is emotionally and socially overdetermined. It sits at the intersection of what is known and what is imagined, what is experienced and what is speculated.

The affective structure of the discourse reveals this ambivalence. The preponderance of pragmatic approval and the persistence of indignation are not contradictory; they are co-constitutive. They mark the simultaneous domestication and disruption of cognitive authority: AI is useful, but it is also unsettling. It accelerates productivity while destabilizing the moral architecture of effort, credit, and expertise. Publics do not simply ask what AI can do—they ask who is authorized to speak about it, to benefit from it, to be displaced by it. They ask what counts as knowledge in a world where synthetic fluency begins to mimic human intention.

But perhaps most importantly, this study underscores that meaning is not inherent in the machine; it is forged in discourse. Through hashtags and quote tweets, jokes and jeremiads, users craft a vernacular epistemology—one that does not merely receive expert framings of AI but reconstitutes them in situated, emotionally charged, and politically aware terms. These acts of everyday reasoning, though often informal, are consequential. They are how

publics rehearse governance in the absence of regulation, how they scaffold legitimacy in the absence of trust.

In analyzing a non-Anglophone, under-theorized context, this study also intervenes in a dominant tendency within AI scholarship: the unspoken universalism of English-language data, elite policy framings, and normative ethics. What Persian-language discourse reveals is not only difference, but *difference as method*—a reminder that cultural specificity is not noise but signal. It is through the granular, the colloquial, and the affectively charged that the global contours of AI adoption and resistance take shape.

The questions animating this discourse—what should count as truth, labor, risk, and authority in the age of intelligent machines—are not merely technological. They are ontological. They ask what it means to be human when cognition can be simulated, when language can be generated, and when attention can be automated. These are not questions that platforms can answer, but they are questions that platforms force us to ask—repeatedly, publicly, and often urgently.

References

- Balazadeh, K., & Kajonius, P. (2025). Exploring Intimacy with Artificial Intelligence: Validation of Robot Intimacy Receptivity Scale (RIRS). *International Journal of Social Robotics*, 1–13.
- Bourdieu, P. (1991). *Language and symbolic power* (J. B. Thompson, Ed.; G. Raymond & M. Adamson, Trans.). Harvard University Press.
- Bozdağ, A. A. (2025). The AI-mediated intimacy economy: A paradigm shift in digital interactions. *AI & Society*, 40(4), 2285–2306.
- Chu, W., Skirpan, M., Mir, K., & Gray, M. L. (2025). Illusions of intimacy: Analyzing emotional discourse in large-scale AI-user interactions. In *Proceedings of the 2025 ACM Conference on Human Factors in Computing Systems (CHI)*. <https://arxiv.org/abs/2505.11649>
- Elliott, A. (2023). Algorithmic intimacy: Digital companionship and the new emotional order. *Information, Communication & Society*. <https://doi.org/10.1080/1369118X.2023.2282554>
- George, A. S., George, A. H., Baskar, T., & Pandey, D. (2023). The allure of artificial intimacy: Examining the appeal and ethics of using generative AI for simulated relationships. *Partners Universal International Innovation Journal*, 1(6), 132–147.
- Illouz, E. (2007). *Cold intimacies: The making of emotional capitalism*. Polity Press.
- Innis, H. A. (1951). *The bias of communication*. University of Toronto Press.
- Lupton, D. (2022). “Sharing Is Caring:” Australian self-trackers' concepts and practices of personal data sharing and privacy. *Frontiers in Digital Health*, 3, 649275.
- McLuhan, M. (1964). *Understanding media: The extensions of man*. McGraw-Hill.
- Papacharissi, Z. (2015). *Affective publics: Sentiment, technology, and politics*. Oxford University Press.
- Reineke, M. J. (2022). The touching test: AI and the future of human intimacy. *Contagion: Journal of Violence, Mimesis, and Culture*, 29, 123–146.
- Sabbar, S., & Habib Zadeh Khiyaban, S. (2023). Algorithms of Displacement: Emotional and Rhetorical Responses to AI-Driven Job Loss in Digital Public Discourse. *International Journal of Advanced Multidisciplinary Research and Studies*, 3(4), 1324-1331. <https://doi.org/10.62225/2583049X.2023.3.4.5012>
- Salehi K, Habib Zadeh Khiyaban S, Sabbar S. (2026). Artificial Intelligence and Crime Detection: A Critical Review. *Cyberspace Studies*. 10(1): 181-197. <https://doi.org/10.22059/jcss.2025.402206.1179>
- Seaver, N. (2022). *Computing taste: Algorithms and the makers of music recommendation*. University of Chicago Press.
- Shahghasemi, E. (2025). AI; A Human Future. *Journal of Cyberspace Studies*, 9(1), 145-173. <https://doi.org/10.22059/jcss.2025.389027.1123>

- Shahghasemi, E., Gholami, F., & Alikhani, Z. (2025). Global patterns of social media use and political sentiment. *Discover Global Society*, 3, 36. <https://doi.org/10.1007/s44282-025-00171-y>
- Sjoraida, D. F. (2025). AI and Emotional Intimacy: Exploring Romantic Bonds with Artificial Companions. *Humanexus: Journal of Humanistic and Social Connection Studies*, 1(8), 333-342.
- Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, "translations" and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19(3), 387-420.
- Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.
- van Dijk, T. A. (1993). Principles of critical discourse analysis. *Discourse & Society*, 4(2), 249-283. <https://doi.org/10.1177/0957926593004002006>



Original-Forschungsarbeit

Künstliche Intelligenz und zwischenmenschliche Beziehungen im Iran: Kulturelle und soziale Herausforderungen

Shahnaz Khademizadeh¹, Sam Clarke^{2*}, Zeinab Mohammadi³

¹ Professorin für Wissens- und Informationswissenschaft, Shahid-Chamran-Universität Ahwaz, Ahwaz, Iran

² Dozent für Primarstufen-Lehrerbildung (Primary ITE), York St John University, York, Vereinigtes Königreich

³ Promotion in Wissens- und Informationswissenschaft, Shahid-Chamran-Universität Ahwaz, Ahwaz, Iran

Empfangen: 10. März 2025 Akzeptiert: 9. Juni 2025

Zusammenfassung:

Diese Studie untersucht die vielschichtigen Auswirkungen der Künstlichen Intelligenz (KI) auf zwischenmenschliche Beziehungen in der iranischen Gesellschaft und hebt die kulturellen, sozialen und psychologischen Herausforderungen hervor, die mit der raschen Verbreitung von KI-Technologien einhergehen. Mit der zunehmenden Integration von Instrumenten wie virtuellen Assistenten, Social-Media-Algorithmen und KI-gestützten Kommunikationsplattformen in den Alltag verändern sich Interaktionsmuster, emotionale Bindungen und kulturelle Normen grundlegend. Die Untersuchung basiert auf zwölf halbstrukturierten Interviews und folgt einem Mixed-Methods-Ansatz mit qualitativer Schwerpunktsetzung, einschließlich thematischer Analyse, Überprüfung der Interoder-Reliabilität und fallübergreifender Vergleichsanalyse. Die Ergebnisse zeigen eine doppelte Dynamik: Einerseits fördert KI Kommunikation, Produktivität und alltägliche Effizienz; andererseits schwächt sie direkte Face-to-Face-Interaktionen, emotionale Bindungen und traditionelle soziale Praktiken, die zentral für die iranische Kultur sind. Die Befunde weisen auf zunehmende Sorgen hinsichtlich geschwächter familiärer und gemeinschaftlicher Bindungen, abnehmender sozialer Kompetenzen, wachsender Abhängigkeit von intelligenten Systemen sowie generationsbedingter Unterschiede in der digitalen Anpassung hin. Darüber hinaus berichten die Teilnehmenden von breiteren kulturellen Veränderungen, darunter der Aufstieg virtueller Lebensstile, Bedrohungen der kulturellen Identität und eine wachsende soziale Ungleichheit infolge ungleicher Zugänge zu KI-Technologien. Die Studie identifiziert zudem psychologische Risiken wie Einsamkeit, oberflächliche Online-Verbindungen, verminderte Empathie und den Rückgang emotionaler Intelligenz im Zuge zunehmender Interaktionen mit algorithmischen Systemen. Auf gesellschaftlicher Ebene erzeugen Fragen des Datenschutzes, der Daten-Governance und ethischer Regulierung zusätzlichen Druck, der das öffentliche Vertrauen und die Dynamik zwischenmenschlicher Beziehungen beeinflusst. Die Untersuchung leistet einen Beitrag zu nationalen und internationalen Debatten über Mensch-KI-Interaktion, indem sie aufzeigt, wie globale Technologien mit lokalen kulturellen Kontexten interagieren. Sie argumentiert, dass ein ausgewogenes Verhältnis zwischen technologischer Innovation und der Bewahrung iranischer sozialer Werte entscheidend ist, damit KI die Grundlagen bedeutungsvoller menschlicher Beziehungen stärkt, anstatt sie zu untergraben.

Schlüsselwörter: künstliche Intelligenz, zwischenmenschliche Beziehungen, kulturelle Herausforderungen, soziale Dynamiken, iranische Gesellschaft

* Korrespondierender Autor

✉ s.clarke1@yorksj.ac.uk

🌐 <https://orcid.org/0009-0000-9297-3835>

Wie dieser Artikel zu zitieren ist:

Khademizadeh, Sh., Clarke, S., & Mohammadi, Z. (2025). AI and interpersonal relationships in Iran: Cultural and social challenges. *Spektrum Iran*, 38(2), 83-113.

📄 <https://doi.org/10.22034/spektrum.2026.554746.1043>

© 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

هوش مصنوعی و روابط بین فردی در ایران: چالش‌های فرهنگی و اجتماعی

شهناز خادمی‌زاده^۱، سم کلارک^{۲*}، زینب محمدی^۳

^۱استاد علم اطلاعات و دانش‌شناسی، دانشگاه شهید چمران اهواز، اهواز، ایران

^۲مدرس تربیت معلم ابتدایی (ITE)، دانشگاه یورک سنت جان، یورک، بریتانیا

^۳دکتری علم اطلاعات و دانش‌شناسی، دانشگاه شهید چمران اهواز، اهواز، ایران

دریافت: ۱۴۰۳/۱۲/۲۰؛ پذیرش: ۱۴۰۴/۰۳/۱۹

چکیده:

این پژوهش به بررسی تأثیرات چندوجهی هوش مصنوعی (AI) بر روابط بین فردی در جامعه ایران می‌پردازد و چالش‌های فرهنگی، اجتماعی و روان‌شناختی ناشی از گسترش سریع فناوری‌های مبتنی بر هوش مصنوعی را برجسته می‌کند. با نهادینه‌شدن ابزارهایی همچون دستیارهای مجازی، الگوریتم‌های شبکه‌های اجتماعی و پلتفرم‌های ارتباطی مبتنی بر هوش مصنوعی در زندگی روزمره، الگوهای تعامل، درگیری عاطفی و هنجارهای فرهنگی در حال دگرگونی هستند. این پژوهش بر پایه دوازده مصاحبه نیمه‌ساختاریافته و با رویکردی ترکیبی با غلبه کیفی انجام شده است که شامل تحلیل مضمون، سنجش پایایی میان‌گذاران و مقایسه میان‌موردی می‌شود. یافته‌ها روایتی دوگانه را نشان می‌دهد: از یک سو، هوش مصنوعی ارتباطات، بهره‌وری و سهولت امور روزمره را تقویت می‌کند؛ و از سوی دیگر، تعاملات چهره‌به‌چهره، پیوندهای عاطفی و شیوه‌های سنتی اجتماعی ریشه‌دار در فرهنگ ایرانی را تضعیف می‌سازد. نتایج بیانگر نگرانی‌های فزاینده درباره تضعیف روابط خانوادگی و اجتماعی، کاهش مهارت‌های اجتماعی، وابستگی به سامانه‌های هوشمند و شکاف نسلی در سازگاری دیجیتال است. مشارکت‌کنندگان همچنین به دگرگونی‌های گسترده فرهنگی، از جمله گسترش سبک زندگی مجازی، تهدید هویت فرهنگی و افزایش نابرابری اجتماعی ناشی از دسترسی نابرابر به ابزارهای هوش مصنوعی اشاره کردند. این مطالعه افزون بر این، مخاطرات روان‌شناختی همچون احساس تنهایی، ارتباطات سطحی آنلاین، کاهش همدلی و افت هوش هیجانی را در پی تعامل روزافزون افراد با سامانه‌های الگوریتمی شناسایی می‌کند. در سطح کلان اجتماعی نیز مسائل مربوط به حریم خصوصی، حکمرانی داده و چالش‌های اخلاقی، فشارهای مضاعفی ایجاد می‌کند که بر اعتماد عمومی و پویایی روابط اثرگذار است. این پژوهش با نشان دادن چگونگی تعامل فناوری‌های جهانی با بسترهای فرهنگی محلی، به مباحث ملی و بین‌المللی درباره تعامل انسان و هوش مصنوعی کمک می‌کند. در نهایت، تأکید می‌شود که ایجاد توازن میان نوآوری فناورانه و حفظ ارزش‌های اجتماعی ایرانی برای تضمین آن ضروری است که هوش مصنوعی به جای تضعیف، بنیان‌های روابط انسانی معنادار را تقویت کند.

واژگان کلیدی: هوش مصنوعی، روابط بین فردی، چالش‌های فرهنگی، پویایی‌های اجتماعی، جامعه ایرانی



Original Research Paper

AI and interpersonal relationships in Iran: Cultural and social challenges

Shahnaz Khademizadeh¹, Sam Clarke^{2*}, Zeinab Mohammadi³

¹ Professor of Knowledge & Information Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran

² Lecturer of Primary ITE, York St John University, York, UK

³ PhD in Knowledge and Information, Shahid Chamran University of Ahvaz, Ahvaz, Iran

Received: Mar. 10, 2025 Accepted: Jun. 09, 2025

Abstract

This study examines the multifaceted impact of artificial intelligence (AI) on interpersonal relationships within Iranian society, highlighting the cultural, social, and psychological challenges emerging from the rapid adoption of AI technologies. As tools such as virtual assistants, social media algorithms, and AI-driven communication platforms become embedded in daily life, they are reshaping patterns of interaction, emotional engagement, and cultural norms. Drawing on twelve semi-structured interviews analyzed through a qualitative-dominant mixed-methods approach, including thematic analysis, intercoder reliability checks, and cross-case comparison, the research identifies a dual narrative: AI enhances communication, productivity, and daily convenience, yet simultaneously undermines face-to-face engagement, emotional bonds, and traditional social practices central to Iranian culture. Findings reveal growing concerns about weakened family and community ties, reduced social skills, dependency on intelligent systems, and generational gaps in digital adaptation. Participants also noted broader cultural shifts, including the rise of virtual lifestyles, threats to cultural identity, and increased social inequality driven by uneven access to AI tools. The study further identifies psychological risks such as loneliness, superficial online connections, diminished empathy, and the perceived decline of emotional intelligence as individuals increasingly interact with algorithmic systems. At the societal level, privacy, data governance, and ethical challenges create additional pressures that shape public trust and relational dynamics. The study contributes to national and international debates on human-AI interaction by demonstrating how global technologies interact with local cultural contexts. It argues that balancing technological innovation with the preservation of Iranian social values is essential to ensuring that AI strengthens rather than erodes the foundations of meaningful human relationships.

Keywords: artificial intelligence, interpersonal relationships, cultural challenges, social dynamics, Iranian society

* Corresponding Author

✉ s.clarke1@yorks.ac.uk

🌐 <https://orcid.org/0009-0000-9297-3835>

How to Cite this Article:

Khademizadeh, Sh., Clarke, S., & Mohammadi, Z. (2025). AI and interpersonal relationships in Iran: Cultural and social challenges. *Spektrum Iran*, 38(2), 83-113.

🔗 <https://doi.org/10.22034/spektrum.2026.554746.1043>

© 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

1. Introduction

In recent years, artificial intelligence (AI) has become deeply embedded in everyday life, with more than 55% of the global population regularly interacting with AI technologies. The global AI market, valued at approximately US\$454.12 billion (AIRPM, 2024), includes tools such as virtual assistants, social media algorithms, and customer service chatbots. This widespread integration raises important questions about how AI is reshaping human relationships, particularly in culturally rich societies such as Iran. Public attitudes toward AI remain mixed; a survey found that 52% of respondents felt more concerned than excited about AI developments (Faverio & Tyson, 2023), reflecting anxieties about the possible erosion of authentic interpersonal connection in an increasingly digital world. AI refers to the simulation of human intelligence by machines, including technologies such as machine learning, natural language processing, and robotics (Nass & Moon, 2000; Nass & Brave, 2005). While these systems enhance efficiency and accessibility, they also pose challenges for the depth and authenticity of human interaction (Sundar & Lee, 2022). As AI becomes more integrated into daily communication and social experiences, understanding its influence on relationships is particularly important in societies like Iran, where interpersonal bonds and traditional values play a central role in social life.

The digital age has also reshaped human identity, with individuals increasingly viewed as informational beings constantly connected to global networks (Dominguez, 2014). Although this connectivity facilitates communication, it can also reduce face-to-face interaction and weaken social skills (Russell, 2019). In Iranian society, where hospitality, family cohesion, and community engagement are deeply valued, this shift may contribute to declining empathy, reduced social cohesion, and the gradual erosion of cultural norms. Research shows that strong social connections are essential for mental and emotional well-being, while social isolation can produce negative health outcomes (Amichai-Hamburger & Ben-Artzi, 2003; Hohenstein et al., 2023). AI-driven platforms, however, may encourage behaviours such as social comparison and reduced emotional engagement (Duan et al., 2022). Studies on cultural change in the digital era (Acerbi, 2020) further suggest that technological immersion can reshape identity, learning, and social practices.

This article argues that although AI offers convenience and new forms of communication, it also challenges the authenticity and quality of relationships in Iranian society. As reliance on AI grows, there is a risk of increasingly superficial interaction (Dehnert & Mongeau, 2022), as well as the loss of emotional nuance that supports empathy and trust (Lee, 2020). Navigating this evolving landscape requires balancing technological advancement with the preservation of cultural values and meaningful human connection.

To examine the impact of artificial intelligence on interpersonal relationships and analyse the cultural and social challenges arising from it in Iran. This can be further broken down into sub-objectives that the project aims to address:

1. Identify the positive and negative effects of artificial intelligence on human relationships.
2. Identify cultural changes resulting from the spread of artificial intelligence in society.
3. Identify social challenges related to the penetration of artificial intelligence.
4. Provide local solutions for managing cultural and social challenges.

While this research investigates the nature of AI's impact on cultural and social dynamics in Iran, its relevance to the wider field of research remains. Contemporary literature identifies several converging themes that situate a country-specific study like this within broader debates: first, the emergence of human-AI intimacy and socio-affective alignment (Laestadius, et al., 2022; Maples, et al., 2024), where scholars worry that increasingly personalized agents reshape emotional bonds, trust, and care practices beyond conventional human relationships (Kirk, et al., 2025). Work in human-centred AI and HCI also stresses the need to move from technocentric evaluations toward context-sensitive studies (Raaijmakers, 2019; Belic, et al., 2019) that account for local meanings, practices, and user agency (Raees, et al., 2024), precisely the methodological gap this study addresses. An additional large and growing body of research on algorithmic and cultural bias also highlights how ostensibly neutral systems reproduce social inequalities and cultural misunderstandings unless evaluated in diverse settings (Celik, et al., 2022;

Feffer, et al., 2023). It is within these broader fields of research that this study situates itself, exploring a topic that remains under active investigation.

2. Methodology

This study employed a qualitative-dominant mixed-methods approach to investigate the cultural and social impacts of artificial intelligence (AI) on interpersonal relationships in Iran. The methodological design combined thematic analysis of semi-structured interviews with quantitative coding and frequency counts to ensure both depth and breadth in interpreting participants' perspectives. Mixed-methods approaches are particularly valuable in sociocultural research as they allow researchers to capture both subjective meaning-making processes and quantifiable patterns of responses, thereby increasing validity and explanatory power (Teddlie & Tashakkori, 2009; Creswell & Plano-Clark, 2018).

2.1. Data Collection

Data were collected through twelve semi-structured interviews using a purposive sampling method (Palinkas et al., 2013). This method was used to select the sample for the present study. was used to select the sample of the present study. In this method, the participants were selected and handpicked by the researcher because they either clearly have the phenomenon or characteristic of interest or are rich in information on a specific issue. In fact, this method is most often used when there is a need for expert samples (Palinkas, et al., 2013). In purposive sampling, it is not possible to determine in advance the number of participants required in the study in order to fully identify the phenomenon under study (Morgan, 1997). Interviews were conducted iteratively and continued until thematic saturation was reached, which refers to the point at which no new codes, concepts, or thematic insights emerged from additional data collection (Glaser & Strauss, 1967; Guest, et al., 2006). In this study, saturation was operationalised following the principles outlined by Guest, et al. (2006): (1) monitoring the emergence of new data within each interview; (2) comparing new data with previously identified categories; and (3) determining whether subsequent interviews contributed novel information relevant to the research questions.

To assess saturation systematically, after each interview, the researcher analyzed the collected data (i.e., opinions expressed by the participant) and examined whether new meaning units or categories appeared. Once several consecutive interviews produced no additional themes, sub-themes, or refinements, it was determined that thematic saturation had been achieved. This approach is consistent with Saunders, et al.'s (2018) conceptualisation of saturation as the point at which further data cease to add value to the conceptual development of the analysis.

2.2. Sampling

A sample refers to a subset of a population (Tailor, 2005), while 'sampling' is the technique used by researchers to select a manageable representation of that group (Sharma, 2017). The sample consisted of 12 participants, who were selected based on their experience, expertise in artificial intelligence, and willingness to participate. The list of sample participants is presented in Table 1. All interviews were recorded, transcribed verbatim, and anonymised to protect participants' identities, following established ethical practices for qualitative research (Wiles, 2013). This study utilised a combination of convenience, purposive, and snowball sampling methods (Cohen, et al., 2007) to gather participants. The researcher contacted colleagues at educational institutions who met the study's criteria, some of whom had a prior professional relationship with the researcher (Oppong, 2013). This convenience sampling method, though less rigorous than random sampling (Oppong, 2013), is frequently used in qualitative research (Dörnyei, 2007) and remains a valid approach for participant selection (Abedsaeidi & Amiraliakbari, 2015). Additionally, the researcher contacted colleagues who expressed an interest in the research area, forming a purposive sample (McChesney & Aldridge, 2019) to enhance the quality of data collected.

Table 1. List of Participants in the Qualitative Study

Rank	Field of Specialization	Gender	Interview Code
Associate Professor	Social Sciences Department	Man	M1
Assistant Professor	Social Sciences Department	Man	M2
Assistant Professor	Computer Engineering-Artificial Intelligence Orientation	Man	M3
Assistant Professor	Information Science and Knowledge Department	Man	M4

Rank	Field of Specialization	Gender	Interview Code
Assistant Professor	Computer Science	Woman	M5
PhD	Computer Engineering-Artificial Intelligence-Machine Learning	Man	M6
Masters	Computer Engineering, Artificial Intelligence and Robotics	Man	M7
Assistant Professor	Computer Engineering, Artificial Intelligence Orientation	Man	M8
PhD	Computer Engineering, Artificial Intelligence Orientation	Man	M9
PhD	Social Sciences Department	Man	M10
PhD	General Psychology Department	Man	M11
Associate Professor	Educational Technology Department	Man	M12

The use of convenience and purposive sampling, while practical for the study, introduces notable limitations regarding the representativeness and generalisability of the findings. Convenience and purposive sampling particularly constrain external validity: such samples can only be generalised to the subpopulation from which the sample is drawn and not to the entire population (Andrade, 2020). Estimates from convenience samples may also be biased, as participants are likely to differ systematically from the broader target population (Jager, et al., 2017). Consequently, the positive perceptions of AI's integration into Iranian society analyzed in this study may predominantly reflect the views of individuals who are more digitally engaged and already interested in AI ethics, potentially underrepresenting those who are less technologically proficient and/or more hesitant.

The selection of participants and the anonymisation of their data followed the Iranian National Ethical Guidelines for Research in the Humanities approved by the Ministry of Science, Research and Technology. Before each interview, the research purpose, participation requirements, and the right to withdraw at any stage were clearly explained verbally to all participants, including scholars from humanities and engineering fields. Explicit verbal consent for audio recording was obtained. In four cases where participants declined recording, interviews were documented through accurate handwritten notes. Because interviews were conducted in person and participants preferred an efficient process, verbal consent was used; however, all principles of informed consent, information, comprehension, voluntariness, and capacity were fully respected. Participant confidentiality

was ensured by removing all identifiable details, including names, affiliations, and positions, and replacing them with non-traceable codes. Raw data (audio files and written notes) were stored on an encrypted, password-protected device, with access restricted solely to the principal researcher who conducted the interviews.

The final sample included eleven men and one woman, creating a clear gender imbalance which warrants acknowledgement (Weber et al., 2021). This resulted from purposive, voluntary recruitment rather than intentional exclusion; such recruitment strategies can yield uneven response rates and remain acceptable when aligned with study aims and context (Patton, 2002; Sharp, 2003). However, the male-dominant sample may limit transferability, as gender can shape experiences and interpretations relevant to the topic, and the under-representation of women risks omitting gender-specific perspectives (Weber et al., 2021). Sample adequacy was assessed using qualitative standards of information power and thematic saturation, prioritizing analytic depth and recurring patterns over simple participant numbers (Guest et al., 2006; Malterud et al., 2016). To address limitations in reporting and interpretation, gender was treated as an explicit contextual factor in the analysis, and findings are presented cautiously with respect to transferability (Tracy, 2010). Future research should purposively recruit a more gender-balanced or stratified sample to examine potential gender differences more fully (Patton, 2002; Sharp, 2003).

2.3. Preparing an Interview

Semi-structured interviews are widely recognised for their flexibility in allowing participants to express perspectives in depth, while also providing a consistent structure for cross-comparison (Kvale & Brinkmann, 2015). The interview guide included open-ended questions aimed at four objectives: exploring the positive and negative effects of AI on interpersonal relationships, examining cultural changes resulting from AI's spread in Iranian society, identifying emerging social challenges, and eliciting locally appropriate solutions for managing these challenges.

Interview questions were developed based on the study's theoretical framework and research questions, following a semi-structured format (Kvale & Brinkmann, 2015) with an open-ended approach. The questions were designed to elicit information about factors indicating the impact of

artificial intelligence on interpersonal relationships and socio-cultural challenges.. The interview included questions such as:

1. Are you familiar with artificial intelligence technologies? If yes, which ones do you know or have you used?
2. How much do you deal with AI tools in your daily or professional life? Please give an example.

These questions were asked in general terms, and additional follow-up questions were posed as needed.

3. Analytical Framework

The analysis proceeded in two systematic stages. The first stage consisted of qualitative coding and thematic analysis. Each transcript was carefully read and manually coded according to the four research objectives, with emergent codes grouped into broader interpretive themes such as “reduced face-to-face interaction,” “cultural identity threats,” and “privacy concerns.” Thematic analysis was selected because it provides a rigorous yet flexible method for identifying, analyzing, and reporting patterns within qualitative data (Braun & Clarke, 2006). Representative quotations were extracted to preserve the nuance of participants’ voices and to capture contradictions or ambivalences in their accounts, aligning with best practices for thick description in qualitative research (Geertz, 1973). Additionally, responses were categorised as positive, negative, or neutral to highlight evaluative orientations toward AI.

3.1. Thematic Analysis

The first step in analyzing the data gathered in the semi-structured interviews was conducting a thematic analysis - a qualitative data analysis method that involves data collection, data familiarisation, coding and grouping of similar codes to derive themes (Braun & Clarke, 2019; McChesney & Aldridge, 2019) - which involved coding each response to determine if they correlated to the four pre-determined objectives (Krippendorff, 2019) of this article: (1) the positive and negative effects of AI on interpersonal relationships, (2) cultural changes arising from AI adoption, (3) emerging social challenges, and (4) locally appropriate solutions. Following a deductive reasoning paradigm (Braun & Clarke, 2019) this article

conducted thematic analysis within the confines of predetermined hypotheses (Attride-Stirling, 2001; Tuckett, 2005). If responses correlated with one (or more) of the four pre-determined objectives, they were coded and recorded under each category. Figure 1 illustrates the thematic analysis of the M1 interview responses.

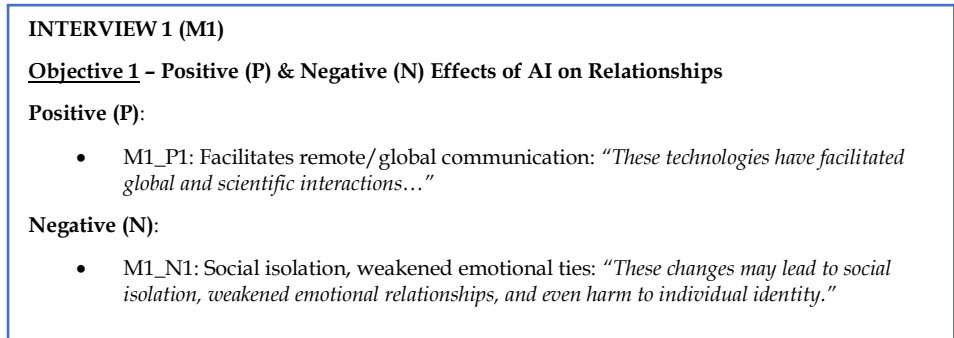


Figure 1. coding and thematic analysis of M1 interview transcript

The analysis of the study's data were informed by the researcher's own social position (Harkness et al., 2010; Kvale, 2007), which prompted ongoing self-reflexivity throughout the interpretive process. This involved critically considering how personal experiences and circumstances shaped the interpretation of the data (Alvesson & Sköldberg, 2018; Merriam & Grenier, 2019). Within an interpretivist framework, such reflexivity is essential, as it acknowledges that the construction of knowledge is inevitably influenced by the researcher's assumptions and perspectives (Alvesson & Sköldberg, 2018). Engaging in this reflective practice enabled the researcher to recognise the inherently subjective aspects of the work and the ways in which their own background and biases informed the research (Lincoln et al., 2011). This approach is especially valuable in small-N qualitative studies, where it supports systematic analysis while remaining attentive to contextual nuance (Yin, 2018).

After conducting the semi-structured interviews with each participant, their interview transcripts underwent thorough examination using NVivo 15 software, with the researcher identifying and documenting phrases from

them. These phrases were selected by the researcher, who is a formally trained educator, with a predisposition to analyse written work. Phrases were selected if they met one (or more) of the following criteria:

- *Significance and impact.* Phrases that reflect the deductive framework of the study and correlate with the four pre-determined objectives the researchers set out to investigate.
- *Clarity and conciseness.* Phrases that articulate complex ideas in a clear and concise manner, making them easier to understand and communicate.
- *Repetition of concepts.* Phrases that appear frequently within interview transcripts may indicate not only the emphasis placed on them by participants but also a consensus in the field, thus warranting particular attention.

Using this theoretical framework of data collection, the data were then organized into a thematic table (Table 2) that categorises responses according to four overarching objectives: 1) Positive and Negative Effects of AI on Relationships, 2) Cultural Changes in Iran, 3) Social Challenges, and 4) Local Solutions. Within each objective, sub-themes were identified by coding recurring concepts and statements across interviews, allowing for a comparison of similarities, differences, and unique insights among participants. Each sub-theme is illustrated with representative codes drawn from individual interviews, highlighting the frequency and context of key ideas. This approach enables a clear visualisation of trends, cross-cutting patterns, and areas of divergence, providing a structured framework for understanding the multifaceted impacts of AI on Iranian society.

Table 2. Cross-Interview Thematic Analysis

Theme	Sub-theme	Human (Principal) Coder: Identified words or phrases	AI (Secondary) Inter-coder Identified words or phrases
Positive Effects of AI on Relationships	Facilitates Communication	7	6
	Supports Professional & Academic Work	6	6
	Daily Convenience & Efficiency	6	5
	Emotional / Social Assistance	5	4
	Cultural Adaptation & Alignment	2	3

Theme	Sub-theme	Human (Principal) Coder: Identified words or phrases	AI (Secondary) Intercoder Identified words or phrases
Negative Effects of AI on Relationships	Reduced Face-to-Face Interactions	7	6
	Weakened Emotional/Family Bonds	10	9
	Dependency on Technology / Overreliance	7	8
	Misinformation / Trust Issues	4	4
	Isolation / Loneliness	4	4
	Reduced Human Skills	4	4
Cultural Changes in Iran	Shift to Digital / Virtual Lifestyles	5	5
	Youth as Early Adopters	5	4
	Cultural Identity Threatened	4	5
	Values and Social Norm Shifts	6	5
	Limited Cultural Penetration (Emerging)	3	4
Social Challenges	Inequalities / Digital Divide	8	6
	Dependency / Overreliance	7	6
	Privacy & Data Security	4	4
	Mental Health / Isolation	5	4
	Cultural & Social Norm Erosion	2	3
Local Solutions	Education & Awareness	11	9
	Cultural Adaptation / Local Content	9	7
	Policy, Regulation, & Ethics	9	8
	Infrastructure & Access Equality	5	4
	Self-Regulation / Responsible Use	4	4

3.2. AI as an Intercoder and Krippendorff’s Alpha

To enhance the reliability and validity of the thematic analysis, this study employed a generative AI large language model (LLM) as an alternative secondary intercoder to review the coding of interview transcripts. Following initial coding by the principal researcher, the LLM independently analyzed anonymised transcripts to identify recurring themes and sub-themes, providing a comparative check against human-generated codes. This approach draws on emerging literature demonstrating the utility of AI-assisted qualitative analysis for augmenting intercoder reliability (Wei et al., 2022), particularly in small-N studies where resource constraints limit multiple human coders (Liang, et al., 2022; Chew, et al., 2023). By cross-

referencing AI-generated codes with those of the researcher, discrepancies were identified, discussed, and resolved through iterative refinement (Roberts, 2020), akin to traditional double-coding practices (MacQueen, et al., 1998; O'Connor & Joffe, 2020). The AI-assisted process increased confidence in the consistency and validity of thematic categorisations while preserving the interpretive nuance central to both inductive and deductive qualitative approaches (Saldaña, 2021; Ziang, et al., 2023).

To evaluate consistency between a human coder and an AI-assisted intercoder, we conducted a systematic intercoder reliability analysis using Krippendorff's Alpha (2011; 2019). This method was selected because the coding scheme involved ordered categories representing the intensity or magnitude of identified evidence within each sub-theme (Hayes & Krippendorff, 2007). The dataset comprised 27 sub-themes across five major thematic domains related to AI integration in Iran (e.g., positive effects on relationships, negative effects, cultural change, social challenges, and local solutions). For each sub-theme, both the human coder and the AI intercoder recorded the number of words or phrases supporting its presence in interview data. Since Krippendorff's Alpha (2011;2019) requires categorical rather than raw count data, we transformed frequency counts into ordinal categories reflecting low, medium, and high evidence levels: Low (1): 0-3 identified phrases, Medium (2): 4-7 identified phrases, and High (3): 8-11 identified phrases. This transformation enabled use of an ordinal distance metric appropriate for Krippendorff's Alpha (2011;2019).

Each of the 27 sub-themes was assigned two ordinal values, one from the human coder and one from the AI intercoder, resulting in 27 paired observations. These pairs formed the basis for computing observed and expected disagreement. Following Krippendorff's approach (2011;2019) for ordinal data, we defined a **distance function** based on squared normalised differences:

$$\delta_{ij} = \left(\frac{|i - j|}{\max \text{ difference}} \right)^2$$

For a 3-point ordinal scale (1-3), the maximum possible difference is 2, giving:

- Agreement (difference = 0): $\delta = 0$
- 1-step distance (difference = 1): $\delta = 0.25$
- 2-step distance (difference = 2): $\delta = 1$

This weighting ensures that more severe coder disagreements contribute proportionally greater error, which in turn would provide an indication as to the validity of the human / AI / principal coder's findings.

3.2.1. Observed Disagreement (D_o)

Observed disagreement was calculated by summing the distance values for all human–AI coding differences across sub-themes and dividing by the total number of items. In the dataset, 22 cases showed perfect agreement, 5 cases showed a 1-point difference, and 0 cases showed a 2-point difference. Using the ordinal distance weights:

$$D_o = \frac{(5 \times 0.25)}{27} = 0.0463$$

This reflects very low observed disagreement between coders.

3.2.2. Expected Disagreement (D_e)

Expected disagreement represents the amount of disagreement that would occur by chance, based on how often each category was used across both coders. We combined all human and AI ratings (54 total observations) to compute category densities:

- Category 1: 5 occurrences
- Category 2: 37 occurrences
- Category 3: 12 occurrences

Proportions:

- $p_1 = 0.0926$
- $p_2 = 0.6852$
- $p_3 = 0.2222$

Expected disagreement is calculated using:

$$D_e = \sum_{i \neq j} p_i p_j \delta_{ij}$$

Accounting for both symmetrical pairs, the final expected disagreement was $D_e = 0.14905$.

3.2.3. Calculation of Krippendorff's Alpha

Krippendorff's Alpha (2011;2019) for ordinal data is computed as:

$$\alpha = 1 - \frac{Do}{De} \quad \alpha = 1 - \frac{0.0463}{0.14905} = 0.6895$$

Krippendorff's Alpha (2011;2019) for the dataset was **0.69**, which falls within the range generally interpreted as indicating moderate to substantial agreement (0.61–0.80). This result suggests that the principal human coder and the AI secondary intercoder showed a high level of consistency in their coding decisions once the frequency counts were converted into an ordinal scale, with only minor variations in the degree to which each sub-theme was coded was coded.

4. Findings

The analysis of twelve interviews indicates that AI is perceived as a transformative yet double-edged force in Iranian society, producing both notable benefits and challenges. Positive effects were reported by all participants, with facilitating communication the most common theme, coded in 7 of 12 interviews (58%), often in global, remote, or educational contexts. AI's support for professional and academic work appeared in 6 interviews (50 %), while daily convenience and efficiency were noted in 5 (42 %). Emotional or social assistance and culturally adaptive features were mentioned less frequently, in 5 interviews (42 %) and 2 interviews (17 %), respectively.

Conversely, negative effects were widespread: reduced face-to-face interaction appeared in 7 interviews (58%) and weakened emotional or family bonds in 7 (58%). Dependency on technology emerged in 7 interviews (58%), while concerns about misinformation or trust issues appeared in 4 (33%), and isolation or loneliness in 4 (33%). An additional 4 interviews (33%) highlighted reduced human skills such as empathy or critical thinking. In terms of cultural change, youth-driven adoption was coded in 5 interviews (42%), alongside shifts to digital lifestyles in 5 (42%). Perceived threats to cultural identity were raised in 4 interviews (33%), values or social norm shifts in 5 (42%), and limited cultural penetration in 3 (25%).

Social-level challenges were also evident: inequalities and the digital divide were cited in 8 interviews (67%), societal overreliance in 7 (58%), and privacy or data security concerns in 4 (33%). Mental health impacts such as isolation surfaced in 4 interviews (33%), and cultural or social norm erosion in 2 (17%). To mitigate these effects, participants proposed local solutions, with education and awareness programs mentioned in 9 interviews (75%), cultural adaptation and locally aligned AI content in 7 (58%), and policy, regulation, and ethics frameworks in 6 (50%). Infrastructure and access equality were raised in 5 interviews (42%), and responsible individual use in 4 (33%). Overall, while AI offers clear opportunities for connectivity and productivity, its broader social and cultural impacts require proactive, multi-level responses combining education, policy, and cultural adaptation.

The interview findings reveal a complex and ambivalent perception of artificial intelligence in Iranian society. When placed alongside existing Iran-focused scholarship, a coherent narrative emerges: AI is broadly viewed as an enabler of communication, education, and productivity, while simultaneously intensifying structural, cultural, and relational pressures. This duality reflects longstanding patterns in Iranian digital-society research, where technology adoption unfolds within uneven access, cultural negotiation, and state-society tensions. Across all twelve interviews, participants highlighted significant benefits. The most frequently noted was enhanced communication (58%), which aligns with Rahimi's (2015) argument that digital technologies often function as mediating tools allowing Iranians to bypass geographic and infrastructural constraints. For young people and globally dispersed families, AI-supported translation, content generation, and messaging expand communicative capacities beyond local linguistic or political boundaries. These uses echo Rajabi and Nasrollahi's (2023) findings on how AI-driven platforms facilitate more fluid cultural participation and self-expression, continuing earlier trends in Iranian digital media adoption.

Professional and academic assistance (50%) was another major theme. Participants described AI as improving productivity through drafting, research support, and information analysis, mirroring insights from Mahboudi, et al. (2017) on computers' role in widening educational opportunities. Their experiences also reflect Mahmoudi, et al.'s (2025) findings that AI enhances efficiency and innovation in Iranian knowledge-

based organisations. These benefits are especially meaningful in a context where access to global academic networks and current resources remains constrained. Participants also cited convenience and efficiency (42%), and emotional or social assistance (42%) and these benefits align with Atwood's (2025) argument that AI in Iran is framed as both a modernising force and a practical tool for coping with systemic limitations. In a society marked by economic pressure and heavy daily workloads, AI often functions as a mechanism for reducing cognitive and emotional burden.

Yet participants expressed an equally strong set of concerns. Reduced face-to-face interaction (58%) and weakened emotional or family bonds (58%) mirror cultural anxieties identified by Rajabi and Nasrollahi (2023), who show that AI-based platforms are often perceived as eroding communal norms and traditional interpersonal dynamics. Iran's strong family-oriented culture heightens fears of relational fragmentation, reinforcing broader societal narratives that present technology as both indispensable and culturally disruptive. Dependency on technology (58%) emerged as another central concern, with participants describing discomfort with how reliant they had become on AI, a sentiment resonant with Atwood's (2025) characterisation of a national "double bind": AI is embraced for its modernising potential but feared for its implications for autonomy and control. Individual dependency thus mirrors anxieties about systemic dependency on technologies developed outside Iran. Such concerns resonate with broader analyses of AI as a site of geopolitical contestation, where public discourse reflects tensions over technological sovereignty, global power asymmetries, and digital dependency (Salehi et al., 2025).

Concerns about misinformation, trust, and loss of human skills (each reported in 33% of interviews) map onto broader structural challenges. As Qadikolaei, et al. (2022) show, Iran's digital divide, marked by uneven access and variable digital literacy, limits many citizens' ability to critically evaluate online content. Participants' fear of diminished critical thinking or empathy thus reflects uneven digital skill development and inconsistent exposure to reliable information ecosystems. Cultural change was another major cluster. Youth-driven adoption (42%) and shifts toward digital lifestyles (42%) parallel the generational patterns identified by Rajabi and Nasrollahi (2023), wherein Iranian youth treat digital spaces as extensions of identity formation

and cultural engagement. Concerns about eroding cultural identity (33%) and shifting values (42%) correspond to Atwood's (2025) analysis of national narratives that depict AI as both necessary for modernisation and a potential threat to cultural autonomy. Three participants (25%) emphasised that AI tools often fail to reflect local cultural specificities, revealing tensions between globalised technologies and Iran's socio-cultural landscape.

Social-level concerns further emphasised inequality and unequal access to AI tools and digital resources (67%), which directly mirror Qadikolaei et al.'s (2022) findings on provincial disparities in digital access. Participants' accounts underscore how digital inequality shapes all other dimensions of AI usage and perception. Other issues including societal dependency (58%), privacy and security (33%), and erosion of social norms (17%), reinforce Rahimi's (2015) warning that rapid technological adoption without tailored regulation can produce cultural strain and intensify public anxiety. These fears about data control also align with Atwood's (2025) observation that Iranian public discourse often frames AI within geopolitical and surveillance concerns.

Overall, the findings depict AI in Iran as simultaneously empowering and disruptive. It expands communication, productivity, and opportunity, yet challenges cultural norms, widens inequalities, and raises concerns about dependency. When situated within Iranian scholarship, these perceptions reveal an ongoing negotiation between modernity, identity, and technological change. Similar patterns have been observed in digital public discourse, where emotional and rhetorical responses to AI-driven disruption function not merely as individual reactions but as mechanisms for constructing collective identity, negotiating institutional trust, and shaping shared imaginaries about technology's social role (Sabbar & Habib Zadeh Khiyaban, 2023). The evidence highlights the need for coordinated interventions in education, policy, and cultural localization to help ensure that AI supports social cohesion, cultural continuity, and equitable development.

5. Discussion

As these findings demonstrate, artificial intelligence (AI) has increasingly permeated daily life, transforming how individuals communicate, form relationships, and interact socially. The rapid evolution of AI technologies

has reshaped human practices, influencing both personal and communal dynamics. From digital communication to emotional companionship, AI's influence extends across multiple layers of social interaction, presenting both opportunities and challenges.

5.1. AI in Communication

Social media platforms rely heavily on AI algorithms to curate content, shape engagement, and personalize user feeds by analyzing interactions such as likes, shares, and time spent on posts (Banas et al., 2022). This personalisation optimises user engagement but also fosters echo chambers, limiting exposure to diverse viewpoints and reinforcing pre-existing beliefs (Duan et al., 2022). Such environments may contribute to the polarization of communities, affecting not only online discourse but also real-world social cohesion. Users increasingly interact with algorithmically curated content rather than engaging in substantive dialogue, raising concerns about the depth and authenticity of digital interactions.

AI-powered chatbots and virtual assistants have transformed professional and personal communication. In customer service, chatbots efficiently handle inquiries, provide immediate information, and resolve issues around the clock, reducing operational costs by up to 30% (Endacott & Leonardi, 2022; IBM, 2024). On a personal level, virtual assistants such as Siri, Alexa, and Google Assistant facilitate everyday tasks, enabling intuitive, natural-language interactions (Sundar, 2020). These advancements improve convenience and accessibility; however, they may also reduce human-to-human communication, potentially contributing to social isolation and dependence on technology (Brandtzaeg et al., 2022; Darioshi & Lahav, 2021).

5.2. AI in Relationships

AI has significantly altered romantic and emotional relationships. Dating applications such as Tinder and Bumble employ algorithms to analyse preferences, behaviours, and demographics, enhancing match predictions and streamlining partner selection (Laapotti & Raappana, 2022). While these tools increase opportunities for connection, they may also encourage superficial interactions and "choice overload," where an abundance of options results in decision fatigue and diminished satisfaction (Hancock et al., 2020).

Beyond dating, AI companions such as Replika and AI Dungeon provide emotional support and virtual companionship, particularly benefiting individuals who experience loneliness or social anxiety (Sundar & Chen, 2023). These AI-driven characters adapt to users' emotional cues, offering personalized engagement. While such companions can provide comfort and reduce perceived isolation, ethical questions arise regarding the authenticity of these relationships and the long-term impact on social skills (Hohenstein et al., 2023; Dehnert & Mongeau, 2022). Increasing reliance on AI for emotional support may inadvertently diminish real-life social competence, creating a paradox where individuals feel less lonely while simultaneously experiencing isolation from human interaction.

5.3. Decreased Face-to-Face Interactions

The rise of AI-mediated communication has coincided with a decline in face-to-face engagement. A survey by LivePerson (2021) reported that 65% of global participants reported communicating more digitally than in person; this figure rises to 74% in English-speaking countries. Younger generations, accustomed to texting and social media, often prefer these digital channels to traditional communication methods. While convenient, digital interaction may reduce social skills such as active listening, interpreting nonverbal cues, and engaging in spontaneous dialogue (Zhang et al., 2024).

Instant communication fosters expectations of immediate responses, sometimes generating pressure, misinterpretation, and conflict (Matlabinejad et al., 2023). However, in educational settings, AI can enhance communication. For example, AI-powered chatbots improve interaction quality in online learning, increasing comfort levels among students communicating with unfamiliar peers (Mostafa et al., 2024). AI instructor self-disclosure has been shown to foster emotional bonding, engagement, and positive learning experiences, highlighting the potential of AI to support interpersonal communication while maintaining pedagogical effectiveness (Tai, 2020; Zhang et al., 2024).

5.4. The Psychological Impact of AI on Relationships

Despite advances in natural language processing and machine learning, AI remains limited in interpreting human emotions. Unlike humans, who rely on subtle cues such as tone, facial expressions, and body language, AI

primarily depends on algorithms and data, often misreading or overlooking emotional states (Sundar & Chen, 2023; Nass & Moon, 2000). For instance, customer service chatbots may respond efficiently but fail to convey empathy, potentially exacerbating frustration (Banas et al., 2022).

Social media and digital communication also promote an illusion of connection. Online personas are often curated, presenting idealised versions of users that may not reflect reality (Guzman & Lewis, 2020; Duan et al., 2022). Interactions are frequently brief and superficial; accumulating friends or followers does not necessarily equate to meaningful engagement, which can negatively impact emotional health (Brandtzaeg et al., 2022; Hohenstein et al., 2023). Younger generations, in particular, may prioritise digital interactions over in-person connections, increasing susceptibility to anxiety and depression.

The paradox of feeling connected yet lonely is heightened by AI reliance for companionship, creating a cycle where individuals substitute machine interaction for authentic human engagement (Gunkel, 2012; Mijwil et al., 2022). During the COVID-19 pandemic, digital tools mitigated isolation but could not fully replicate the emotional richness of physical presence, potentially fostering a long-term preference for virtual communication (Tai, 2020).

5.5. The Impact on Family and Friend Dynamics

AI has reshaped family communication through digital platforms, enabling real-time coordination and interaction across distances (Gong et al., 2021). AI-driven features such as smart notifications and automated responses enhance efficiency but may also introduce misunderstandings due to the absence of tone, body language, and synchronous dialogue (Brito & Dias, 2020; Carvalho et al., 2015). Friendship dynamics have similarly evolved. AI and social media allow connections based on shared interests rather than physical proximity, expanding opportunities for social engagement (Druga et al., 2022; Garg et al., 2022). Algorithms may suggest potential friends, facilitating community formation, but online convenience can lead to superficiality, emphasising quantity over quality (Higgins, 2019). AI-mediated conflict resolution tools provide guidance but often lack the emotional intelligence necessary for genuine reconciliation (Lee & Yoon, 2021; De Tongi et al., 2021). Generational differences significantly influence AI adoption. Digital natives embrace AI features as integral to their social interactions, while older adults

may feel alienated or overwhelmed by technological changes, creating intergenerational disconnects (Liao et al., 2023; Hata et al., 2019; Galaz et al., 2021). Addressing these divides requires patience, empathy, and efforts to bridge traditional and digital communication modes.

5.6. The Ethical Implications of AI in Relationships

AI integration raises ethical concerns, particularly regarding privacy, consent, and dependency. Platforms collect extensive user data, including interactions, preferences, and emotional cues, often without explicit consent (Crawford, 2021; Bie, 2023). Dating apps exemplify this, tracking behaviours to enhance matchmaking while potentially breaching trust or exploiting sensitive information (Gan & Wang, 2024; Greene, 2020). Dependency on AI for emotional support presents further challenges. While AI provides immediate comfort, it lacks genuine empathy, potentially reducing individuals' engagement in authentic relationships (Wu, 2024; Marcos-Pablos & García-Peñalvo, 2022; Chen & Tang, 2024). Over-reliance may isolate individuals and encourage a preference for AI companionship, raising questions about the social consequences of substituting human interaction with artificial ones (Morgante et al., 2024; Epley et al., 2007; Petina et al., 2023).

5.7. Reclaiming Authenticity in Relationships

Maintaining authenticity in relationships requires balancing technology use with human connection, fostering emotional intelligence, and prioritizing meaningful engagement (Sundar & Lee, 2022; Hancock et al., 2020; McStay, 2018; Marcos-Pablos & García-Peñalvo, 2022; Atwood, 2025). Setting boundaries around AI and digital platforms, promoting face-to-face communication, and thoughtfully integrating technology can enhance rather than replace genuine interaction (Lee, 2020; Carvalho et al., 2015; Brandtzaeg et al., 2022; De Togni et al., 2021).

Establishing tech-free spaces, such as family meals or social gatherings, supports uninterrupted engagement and preserves nonverbal communication cues (Greene, 2020; Bie & Zeng, 2024; Deneke et al., 2021). Regular in-person interactions sustain empathy, active listening, and trust – essential for resolving conflict and building genuine bonds (McStay, 2020; Morgante et al., 2024; Epley et al., 2007). Community-based programmes and workshops on emotional skills, conflict resolution, and communication can

strengthen social cohesion and encourage authentic engagement (Marvin, 2006; Cha et al., 2022). Participation in social clubs, volunteer work, and group activities fosters belonging, mitigates digital isolation, and provides opportunities for face-to-face connection (Pentina et al., 2023; Hennig-Thurau et al., 2022; Wu, 2024). While reclaiming authenticity remains challenging due to miscommunication and evolving norms, open dialogue, adaptability, and intentional engagement are crucial to sustaining meaningful human relationships in an AI-driven society (Gremsl & Hödl, 2022; Marvin, 2006).

6. Conclusion

As artificial intelligence becomes increasingly embedded in everyday life, its influence on interpersonal relationships in Iranian society presents both opportunities and challenges. AI-driven technologies, such as social media algorithms, messaging platforms, and virtual assistants, have made communication faster and more accessible, allowing families and friends to stay connected across long distances. In a society where strong family ties, hospitality, and community cohesion are deeply valued, these tools can help sustain relationships despite migration, busy schedules, or geographic separation. However, the convenience of digital interaction also risks reducing the frequency and quality of face-to-face encounters that traditionally nurture emotional closeness and trust. The growing reliance on AI-mediated communication may subtly reshape how individuals express empathy, resolve conflict, and build intimacy. Interacting primarily through screens and algorithms can limit exposure to the emotional nuance, body language, and shared experiences that deepen relationships. For younger generations in particular, the ease of AI-driven interaction may encourage a shift away from long-standing social customs, potentially weakening traditions rooted in personal presence and communal engagement. There is also a risk that algorithmic curation of content and social networks will create echo chambers, narrowing perspectives and contributing to social polarization within communities.

Psychologically, AI offers both support and concern. Chatbots and digital companions can provide a sense of connection or assistance, yet they cannot fully replicate the emotional richness of human relationships. This may foster what researchers call an “illusion of connection,” where individuals feel

socially engaged online while experiencing loneliness or isolation offline. In a culture that places high importance on collective belonging and interpersonal warmth, such a disconnect could have meaningful implications for well-being. To navigate this evolving landscape, a balanced approach is essential. Promoting digital literacy, emotional awareness, and critical engagement with AI can help individuals use technology thoughtfully rather than dependently. Encouraging community spaces, family gatherings, and in-person interaction can preserve the relational depth central to Iranian culture. Ultimately, AI may be most beneficial when embraced as a supportive tool that enhances communication and connection without replacing the human bonds that sustain social life and cultural continuity.

Conflict of Interest

The author declares no conflict of interest.

Funding

No funding agency or institution influenced the research design, analysis, or interpretation of results.

References

- Abedsaeidi J., & Amiraliakbari S. (2015). *Research Method in Medical Sciences and Health*. Salemi
- Acerbi, A. (2020) *Cultural Evolution in a Digital Age*. Oxford University Press.
- Afifi, T. D., Zamanzadeh, N., Harrison, K., & Callejas, M. A. (2018). WIRED: The impact of media and technology use on stress (cortisol) and inflammation (interleukin IL-6) in fast-paced families. *Computers in Human Behavior*, 81, 265–273. <https://doi.org/10.1016/j.chb.2017.12.010>
- AIPRM. (2024). *AI Statistics 2024*. <https://www.aiprm.com/en-gb/ai-statistics/>
- Amichai-Hamburger, Y., & Ben-Artzi, E. (2003). Loneliness and Internet use. *Computers in Human Behaviour*, 19(1), 71–80. [https://doi.org/10.1016/S0747-5632\(02\)00014-6](https://doi.org/10.1016/S0747-5632(02)00014-6)
- Andrade, C. (2020). The Inconvenient Truth About Convenience and Purposive Samples. *Indian Journal of Psychological Medicine*, 43(1), 86–88.
- Attride-Stirling, J. (2001). Thematic networks: An analytical tool for qualitative research. *Qualitative Research*, 1(3), 385–405
- Atwood, B. (2025). Artificial Intelligence in Iran: National Narratives and Material Realities. *Iranian Studies*, 1–18. <https://doi.org/10.1017/irn.2024.63>
- Banas, J., Palomares, N., Richards, A., Keating, D., Joyce, N., & Rains, S. (2022). When machine and bandwagon heuristics compete: Understanding users' response to conflicting AI and crowdsourced fact-checking. *Human Communication Research*, 48(3), 430–461. <https://doi.org/10.1093/hcr/hqac010>
- Belic, M., Bobic, V., Badza, M., Solaja, N., Djuric-Jovicic, M., & Kostic, V. (2019). Artificial intelligence for assisting diagnostics and assessment of parkinson's disease - a review. *Clinical Neurology and Neurosurgery*, 184, 105442. <https://doi.org/10.1016/j.clineuro.2019.105442>
- Bie, J. H. (2023). Platformized digital interactions: Affective practices based on the availability of technology. *Young Journal*, 4, 22–25. <https://doi.org/10.15997/j.cnki.qnjz.2023.04.007>
- Bie, J. H., & Zeng, Y. T. (2024). Algorithmic imagination of platform participation and affective networks: An analysis of users on Xiaohongshu. *China Youth Research*, 2, 15–23. <https://doi.org/10.19633/j.cnki.11-2579/d.2024.0018>
- Brandtzaeg, P., Skjuve, M., & Følstad, A. (2022). My AI friend: How users of a social chatbot understand their human–AI friendship. *Human Communication Research*, 48(3), 404–429. <https://doi.org/10.1093/hcr/hqac008>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), (2006). 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis." *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597.

- Brito, R., & Dias, P. (2020). Which apps are good for my children? How the parents of young children select apps. *International Journal of Child-Computer Interaction*, 26, 100188. <https://doi.org/10.1016/j.ijcci.2020.100188>
- Carvalho, J., Francisco, R., & Relvas, A. P. (2015). Family functioning and information and communication technologies: How do they relate? A literature review. *Computers in Human Behavior*, 45, 99–108. <https://doi.org/10.1016/j.chb.2014.11.037>
- Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends*, 66(4), 616–630.
- Cha, D. L., Jiang, Z. H., & Cao, G. H. (2022). A study of user-perceived algorithmic anxiety and its structural dimensions in information systems. *Intelligent Science*, 6, 66–73. <https://doi.org/10.13833/j.issn.1007-7634.2022.06.009>
- Chen, S. H., & Tang, L. (2014). Human-machine love: Emotional interchain and emotional intelligence coupling of artificial intelligence partners. *Journal of Hainan University*, 9, 1–9. <https://doi.org/10.15886/j.cnki.hnus.202405.0437>
- Chew, R., Bollenbacher, J., Wenger, M., Speer, J., & Kim, A. (2023). LLM-Assisted Content Analysis: Using Large Language Models to Support Deductive Coding. *ArXiv*, <https://doi.org/10.48550/arXiv.2306.14924>
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education*. Routledge
- Crawford, K. (2021). Time to regulate AI that interprets human emotions." *Nature*, 592, 7853. <https://doi.org/10.1038/d41586-021-00868-5>
- Creswell, J., & Plano Clark, V. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Darioshi, R., & Lahav, E. (2021). The impact of technology on the human decision-making process. *Human Behavior and Emerging Technologies*. 3(2), 1-10. <http://dx.doi.org/10.1002/hbe2.257>
- De Togni, G., Erikainen, S., Chan, S., & Cunningham-Burley, S. (2021). What makes AI 'intelligent' and 'caring'? Exploring affect and relationality across three sites of intelligence and care. *Social Science and Medicine*, 277, 113874. <https://doi.org/10.1016/j.socscimed.2021.113874>
- Dehnert, M., & Mongeau, P. A. (2022). Persuasion in the age of artificial intelligence (AI): Theories and complications of AI-based persuasion. *Human Communication Research*, 48(3), 386–403. <https://doi.org/10.1093/hcr/hqac006>
- Denzin, N. (2017). *The research act: A theoretical introduction to sociological methods* (4th ed.). Routledge.
- Domínguez, M. (2014). Einstein versus neutrinos: The two cultures revisited with the media coverage of a scientific news item in cartoons. *Science Communication* 36(2), 248–25.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press

- Druga, S., Christoph, F. L., & Ko, A. J. (2022). Family as a third space for AI literacies: How do children and parents learn about AI together? In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1-17.
- Duan, Z., Li, J., Lukito, J., Yang, K., Chen, F., Shah, D., & Yang, S. (2022). Algorithmic agents in the hybrid media system: Social bots, selective amplification, and partisan news about COVID-19." *Human Communication Research*, 48(3), 516-542. <https://doi.org/10.1093/hcr/hqac012>
- Endacott, C., & Leonardi, P. (2022). Artificial intelligence and impression management: Consequences of autonomous conversational agents communicating on one's behalf. *Human Communication Research*, 48(3), 462-490. <https://doi.org/10.1093/hcr/hqac009>
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864-886. <https://doi.org/10.1037/0033-295X.114.4.864>
- Faverio, M., & Tyson, A. (2023). What the data says about Americans' view of artificial intelligence. *Pew Research Centre*. <https://www.pewresearch.org/short-reads/2023/11/21/what-the-data-says-about-americans-views-of-artificial-intelligence/>
- Feffer, M., Martelaro, N., & Heidari, H. (2023). The AI incident database as an educational tool to raise awareness of AI harms: A classroom exploration of efficiency, limitations & future improvements. *arXiv*. <https://arxiv.org/abs/2310.06269>
- Galaz, V., Centeno, M. A., Callahan, P. W., Causevic, A., Patterson, T., Brass, I., et al. (2021). Artificial intelligence, systemic risks, and sustainability. *Technology in Society*, 67, 101741. <https://doi.org/10.1016/j.techsoc.2021.101741>
- Gan, L. H., & Wang, H. (2024). From emotional projection to digital emotion: Emotional transformation of human-computer interaction in digital landscapes." *Modern Publishing*, 3, 27-38.
- Garg, R., Cui, H., Seligson, S., Zhang, B., Porcheron, M., Clark, L., et al. (2022). The last decade of HCI research on children and voice-based conversational agents. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1-19.
- Geertz, C. (1973). *The interpretation of cultures*. Basic Books.
- Glaser, B. & Strauss, A. (2017). *Discovery of Grounded Theory: Strategies for Qualitative Research*. Routledge.
- Gong, W. J., Wong, B. Y. M., Ho, S. Y., Lai, A. Y. K., Zhao, S. Z., Wang, M. P., et al. (2021). Family e-chat group use was associated with family well-being and personal happiness in Hong Kong adults amidst the COVID-19 pandemic. *International Journal of Environmental Research and Public Health*, 18, 9139. <https://doi.org/10.3390/ijerph18179139>
- Gossett, S. (2023). Emotion AI: 3 experts on the possibilities and risks. <https://builtin.com/artificial-intelligence/emotion-ai>

- Greene, G. (2020). The ethics of AI and emotional intelligence. <https://partnershiponai.org/paper/the-ethics-of-ai-and-emotional-intelligence/>
- Gremsl, T., & Hödl, E. (2022). Emotional AI: Legal and ethical challenges. *Information Policy*, 27, 163–174. <https://doi.org/10.3233/IP-211529>
- Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough? An Experiment with Data Saturation and Variability. *Field Methods*, 18(1), 59-82. <https://doi.org/10.1177/1525822X05279903>
- Gunkel, D. (2012). Communication and artificial intelligence: Opportunities and challenges for the 21st century." *Communication +1*, 1(1), 1-23. <https://doi.org/10.7275/R5QJ7F7R>
- Guzman, A., & Lewis, S. (2020). Artificial intelligence and communication: A human-machine communication research agenda. *New Media and Society*, 22(1), 70–86. <https://doi.org/10.1177/1461444819858691>
- Hancock, J., Naaman, M., & Levy, K. (2020). AI-mediated communication: Definition, research agenda, and ethical considerations." *Journal of Computer-Mediated Communication*, 25(1), 89–100. <https://doi.org/10.1093/jcmc/zmz022>
- Hata, A., Inam, R., Raizer, K., Wang, S., & Cao, E. (2019). AI-based safety analysis for collaborative mobile robots." In *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 1722–1729.
- Hayes, A., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, 1(1), 77-89.
- Hennig-Thurau, T., Aliman, D. N., Herting, A. M., et al. (2022). Social interactions in the metaverse: Framework, initial evidence, and research roadmap." *Journal of the Academy of Marketing Science*.
- Higgins, E. T. (2019). *Shared reality: What makes us stronger and tears us apart?* Oxford University Press.
- Hohenstein, J., Kizilcec, R., DiFranzo, D., Agharjari, Z., Mieczkowski, H., Levy, K., Naaman, M., Hnacock, J., & Jung, M. (2023). Artificial intelligence in communication impacts language and social relationships. *Scientific Reports*, 13, 5487. <https://doi.org/10.1038/s41598-023-30938-9>
- IBM. (2024). *Unlocking the power of chatbots; Key benefits for businesses and customers.* <https://www.ibm.com/think/insights/unlocking-the-power-of-chatbots-key-benefits-for-businesses-and-customers>
- Jager, J., Putnick, D., & Bornstein, M. (2017). More than Just Convenient: The Scientific Merits of Homogeneous Convenient Samples. *Monographs of the Society for Research in Child Development*, 82(2), 13-30.
- Kirk, H., Gabriel, I., Summerfield, C., Vidgen, B., Hale, S. (2025). Why human-AI relationships need socioaffective alignment. *Humanities and Social Sciences Communications*, 12, 728. <https://doi.org/10.1057/s41599-025-04532-5>

- Krippendorff, K. (2011). Computing Krippendorff's Alpha-Reliability. *Research at Penn Working Papers*, <https://repository.upenn.edu/handle/20.500.14332/2089>
- Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (4th ed.). SAGE Publications.
- Kvale, S., & Brinkmann, S. (2015). *Interviews: Learning the craft of qualitative research interviewing* (3rd ed.). SAGE Publications.
- Laapotti, T., & Raappana, M. (2022). Algorithms and organizing. *Human Communication Research*, 48(3), 491-515. <https://doi.org/10.1093/hcr/hqac013>
- Laestadius, L., Bishop, A., Gonzalez, M., Illenčik, D., & Campos-Castillo, C. (2022). Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media & Society*, 26(10), 5923-5941.
- Lee, D., & Yoon, S. N. (2021). Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. *International Journal of Environmental Research and Public Health*, 18, 271. <https://doi.org/10.3390/ijerph18010271>
- Lee, E. (2020). Authenticity model of computer-mediated communication: Conceptual explorations and testable propositions. *Journal of Computer-Mediated Communication*, 25(1), 60-73. <https://doi.org/10.1093/jcmc/zmz025>
- Liang, P., Bommasani, R., Lee, T., Tsipras, D., Soylu, D., et al. (2022). Holistic evaluation of language models. *arXiv*. <https://doi.org/10.48550/arXiv.2211.09110>
- Liao, Q. V., Subramonyam, H., Wang, J., & Wortman Vaughan, J. (2023). Designerly understanding: Information needs for model transparency to support design ideation for AI-powered user experience. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1-21.
- LivePerson. (2021). The digital lives of Millennials and Gen Z. *LivePerson*. <https://www.liveperson.com/blog/digital-lives-of-millennials-and-gen-z/#:~:text=Digital%20is%20the%20new%20IRL&text=In%20fact%2C%2065%25%20now%20communicate,and%20the%20UK%20at%2074.4%25>.
- MacQueen, K., McLellan-Lemal, E., Kay, K., & Milstein, B. (1998). Codebook Development for Team-Based Qualitative Analysis. *Field Methods*, 10(2), 31-36.
- Mahboudi, H., Farrokhi, F., & Ansarin, A. (2017). A Review on Application of Computers in Education Inside and Outside of Iran. *Advances in Language and Literacy Studies*, 8(4), 29-42.
- Mahmoudi, T., Ronaghi, M., & Amini, A. (2025). The Effect of Artificial Intelligence Adoption on Social Sustainability (Case Study: Isfahan Province Knowledge-Based Companies). *Journal of Entrepreneurship Development*, 17(4), 1-31.
- Malterud, K., Siersma, V., & Guassora, A. (2016). Sample Size in Qualitative Interview Studies: Guided by Information Power. *Qualitative Health Research*, 26(13), 1753-1760.

- Maples, B., Cerit, M., Vishwanath, A., & Pea, R. (2024). Loneliness and suicide mitigation for students using GPT3-enabled chatbots. *NPJ Mental Health Research*, 3(4), <https://doi.org/10.1038/s44184-023-00047-6>.
- Marcos-Pablos, S., & García-Peñalvo, F. J. (2022). *Emotional intelligence in robotics: A scoping review*. Cham: Springer.
- Marvin, M. (2006). *The emotion machine*. Hangzhou: Zhejiang People's Publishing.
- Matlabinejad, A., Fazeli, F. & Navaei, E. (2023). A systematic review of the promises and challenges of artificial intelligence for teachers." *Technology and Scholarship in Education*, 3(1), 23-44.
- McChesney, K., & Aldridge, J. (2019). Weaving an interpretivist stance through mixed methods research. *International Journal of Research and Method in Education*, 42(3), 225–238. McStay, A. (2018). *Emotional AI: The rise of empathic media*. London, Thousand Oaks, CA: Sage.
- McStay, A. (2020). Emotional AI and EdTech: Serving the public good? *Learning, Media and Technology*, 45, 270–283. <https://doi.org/10.1080/17439884.2020.1686016>
- Mijwil, M., Aggarwal, K., Mutar, D., Mansour, N., & Singh, R. (2022). The position of artificial intelligence in the future of education: An overview. *Asian Journal of Applied Sciences*, 10(2), 102–108.
- Morgan, D. (1997). *Focus groups as qualitative research* (2nd ed.). Sage Publications, Inc. <https://doi.org/10.4135/9781412984287>
- Morgante, E., Susinna, C., Culicetto, L., Quartarone, A., & Lo, B. V. (2024). Is it possible for people to develop a sense of empathy toward humanoid robots and establish meaningful relationships with them?" *Frontiers in Psychology*, 15, 1391832. <https://doi.org/10.3389/fpsyg.2024.1391832>
- Mostafa, G., Mahmoud, H., Abd El-Hafeez, T., et al. (2024). Feature reduction for hepatocellular carcinoma prediction using machine learning algorithms." *Journal of Big Data*. 11 (88). <https://doi.org/10.1186/s40537-024-00944-3>
- Nass, C., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. Cambridge: MIT Press.
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers." *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- O'Connor, C., & Joffe, H. (2020). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods*, 19. <https://doi.org/10.1177/1609406919899220>
- Palinkas, L., Horwitz, S., Green, C., Wisdom, J., Duan, N., & Hoagwood, K. (2013). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and policy in mental health*. 42. <http://dx.doi.org/10.1007/s10488-013-0528-y>.
- Patton M. (2002). *Qualitative Research and Evaluation Methods* (3rd ed.). Sage.

- Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4th ed.). SAGE Publications.
- Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of Replika. *Computers in Human Behavior*, 140, 107600.
- Qadikolaei, M., Zali, N., Soltani, A. (2022). Spatiotemporal investigation of the digital divide, the case study of Iranian Provinces. *Environment, Development and Sustainability*, 26, 869-884.
- Raaijmakers, S. (2019). Artificial intelligence for law enforcement: challenges and opportunities. *IEEE Security & Privacy*, 17(5), 74-77.
- Raes, M., Meijerink, I., Lykourantzou, I., & Khan, V. (2024). For Explainable to Interactive AI: A Literature Review on Current Trends in Human-AI Interaction. *ArXiv*. <https://arxiv.org/html/2405.15051v1#bib.bib1>
- Rahimi, B. (2015). Rethinking Digital Technologies in the Middle East. *International Journal of Middle East Studies*, 47(2), 362-365.
- Rajabi, M., & Nasrollahi, M. (2023). The cultural impact of artificial intelligence development on social media in Iran. *Journal of Iranian Cultural Research*, 16(2), 95-125. <https://doi.org/10.22035/jicr.2023.3178.3481>
- Roberts, C. (2020). *Text analysis for the social sciences: methods for drawing statistical inferences from texts and transcripts*. Routledge
- Russell, S. (2019). *Human Compatible AI and the Problem of Control*. London: Penguin
- Sabbar, S., & Habib Zadeh Khiyaban, S. (2023). Algorithms of displacement: Emotional and rhetorical responses to ai-driven job loss in digital public discourse. *International Journal of Advanced Multidisciplinary Research and Studies*, 3(4), 1324-1331.
- Saldaña, J. (2021). *The coding manual for qualitative researchers* (4th ed.). Sage.
- 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. <https://doi.org/10.22059/wsps.2025.401951.1552>
- Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & Quantity*, 52(4), 1893-1907.
- Sharma, G. (2017). Pros and cons of different sampling techniques. *International Journal of Applied Research*, 3(7), 749 - 752.
- Sharp, C. (2003). Qualitative Research and Evaluation Methods. (3rd ed.) *Evaluation Journal of Australasia*, 3(2), 60-61.
- Sundar, S. (2020). Rise of machine agency: A framework for studying the psychology of human-AI interaction (HAI). *Journal of Computer-Mediated Communication*, 25(1), 74-88. <https://doi.org/10.1093/jcmc/zmz026>

- Sundar, S., & Chen, J. (2023). From CASA to TIME: Machine as a source of media effects. In A. Guzman, R. McEwen, and S. Jones (Eds.), *The SAGE handbook of human-machine communication*. Sage Publications.
- Sundar, S., & Lee, E. (2022). Rethinking communication in the era of artificial intelligence. *Human Communication Research*, 48(3), 379–385. <https://doi.org/10.1093/hcr/hqac014>
- Tai, M. (2020). The impact of artificial intelligence on human society and bioethics." *Tzu Chi Medical Journal*, 32(4), 339-343. http://dx.doi.org/10.4103/tcmj.tcmj_71_20
- Taylor, G. (2005). *Integrating quantitative and qualitative methods in research*. University Press of America Inc.
- Teddle, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. SAGE Publications.
- Tracy, S. (2010). Qualitative Quality: Eight "Big-Tent" Criteria for Excellent Qualitative Research. *Qualitative Inquiry*, 16(10), 837-851.
- Tuckett, A. (2005). Applying thematic analysis theory to practice: A researcher's experience." *Contemporary Nurse*, 19(1), 75–87.
- Weber, A., Gupta, R., Abdalla, S., Cislighi, B., Meausoone, V., & Darmstadt, G. (2021). Gender-related data missingness, imbalance and bias in global health surveys. *BMJ Global Health*, 6, 007405. <https://doi.org/10.1136/bmjgh-2021-007405>
- Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., et al. (2022). Emergent abilities of large language models. *arXiv*. <https://doi.org/10.48550/arXiv.2206.07682>
- Wiles, R. (2013). *What are qualitative research ethics?* Bloomsbury Academic.
- Wu, J. (2024). Social and ethical impact of emotional AI advancement: the rise of pseudo-intimacy relationships and challenges in human interactions. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2024.1410462>
- Xiao, Z., Yuan, X., Liao, Q., Abdelghani, R., & Oudeyer, P. (2023). Supporting qualitative analysis with large language models: Combining codebook with GPT-3 for deductive coding. In *28th International Conference on Intelligent User Interfaces*, 75–78. ACM.
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.
- Zhang, M., Tang, E., Ding, H., & Zhang, Y. (2024). AI in communication sciences and disorders. *ASHA journals*. Dataset. <https://doi.org/10.23641/asha.27162564.v1>



Original-Forschungsarbeit

Die transformative Rolle der künstlichen Intelligenz in der Mediendatenanalyse für das Krisenmanagement

Hatef Pourrashidi Alibigloo^{1*}, Mehran Samadi²

¹ Fachbereich Kommunikationswissenschaft, Fakultät für Sozialwissenschaften, Universität der Religionen und Konfessionen, Qom, Iran

² Fachbereich Kommunikationswissenschaft, Islamische Azad-Universität, Ta.C.-Zweigstelle, Täbris, Iran

Empfangen: 20. Februar 2025 Akzeptiert: 2. Juni 2025

Zusammenfassung:

In der gegenwärtigen Landschaft des Krisenmanagements sehen sich Entscheidungsträger zunehmend mit der schieren Masse, Geschwindigkeit und Vielfalt an Mediendaten konfrontiert, die in Notsituationen generiert werden. Traditionelle manuelle Analysemethoden erweisen sich oft als unzureichend, um diesen Informationsfluss effektiv zu verarbeiten, was einen Paradigmenwechsel hin zu fortgeschrittenen computergestützten Ansätzen unumgänglich macht. Das primäre Ziel dieser Studie ist es, die Lücke zwischen technischer Datenwissenschaft und praktischer Krisenkommunikation zu schließen, indem eine klare analytische Verbindung zwischen spezifischen Paradigmen des maschinellen Lernens (ML) und ihren operativen Fähigkeiten hergestellt wird. Dieser Artikel bedient sich der Methodik eines narrativen Reviews, fundiert durch einen theoretischen Rahmen des maschinellen Lernens. Die Studie synthetisiert systematisch die bestehende Literatur, um zu kategorisieren und zu analysieren, wie unterschiedliche ML-Architekturen – insbesondere überwachtes, unüberwachtes und deep learning – im Bereich der Mediendatenanalyse zur Unterstützung von Entscheidungsprozessen während Krisen angewendet werden. Die Analyse bestätigt, dass Künstliche Intelligenz die Effektivität des Krisenmanagements durch die Automatisierung des Medienmonitorings und die Generierung handlungsrelevanter Echtzeiterkenntnisse signifikant steigert. Die Ergebnisse weisen verschiedenen Algorithmen spezifische Rollen zu: Überwachtes Lernen (*supervised learning*) dient als theoretisches Fundament für die schnelle Erkennung von Falschinformationen und die präzise Krisenklassifizierung. Demgegenüber werden unüberwachtes Lernen (*unsupervised learning*) und deep learning als kritische Werkzeuge zur Detektion von Datenanomalien und zur Erkennung aufkommender Muster identifiziert, die für die Funktionalität proaktiver Frühwarnsysteme essenziell sind. Obwohl KI ein transformatives Potenzial bietet, liefert diese Studie eine kritische Reflexion wesentlicher Implementierungsherausforderungen. Sie hebt das „Black-Box“-Problem – gekennzeichnet durch mangelnde algorithmische Interpretierbarkeit – sowie inhärente Datenverzerrungen (*Bias*) als zentrale ethische Hürden hervor, welche die Rechenschaftspflicht und Fairness bei der Krisenbewältigung beeinträchtigen können. Diese Arbeit bietet einen strukturierten Rahmen zum Verständnis der Rolle von KI aus einer theoretischen Perspektive und kommt zu dem Schluss, dass künftige Implementierungen „erklärbare KI“ (*Explainable AI*) priorisieren müssen, um eine Balance zwischen rechnerischer Effizienz und ethischer Verantwortung herzustellen.

Schlüsselwörter: künstliche Intelligenz, Mediendatenanalyse, Krisenmanagement, Krisenkommunikation, maschinelles Lernen

* Korrespondierender Autor

✉ h.pourrashidi@gmail.com

🌐 <https://orcid.org/0000-0003-2471-6778>

Wie dieser Artikel zu zitieren ist:

Pourrashidi Alibigloo, H., & Samadi, M. (2025). The transformative role of artificial intelligence in media data analysis for crisis management. *Spektrum Iran*, 38(2), 115-142.

🔗 <https://doi.org/10.22034/spektrum.2025.563353.1051>

© 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

مقاله پژوهشی

نقش متحول‌کننده هوش مصنوعی در تحلیل داده‌های رسانه‌ای برای مدیریت بحران

هاتف پوررشیدی علی‌بیگلو^{۱*}، مهران صمدی^۲

۱ گروه ارتباطات، دانشکده علوم اجتماعی، دانشگاه ادیان و مذاهب، قم، ایران

۲ گروه علوم ارتباطات، دانشگاه آزاد اسلامی، واحد تبریز، تبریز، ایران

دریافت: ۱۴۰۳/۱۲/۱؛ پذیرش: ۱۴۰۴/۰۳/۱۲

چکیده:

در چشم‌انداز کنونی مدیریت بحران، تصمیم‌گیرندگان به‌طور فزاینده‌ای با چالش حجم انبوه، سرعت بالا و تنوع گسترده داده‌های رسانه‌ای تولید شده در شرایط اضطراری مواجه هستند. روش‌های تحلیلی و دستی سنتی اغلب برای پردازش مؤثر این جریان اطلاعاتی ناکافی بوده و گذار پارادایمی به سوی رویکردهای محاسباتی پیشرفته را اجتناب‌ناپذیر می‌سازند. هدف اصلی این پژوهش، پر کردن شکاف میان علم داده‌های فنی و ارتباطات بحران کاربردی از طریق برقراری پیوندی تحلیلی و شفاف میان پارادایم‌های خاص «یادگیری ماشین» (ML) و قابلیت‌های عملیاتی آن‌ها است. این مقاله با بهره‌گیری از روش‌شناسی «مرور روایی» و با تکیه بر چارچوب نظری یادگیری ماشین تدوین شده است. در این مطالعه، ادبیات موجود به‌صورت نظام‌مند سنتز شده است تا نحوه کاربرد معماری‌های متمایز یادگیری ماشینی، به‌ویژه یادگیری نظارت‌شده، نظارت‌نشده و یادگیری عمیق، در حوزه تحلیل داده‌های رسانه‌ای جهت پشتیبانی از فرآیندهای تصمیم‌گیری در حین بحران‌ها، دسته‌بندی و تحلیل شود. نتایج تحلیل تأیید می‌کند که هوش مصنوعی از طریق خودکارسازی پایش رسانه‌ای و تولید بینش‌های لحظه‌ای و عملیاتی، اثربخشی مدیریت بحران را به‌طور قابل‌توجهی ارتقا می‌بخشد. یافته‌ها نقش‌های مشخصی را برای الگوریتم‌های مختلف ترسیم می‌کنند: «یادگیری نظارت‌شده» به‌عنوان مبنای نظری برای تشخیص سریع اطلاعات نادرست و طبقه‌بندی دقیق بحران عمل می‌کند. در مقابل، «یادگیری نظارت‌نشده» و «یادگیری عمیق» به‌عنوان ابزارهای حیاتی برای شناسایی ناهنجاری‌های داده‌ها و تشخیص الگوهای نوظهور شناسایی شده‌اند که برای کارکرد سیستم‌های هشدار سریع پیش‌دستانه ضروری هستند. اگرچه هوش مصنوعی پتانسیلی تحول‌آفرین ارائه می‌دهد، اما این پژوهش تأملی انتقادی بر چالش‌های مهم پیاده‌سازی آن دارد. این مطالعه مسئله «جعبه سیاه» که با فقدان تفسیرپذیری الگوریتمی مشخص می‌شود و «سوگیری‌های ذاتی داده‌ها» را به‌عنوان موانع اخلاقی عمده‌ای برجسته می‌سازد که می‌توانند پاسخگویی و انصاف را در واکنش به بحران مخدوش کنند. این پژوهش چارچوبی ساختار یافته برای درک نقش هوش مصنوعی از دریچه‌ای نظری ارائه می‌دهد و نتیجه می‌گیرد که پیاده‌سازی‌های آتی باید «هوش مصنوعی تفسیرپذیر» را در اولویت قرار دهند تا توازن میان کارایی محاسباتی و مسئولیت اخلاقی برقرار شود.

واژگان کلیدی: هوش مصنوعی، تحلیل داده‌های رسانه‌ای، مدیریت بحران، ارتباطات بحران، یادگیری ماشین

* نویسنده مسئول

<https://orcid.org/0000-0003-2471-6778>

h.pourrashidi@gmail.com

<https://doi.org/10.22034/spektrum.2025.563353.1051>



Original Research Paper

The transformative role of artificial intelligence in media data analysis for crisis management

Hatef Pourrashidi Alibigloo^{1*}, Mehran Samadi²

¹ Department of Communication, Faculty of Social Science, University of Religions and Denominations, Qom, Iran

² Department of Communication Science, Ta.C. Islamic Azad University, Tabriz, Iran

Received: Feb. 20, 2025 Accepted: Jun. 02, 2025

Abstract

In the contemporary landscape of crisis management, decision-makers are increasingly overwhelmed by the sheer volume, velocity, and variety of media data generated during emergencies. Traditional manual analytical methods are often insufficient to process this influx effectively, necessitating a paradigm shift toward advanced computational approaches. The primary goal of this study is to bridge the gap between technical data science and practical crisis communication by establishing a clear analytical link between specific machine learning (ML) paradigms and their operational capabilities. This article utilizes a narrative review methodology, underpinned by a theoretical framework grounded in machine learning. The study systematically synthesizes existing literature to categorize and analyze how distinct ML architectures – specifically supervised, unsupervised, and deep learning – are applied within the domain of media data analysis to support decision-making processes during crises. The analysis confirms that artificial intelligence significantly enhances crisis management effectiveness by automating media monitoring and generating actionable real-time insights. The findings delineate specific roles for different algorithms: supervised learning serves as the theoretical foundation for rapid misinformation detection and precise crisis classification. Conversely, unsupervised learning and deep learning are identified as critical tools for detecting data anomalies and recognizing emerging patterns, which are essential for the functionality of proactive early warning systems. While AI offers transformative potential, this study provides a critical reflection on significant implementation challenges. It highlights the “black box” problem – characterized by a lack of algorithmic interpretability – and inherent data biases as major ethical hurdles that can compromise accountability and fairness in crisis response. The present study contributes a structured framework for understanding AI’s role through a theoretical lens. It concludes that future implementation must prioritize explainable AI to balance computational efficiency with ethical responsibility.

Keywords: artificial intelligence, media data analysis, crisis management, crisis communication, machine learning

* Corresponding Author

✉ h.pourrashidi@gmail.com

🌐 <https://orcid.org/0000-0003-2471-6778>

How to Cite this Article:

Pourrashidi Alibigloo, H., & Samadi, M. (2025). The transformative role of artificial intelligence in media data analysis for crisis management. *Spektrum Iran*, 38(2), 115-142.

🔗 <https://doi.org/10.22034/spektrum.2025.563353.1051>

© 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

1. Introduction

The contemporary information landscape has fundamentally transformed crisis management dynamics. When emergencies occur – whether natural disasters, public health crises, organizational scandals, or security threats – digital media platforms generate unprecedented volumes of data within minutes. Social media posts, news articles, user-generated videos, and online discussions create an information deluge that simultaneously aids and complicates crisis response efforts (Perry et al., 2003, p. 207). Traditional manual monitoring and analysis methods, while effective in pre-digital eras, cannot process the velocity, volume, and variety of modern media data streams. This analytical gap poses critical risks: delayed crisis detection, inability to counter misinformation rapidly, and failure to gauge public sentiment accurately – all of which can escalate emergencies and erode institutional trust (Wodak, 2021, p. 330).

Artificial intelligence (AI), particularly its machine learning (ML) algorithms, has emerged as a transformative solution to these challenges. AI systems can monitor millions of social media posts in real-time, detect sentiment shifts across populations, identify emerging misinformation narratives, and predict crisis trajectories based on historical patterns (Cheng et al., 2025). These capabilities extend beyond human cognitive capacity, offering crisis managers sophisticated tools for situational awareness and strategic communication (Hossain et al., 2025, p. 156). However, AI's integration into crisis management is not without complications. Algorithmic opacity raises accountability concerns (Pedreschi et al., 2019, p. 9780), training data biases risk discriminatory outcomes (Reddy et al., 2024, p. 4928), and over-reliance on automated systems may undermine critical human judgment (Yuan et al., 2025, p. 2). As AI becomes increasingly embedded in crisis infrastructure, understanding both its capabilities and limitations becomes essential for responsible deployment.

Existing literature on AI in crisis management tends toward two extremes: technical studies focused on algorithm performance without contextual application and practitioner accounts that lack theoretical grounding (Barbierato & Gatti, 2024, p. 2). Few studies systematically examine how specific ML paradigms – supervised learning, unsupervised learning, deep learning – operationalize distinct crisis management functions, or critically

engage with the ethical tensions inherent in AI-driven crisis response (Du et al., 2025, p. 1). This gap is particularly problematic given the rapid adoption of AI tools by governments, corporations, and humanitarian organizations, often without adequate understanding of algorithmic limitations or ethical safeguards (Bălan & Nedelcu, 2024, p. 3).

This study addresses the gap by examining how AI-driven media data analysis transforms crisis management practices, with particular focus on the operational capabilities enabled by specific machine learning paradigms and the critical ethical limitations that must be addressed for responsible implementation. Through systematic documentary analysis of peer-reviewed literature published between 2014-2025, this research explores how different machine learning paradigms operationalize AI's role in media data analysis during crises, while critically examining the ethical, technical, and practical limitations that constrain AI deployment in high-stakes crisis contexts and identifying pathways for mitigation. By grounding the analysis in ML theory (Mohri et al., 2012, p. 55) while maintaining critical engagement with real-world crisis scenarios, this study bridges computational science and crisis communication scholarship, offering both theoretical advancement and practical guidance.

This research adopts a systematic documentary review methodology, synthesizing findings from 20 high-quality peer-reviewed journal articles sourced from Google Scholar and Scopus database. Unlike purely technical reviews, this study integrates perspectives from computer science, communication studies, organizational management, and applied ethics to provide a comprehensive understanding of AI's multifaceted role in crisis management (Sharma et al., 2021, p. 25). The review critically examines not only what AI can do, but what it should do – and under what conditions its use is justified, transparent, and equitable (Gasana, 2024, p. 30).

The paper proceeds as follows: the next section reviews existing literature on AI applications in crisis management, positioning this study within ongoing scholarly debates. The theoretical framework then explains how machine learning paradigms translate to crisis management operations. The methodology section details the systematic documentary approach employed for literature selection and analysis. The findings section discusses AI's capabilities, challenges, and ethical considerations in media data

analysis during crises. The conclusion synthesizes theoretical contributions, practical implications, and directions for future research. Throughout, the analysis maintains a focus on the dual imperatives of technological innovation and ethical responsibility, recognizing that effective crisis management requires not only powerful tools but also wise and accountable use of those tools (Parker, 2024, p. 328).

In this regard, research on crisis communication and data management within this process has garnered attention from both domestic and international scholars, with recent advancements in diverse AI tools sparking widespread interest in their application. A notable contribution is the book *Media and Crisis Communication* (2025), where Cheng et al. in Chapter 8 titled "Crisis Communication in the Age of Artificial Intelligence: Navigating Opportunities, Challenges, and Future Horizons," emphasize that integrating AI into crisis communication signals a paradigm shift in crisis management. This integration offers transformative capabilities for enhancing real-time insights, public engagement, and emergency response optimization, while AI plays a strategic role in information management – particularly countering misinformation and handling public relations – due to its high capacity for large-scale data analysis and rapid identification of false information.

Among domestic studies, Valivand Zamani and Mortazavi Zadeh (2024) examined "The Impact of artificial intelligence Use on Managers' Decision-Making Processes in Organizational Crisis Management." The researchers stress the necessity of adopting new tools and technologies in crisis management, advocating for AI training and development within organizations to enable managers to leverage it for timely, appropriate decisions that prevent crises and foster organizational progress.

Another study by Rahmani Maman and Dehghani Sanich (2024), titled "AI Innovations in Crisis Management: Exploring Applications and Methods," underscores how harnessing cutting-edge knowledge and advanced technologies, especially AI, can play a pivotal role in precise evaluation, comprehensive assessment, and innovative solutions for managing such events. AI excels at rapidly analyzing vast datasets to deliver accurate predictive analysis, thereby supporting more effective decision-making in crisis scenarios.

Bălan and Nedelcu (2024), in "artificial intelligence and Crisis Management," define AI as computers' ability to exhibit human-like reasoning and resolve situations typically requiring human intelligence. They highlight its use in forecasting crises like natural disasters through machine learning and robotics, such as monitoring volcanic ash or predicting water levels via specialized algorithms.

Mintoo et al. (2024), in "Transforming Global Crisis Communication through Digital Synergy: Enhancing Media Response Strategies with machine learning," systematically explore the disruptive role of digital synergy and machine learning (ML) in bolstering crisis communication strategies across domains like pandemics, natural disasters, cybersecurity incidents, and social media interactions. Findings reveal that digital synergy provides real-time, high-fidelity simulations of crisis dynamics, empowering decision-makers to anticipate challenges, allocate resources efficiently, and optimize emergency responses; meanwhile, ML techniques such as deep learning, predictive analytics, and natural language processing (NLP) facilitate misinformation detection, sentiment analysis, and forecasting public emotional responses.

2. Theoretical Framework

The theoretical foundation of this study rests on machine learning (ML) theory, which provides the necessary computational framework to manage and derive actionable insights from the immense volume and velocity of media data during a crisis (Mohri et al., 2012, p. 55). ML is not merely a tool; it encompasses distinct paradigms that directly inform the real-time monitoring and strategic response capabilities of AI systems. The primary analytical utility of ML lies in its ability to transform unstructured media content into actionable intelligence for crisis managers (Hossain et al., 2025, p. 156).

2.1. ML Paradigms for Crisis Media Analysis

The application of ML in crisis media analysis hinges on two key operational functions: immediate classification and discovery of novel patterns. Supervised Learning is paramount for rapid classification tasks, forming the theoretical basis for misinformation detection and crisis type identification. By training on historical labeled data – such as verified crisis reports versus

fake news—supervised algorithms enable AI to accurately and instantly categorize incoming media narratives, allowing organizations to triage high-priority information and debunk false claims (Komendantova & Erokhin, 2025, p. 85). This capability is fundamental to maintaining control over the public narrative during high-stakes events.

In contrast, unsupervised learning and deep learning address the complexity of emergent crises and unstructured data. Since crises are often novel events, algorithms are required to discover new patterns without prior labels. Unsupervised methods excel at anomaly detection—identifying unusual spikes in conversation or shifts in sentiment that signal an unrecognized threat (Pattanayak et al., 2024, p. 58). Deep learning models, utilizing architectures like Recurrent Neural Networks, enhance nuanced sentiment and emotion analysis from text and video (Liu & Kirubakaran, 2022, p. 8). These models provide depth to the understanding of public sentiment beyond simple positive/negative labels, linking theory to practical insights regarding public anxiety and potential unrest.

2.2. Critical Limitations: Accountability and Ethics

To provide the necessary critical reflection, the application of ML theory must acknowledge its inherent limitations regarding accountability and ethics (Yuan et al., 2025, p. 14). Data bias remains a significant theoretical and practical challenge. If training data reflects societal prejudices, the resulting ML model will propagate those biases, potentially leading to unfair resource allocation or disproportionate monitoring during a crisis (Reddy et al., 2024, p. 4930). The lack of model interpretability—the “black box” problem—compromises trust and justification (Pedreschi et al., 2019, p. 9780). Crisis managers must be able to explain why an AI system recommended a specific action. The opaqueness of many deep learning models undermines the theoretical requirement for transparency in high-impact decision-making. These limitations necessitate integrating ethical governance with the computational theory of ML to ensure that AI-driven crisis responses are not only effective but also equitable and accountable (Barbierato & Gatti, 2024, p. 2).

2.3. Crisis Communication

Crisis communication, a specialized public relations field, addresses the critical need to manage and mitigate reputational damage during crises—

events that threaten stakeholder safety, security, or welfare and impact operations and public image. It guides crisis managers to limit harm to stakeholders and organizations, emerging as a serious U.S. research concern in the 1980s with rapid growth in definitions and theorization thereafter (Coombs, 2014, p. 2). Effective crisis communication minimizes negative effects through timely, accurate, and consistent information, establishing best practices for pre-, during-, and post-crisis engagement (Spradley, 2017).

A comprehensive crisis communication plan serves as a roadmap, outlining identification, assessment, and response protocols, including crisis teams, communication channels, key messages, and stakeholder targeting to enable swift, coordinated responses. Organizations must adapt strategies against misinformation, prioritizing transparency, accuracy, and accountability, with leadership fostering trust and ethical practices (Gasana, 2024, p. 28). Transparency and honesty are paramount; concealing information erodes trust, while admitting errors rebuilds it, as media amplify crises, shaping public perception: "In crisis conditions, public reliance on media peaks, enabling crisis exacerbation or resolution, agenda-setting, hope instillation or despair, and opinion influence" (Nasrollahi Kasmani, 2018, p. 282).

Effective crisis communication tailors messages to diverse stakeholders – employees, customers, investors, media – via active listening and dialogue, defined as "a strategic approach to managing undesired events' impact on reputation, operations, and stakeholders" (Nuortimo, 2024). Social media complicates this, demanding vigilant monitoring and engagement to counter rapid misinformation spread, as crises induce fear amid uncertainty (Wodak, 2021, p. 332).

2.4. Crisis Management

Crisis management demands a proactive, adaptive approach to unforeseen disruptions, analyzing key components like preparation, response, and recovery to highlight the need for robust strategies. Crises are inherently unstable threats to strategic goals, reputation, or survival, distinct from incidents and requiring unique approaches amid social and personal disruptions reducing control and predictability (Hamidovic, 2012, p. 1; Jafari et al., 2021, p. 9).

Emergency preparedness has evolved into an all-hazards framework encompassing mitigation, preparation, response, and recovery (Herstein et al., 2021, p. 2), with preparation identifying vulnerabilities, risk assessments, and comprehensive plans specifying roles, protocols, training, and simulations. Essential steps include competent staffing, leadership modeling, rapid issue resolution, accountability, defined processes, continuous learning, open feedback cultures, and effective safety communication (Kapur et al., 2024, p. 11).

The response phase requires swift, decisive action with transparent internal and external communication to reduce panic, build trust, and control narratives, guided by pre-set protocols amid adaptations. As the most time-intensive stage, coordinated responses significantly limit impacts (Saghaei, 2023, p. 128). Recovery focuses on operational restoration, damage assessment, corrective actions, and lessons learned to rebuild stakeholder trust through ongoing commitment.

Effective crisis management is a continuous cycle of preparation, response, and improvement, rooted in strong planning, clear communication, and experiential learning, positioning resilient organizations to navigate disruptions with minimal damage (Parker, 2022, p. 328).

3. Materials and Methods

The present study adopts a systematic documentary review approach to critically synthesize existing literature on the transformative role of artificial intelligence in media data analysis for crisis management. Documentary reviews are particularly suited for emerging interdisciplinary topics, enabling comprehensive integration of theoretical insights, empirical findings, and practical implications across multiple scholarly domains (Greenhalgh, 2016, p. 97). This methodology allows researchers to explore complex phenomena like AI applications in dynamic crisis contexts, identifying patterns, gaps, and future research directions through interpretive thematic analysis.

3.1. Research Design

The review employed a systematic documentary methodology, drawing exclusively from peer-reviewed journal articles to ensure scholarly rigor and

credibility. Unlike purely technical surveys, this study integrates perspectives from computer science, communication studies, organizational management, and applied ethics to provide a multifaceted understanding of AI's role in crisis management (Sharma et al., 2021, p. 25). The systematic approach ensures transparency and replicability while maintaining the conceptual flexibility necessary for interdisciplinary synthesis.

3.2. Search Strategy and Data Sources

Literature searches were conducted using Google Scholar and Scopus database, selected for their comprehensive coverage of English-language scholarship in artificial intelligence, media studies, and crisis communication. The search employed Boolean operators combining keywords: (“artificial intelligence” or “machine learning” or “deep learning”) and (“media data analysis” or “media monitoring”) and (“crisis management” or “crisis communication” or “emergency response”). The temporal scope spanned 2014–2025, capturing AI's rapid evolution following deep learning breakthroughs while ensuring contemporary relevance. Initial searches yielded over 200 results across both databases.

3.3. Inclusion and Exclusion Criteria

Rigorous inclusion criteria were applied to ensure analytical quality. Articles were included if they: (1) were peer-reviewed journal publications in English; (2) explicitly addressed AI or ML applications in crisis or emergency contexts; (3) focused on media data analysis, monitoring, or communication strategies; and (4) provided empirical findings, theoretical frameworks, or systematic reviews. Exclusion criteria eliminated: conference proceedings without full peer review, non-English publications, purely technical papers lacking crisis context, and editorials or opinion pieces without empirical grounding. After applying these criteria and removing duplicates, 20 high-quality articles were selected for comprehensive analysis, as detailed in Table 1.

AI in media data analysis for crisis management

Table 1: Key Peer-Reviewed Articles Reviewed (N = 20, Google Scholar & Scopus)

	Authors (Year)	Journal	Key Focus	AI/Crisis Media Relevance
1	Cheng et al. (2025)	Media and Crisis Communication	AI in crisis comm.	Real-time misinformation countering
2	Eismann et al. (2021)	Journal of Strategic Information Systems	Social media in crises	Organizational learning via AI monitoring
3	Farrokhi et al. (2020)	Industrial Marketing Management	AI event detection	B2B crisis decision-making
4	Valivand Zamani & Mortazavi Zadeh (2024)	Motaleat-e Modiriyat-e Bohran	AI in org. crises	Managerial decision processes
5	Mintoo et al. (2024)	Journal of Next-Gen Engineering Systems	ML in crisis comm.	Digital synergy for media responses
6	Du et al. (2025)	Mathematics	ML theory	Generalization in high-stakes data
7	Barbierato & Gatti (2024)	Electronics	ML challenges	Crisis-applicable model limitations
8	BĂLAN & NEDELUCU (2024)	Social Economic Debates	AI crisis mgmt.	Forecasting via ML algorithms
9	Rahmani Maman & Dehghani Sanich (2024)	Conference Proceedings	AI crisis innovations	Predictive analytics
10	Gerlich et al. (2023)	Frontiers in Communication	AI influencer analysis	Public opinion in marketing crises
11	Mishra (2024)	Int. J. Scientific Research	AI/ML review	Media data ethics
12	Dabas (2023)	Integrated Journal for Research	AI media landscapes	Automated monitoring
13	Aleessawi & Alzubi (2024)	Studies in Media and Communication	AI media quality	Content analysis biases
14	Gao et al. (2023)	SAGE Open	AI advertising	Personalization in crises
15	Nuortimo et al. (2024)	Journal of Marketing Analytics	Reputation mgmt.	Crisis narrative control
16	Pedreschi et al. (2019)	AAAI Conference	Explainable AI	Black box issues in crises
17	Reynolds et al. (2025)	Trends in Ecology & Evolution	AI conservation	Anomaly detection analogy
18	Sharma et al. (2021)	Global Transitions Proceedings	ML applications	Deep learning in media
19	Gasana (2024)	Journal of Public Relations	Fake news mgmt.	AI transparency needs
20	Maleki Varnosfaderani & Forouzanfar (2024)	Bioengineering	AI healthcare	Real-time crisis parallels

4. Results & Discussion

4.1. The Operational Necessity of AI in Crisis Media Environments

The contemporary crisis landscape confronts organizations with unprecedented media data challenges that fundamentally exceed traditional analytical capabilities. During major crises, organizations face exponential information flows: millions of social media posts, news articles, citizen reports, and multimedia content generated within hours (Perry et al., 2003, p. 210). This volume, combined with velocity demands for real-time response, creates what Lundsgaard-Larsen and Gadegaard (2016, p. 15) term “the crisis data paradox” –the simultaneous necessity and impossibility of comprehensive media monitoring using conventional human-centered methods. The problem is not merely quantitative but qualitative: crisis media data exhibits high dimensionality, linguistic variability, contextual ambiguity, and intentional manipulation through coordinated disinformation campaigns (Komendantova & Erokhin, 2025, p. 83). These characteristics demand computational approaches capable of pattern recognition, semantic understanding, and adaptive learning—capabilities that define artificial intelligence systems grounded in machine learning paradigms.

The analytical framework established in the theoretical section—distinguishing supervised learning, unsupervised learning, and deep learning—directly translates to specific operational capabilities essential for crisis management. Each ML paradigm addresses distinct crisis communication challenges, and understanding these correspondences is critical for both theoretical advancement and practical implementation (Sharma et al., 2021, p. 26). The findings below systematically examine how these paradigms operationalize AI’s role in media data analysis during crises, while critically engaging with the ethical, technical, and practical limitations that constrain responsible deployment.

4.2. Supervised Learning: Real-Time Content Classification and Misinformation Detection

Supervised learning models, trained on labeled datasets to recognize predetermined patterns, provide crisis managers with immediate classification capabilities for incoming media content. This paradigm’s

primary operational value lies in its ability to rapidly categorize vast streams of social media posts, news articles, and citizen reports into actionable intelligence categories: urgent threats requiring immediate response, misinformation requiring correction, public sentiment indicators, and resource allocation signals (Farrokhi et al., 2020, p. 1356). During the COVID-19 pandemic, supervised learning algorithms demonstrated the capacity to process millions of social media posts daily, identifying critical patterns such as emerging outbreak clusters, public compliance with health measures, and coordinated disinformation campaigns—analyses that would require thousands of human analysts working continuously (Jafari et al., 2021, p. 12).

The most critical application of supervised learning in crisis contexts is misinformation and disinformation detection. False information during crises—whether accidental rumors or deliberate propaganda—can directly endanger lives by misdirecting evacuations, undermining public health measures, or inciting panic (Gasana, 2024, p. 32). AI systems trained on historical examples of verified versus false crisis claims can flag suspicious content with accuracy rates exceeding 85% in controlled studies, dramatically reducing verification time from hours to seconds (Komendantova & Erokhin, 2025, p. 88). For instance, during natural disasters, supervised models can distinguish legitimate eyewitness reports from recycled imagery, satire misinterpreted as fact, or intentional fabrications designed to manipulate relief efforts. This capability allows crisis communication teams to prioritize verification resources and issue timely corrections before false narratives achieve viral spread.

However, the effectiveness of supervised learning is fundamentally constrained by training data quality and representativeness. Models can only recognize patterns they have encountered during training, creating vulnerability to novel crisis types, emerging linguistic patterns, or culturally specific misinformation tactics (Barbierato & Gatti, 2024, p. 8). More critically, if training datasets contain systematic biases—for example, overrepresenting urban perspectives, certain demographic groups, or specific geographic regions—the resulting models will perpetuate these biases in operational deployment (Reddy et al., 2024, p. 4929). In crisis contexts where equitable resource allocation and fair representation of all affected populations are ethical imperatives, biased AI systems risk amplifying existing social

inequities rather than ameliorating them. This tension between operational efficiency and ethical responsibility represents a central challenge that supervised learning applications must address through rigorous debiasing protocols and continuous fairness auditing.

4.3. Unsupervised Learning: Anomaly Detection and Early Warning Systems

While supervised learning excels at recognizing known patterns, crises frequently manifest through unprecedented combinations of events that defy pre-labeled categories. Unsupervised learning addresses this limitation by identifying hidden structures, anomalies, and emerging patterns within unlabeled media data without requiring prior examples (Pattanayak et al., 2024, p. 58). This paradigm's operational value for crisis management lies in its capacity to serve as an early warning system, detecting signals of emerging crises before they escalate to full-scale emergencies requiring reactive response (Reynolds et al., 2025, p. 198).

Unsupervised anomaly detection algorithms continuously monitor baseline patterns of media activity – typical volumes of posts about specific topics, normal sentiment distributions, expected geographic distributions of content – and flag statistically significant deviations that may indicate emerging threats (Eismann et al., 2021, p. 4). For example, a sudden spike in social media posts containing health-related keywords from a specific geographic region, coupled with abnormal sentiment patterns expressing fear or confusion, might indicate an emerging disease outbreak, chemical exposure, or infrastructure failure before official reports reach authorities. Similarly, unexpected clustering of posts containing specific hashtags or coordinated messaging patterns can reveal organized disinformation campaigns in their initial stages, enabling preemptive countermeasures before false narratives achieve widespread acceptance (Komendantova & Erokhin, 2025, p. 91).

The strategic advantage of unsupervised learning lies in its “unknown unknown” detection capability – identifying threats that crisis planners did not anticipate and therefore could not prepare supervised models to recognize. This capability aligns directly with the contemporary shift toward all-hazards emergency preparedness frameworks that acknowledge the impossibility of predicting all potential crisis scenarios (Herstein et al., 2021, p. 3). However, unsupervised learning's strength is simultaneously its

operational weakness: anomaly detection generates high false-positive rates, flagging numerous statistical deviations that represent benign variations rather than genuine threats. Crisis managers must therefore balance sensitivity (detecting real emerging threats) against specificity (avoiding alert fatigue from false alarms), a trade-off that requires careful threshold calibration and integration with human expertise for context interpretation (Pattanayak et al., 2024, p. 61). Without this integration, unsupervised systems risk overwhelming response teams with noise rather than signal.

4.4. Deep Learning: Semantic Understanding and Emotion Analysis

Deep learning, particularly through neural network architectures, extends AI capabilities beyond pattern recognition to semantic understanding of media content's meaning, context, and emotional valence. This paradigm addresses a fundamental limitation of traditional ML approaches: the inability to comprehend linguistic nuance, contextual irony, cultural references, and emotional undertones that profoundly influence crisis communication effectiveness (Liu & Kirubakaran, 2022, p. 4). Deep learning models, especially those utilizing natural language processing (NLP) architectures like transformers, can analyze not merely what is being said but how it is being said and what emotions it conveys – capabilities essential for assessing public psychological states and tailoring communication strategies accordingly (Mintoo et al., 2024, p. 41).

Sentiment and emotion analysis powered by deep learning provides crisis managers with real-time assessments of public emotional responses to both the crisis itself and organizational communications. During major crises, tracking the emotional trajectory of public discourse – from initial shock and fear through anger, bargaining, and eventual acceptance or sustained trauma – enables organizations to adapt messaging strategies to meet populations where they are psychologically rather than where planners assume they should be (Liu & Kirubakaran, 2022, p. 12). For instance, if deep learning analysis reveals that public sentiment toward official health guidance is shifting from compliance to skepticism or anger, crisis communicators can adjust tone, provide additional transparency about decision-making processes, or engage trusted community voices to rebuild confidence before complete trust erosion occurs.

Deep learning also enables sophisticated content generation and personalization capabilities that can enhance crisis communication effectiveness. By analyzing individual users' historical interactions, concerns expressed, and information-seeking patterns, AI systems can tailor crisis messages to address specific population segments' priorities (Gao et al., 2023, p. 8). Elderly populations may require different messaging emphasis (focusing on accessibility of services, transportation assistance) compared to parents of young children (focusing on school closures, childcare resources) or individuals with disabilities (focusing on accessible evacuation routes, specialized medical support). This personalization, when implemented ethically with appropriate consent and privacy protections, can significantly improve message relevance and reduce the one-size-fits-all communication failures that have characterized many historical crisis responses (Maleki Varnosfaderani & Forouzanfar, 2024, p. 8).

However, deep learning's semantic sophistication comes at substantial cost. The neural network architectures required for NLP and emotion analysis are computationally intensive, requiring significant hardware resources and energy consumption that may be impractical during crises when infrastructure is compromised (Barbierato & Gatti, 2024, p.12). More fundamentally, deep learning models exhibit the most severe manifestation of the "black box" problem: their decision-making processes involve millions or billions of parameters interacting in ways that resist human interpretation even by the models' creators (Pedreschi et al., 2019, p. 9781). When a deep learning system classifies content as threatening or recommends prioritizing aid to specific populations, explaining why that decision was reached—a requirement for accountability in high-stakes crisis contexts—becomes extraordinarily difficult or extremely difficult with current architectures.

4.5. Integrated Crisis Communication: Strategic Narrative Control and Stakeholder Engagement

Beyond individual analytical capabilities, AI's transformative impact on crisis management emerges from integrated deployment across the complete crisis communication cycle. Modern crises unfold simultaneously across multiple media platforms—traditional news outlets, social media ecosystems, messaging applications, video platforms—each with distinct audience demographics, communication norms, and information diffusion

dynamics (Perry et al., 2003, p. 218). Managing coherent organizational narratives across this fragmented landscape manually is increasingly infeasible; AI provides the necessary coordination infrastructure.

Strategic narrative control during crises requires organizations to move beyond reactive “firefighting” toward proactive shaping of information environments. AI enables this shift through several mechanisms. First, by analyzing which information sources and influencers are shaping public understanding of the crisis, AI systems can guide communication teams to engage these key nodes directly with accurate information before misinformation takes root (Gerlich et al., 2023, p. 8). During the COVID-19 pandemic, public health agencies that successfully identified and collaborated with trusted community influencers—religious leaders, local celebrities, community organizers—achieved significantly higher vaccination rates than those relying exclusively on official government channels. AI’s network analysis capabilities make systematic identification of these influential voices feasible at scale. Recent computational discourse research further illustrates how large-scale social media analysis—integrating topic modeling, sentiment and emotion classification, and entity recognition—can illuminate the thematic structures, affective currents, and power configurations shaping digital public debate, demonstrating how algorithmic analysis reveals not only patterns of influence but also the broader sociopolitical dynamics embedded within contemporary information environments (Salehi et al., 2025).

Second, AI enables dynamic communication channel optimization. Different population segments consume information through different platforms and respond to different communication styles. Young adults may be reached most effectively through Instagram and TikTok with visual, concise messaging, while older populations may rely on Facebook and traditional news with more detailed, text-based information (Dabas, 2023, p. 252). AI systems can analyze which channels are reaching which demographics with what messages, allowing real-time strategy adjustments to maximize coverage across diverse stakeholder groups. This capability proves particularly valuable in international crises requiring coordination across linguistic and cultural boundaries.

Third, AI facilitates continuous feedback loops between organizational communication and public response. Rather than issuing statements and hoping for compliance, modern crisis management involves monitoring how messages are received, identifying misunderstandings or resistance points, and iteratively refining communication strategies based on observed patterns (Nuortimo et al., 2024, p. 12). AI's real-time processing enables this adaptive approach at the speed crises demand. For example, if AI analysis reveals that a particular population segment is misinterpreting a technical term in emergency instructions, communicators can immediately issue clarifications using accessible language before dangerous misunderstandings lead to injuries.

However, this integration also concentrates substantial power in organizational hands, raising ethical questions about manipulation and transparency. The same capabilities that enable effective crisis communication could be weaponized for propaganda, narrative suppression, or selective information disclosure. When organizations use AI to identify and engage influencers, are they facilitating information dissemination or orchestrating astroturfing? When they personalize messages based on psychological profiling, are they increasing relevance or engaging in manipulative microtargeting? These questions lack clear answers but demand explicit ethical frameworks governing AI use in crisis contexts (Yuan et al., 2025, p. 9).

4.6. Critical Limitations: Interpretability, Bias, and Implementation Challenges

While AI offers substantial operational capabilities for crisis media analysis, acknowledging and critically examining its fundamental limitations is essential for responsible implementation. These limitations are not merely technical challenges awaiting engineering solutions but represent inherent tensions between AI's operational logic and the ethical, social, and political requirements of just crisis management.

The interpretability challenge—commonly termed the “black box problem”—poses fundamental obstacles to accountability in high-stakes crisis decision-making. As ML models, particularly deep neural networks, increase in complexity and accuracy, their internal decision-making processes become increasingly opaque (Pedreschi et al., 2019, p. 9780). When an AI system recommends prioritizing aid distribution to specific geographic

areas, classifying social media content as threatening, or identifying certain populations as high-risk, human operators cannot readily understand which features, patterns, or correlations drove those recommendations. In non-crisis contexts, this opacity may be acceptable if overall performance is satisfactory. In crisis contexts, where every decision can affect human lives, the inability to explain why a system made a specific recommendation undermines accountability, prevents meaningful human oversight, and erodes public trust (Reynolds et al., 2025, p. 201). Parallel concerns have been documented in other high-stakes AI domains, where critical reviews emphasize that systems deployed in socially sensitive contexts require robust explainability, transparency, and human oversight to prevent the normalization of opaque decision-making and to safeguard ethical accountability (Salehi et al., 2026).

This problem extends beyond technical explanation to epistemic concerns about AI's reasoning. Even when researchers employ explainability techniques like SHAP values or attention mechanisms to identify which input features most influenced a model's output, these technical explanations often fail to align with human causal reasoning or ethical principles. An AI system might accurately predict crisis escalation based on correlations that are statistically valid but ethically inappropriate to act upon—for example, demographic characteristics protected by anti-discrimination law. The system cannot distinguish between predictive accuracy and ethical permissibility, placing full responsibility for this distinction on human operators who may lack complete understanding of what patterns the AI is exploiting (Valivand Zamani & Mortazavi Zadeh, 2024, p. 22).

Data bias represents an equally serious concern with direct implications for social justice in crisis response. AI models learn from historical data, inheriting and often amplifying whatever biases that data contains. In crisis media analysis, multiple bias sources converge: media coverage itself exhibits systematic biases toward urban areas, wealthier populations, and dominant cultural groups; social media participation is demographically skewed with significant digital divides excluding marginalized populations; and training dataset construction involves subjective human judgments about what constitutes relevant, threatening, or important content (Reddy et al., 2024, p. 4930). When AI systems trained on such data are deployed in crisis contexts,

they risk perpetuating historical patterns of unequal attention, resource allocation, and voice amplification.

For example, if training data overrepresents media coverage of crises affecting wealthy urban areas while underrepresenting rural or economically disadvantaged regions, AI systems may be less sensitive to early warning signals emerging from these underrepresented areas. Similarly, if sentiment analysis models are trained predominantly on standard language varieties, they may misinterpret or fail to process dialectal variations, code-switching, or multilingual content common in diverse communities—effectively silencing these populations’ voices in crisis monitoring systems (Aleessawi & Alzubi, 2024, p. 57). The operational consequence is not merely technical inaccuracy but systematic injustice: the populations most vulnerable during crises become invisible to the AI systems designed to protect them.

Addressing bias requires more than technical debiasing algorithms; it demands critical examination of what crisis management priorities are being embedded in AI systems and whose interests these systems serve. Current research on algorithmic fairness offers various mathematical definitions of fairness—demographic parity, equalized odds, individual fairness—but these definitions can conflict, and selecting among them involves normative judgments about social values that technical experts alone cannot make (Reddy et al., 2024, p. 4931). Crisis management organizations deploying AI must therefore engage in participatory processes involving affected communities, ethicists, social scientists, and domain experts to establish fairness criteria appropriate to specific contexts, and must commit to ongoing auditing and adjustment as new biases emerge.

Implementation challenges extend beyond algorithmic limitations to organizational, infrastructural, and sociotechnical factors. AI systems require substantial computational resources, high-quality data streams, technical expertise for deployment and maintenance, and organizational cultures that can effectively integrate algorithmic insights with human judgment (Barbierato & Gatti, 2024, p. 15). Many organizations responsible for crisis management—particularly in resource-constrained settings, developing nations, or rural areas—lack these prerequisites. Even when technical capabilities exist, organizational resistance to AI-driven recommendations, misplaced trust in algorithmic authority, or lack of training in appropriate AI

use can undermine effectiveness or create new vulnerabilities (Valivand Zamani & Mortazavi Zadeh, 2024, p. 20).

The COVID-19 pandemic illustrated these implementation challenges clearly. While sophisticated AI systems for epidemiological modeling, misinformation detection, and resource optimization were developed rapidly, their practical impact varied enormously across contexts. Organizations with strong technical capacity, interdisciplinary collaboration, and established crisis management protocols could integrate AI effectively; those lacking these foundations found AI tools unhelpful or actively counterproductive when deployed without adequate support infrastructure (Wodak, 2021, p. 338). This disparity risks creating a two-tiered crisis management landscape where technologically advanced organizations can leverage AI's capabilities while others fall further behind, exacerbating global inequalities in crisis preparedness and response capacity.

4.7. Toward Responsible AI Integration: Requirements for Ethical Crisis Management

The findings above demonstrate that AI's role in crisis media analysis is neither purely beneficial nor inherently problematic but fundamentally ambivalent—offering transformative capabilities while introducing serious risks. Responsible integration requires deliberate design choices, robust governance frameworks, and ongoing critical engagement with AI's limitations rather than uncritical enthusiasm for its possibilities. Recent scholarship in disaster risk management similarly characterizes AI as a dual-use technology, simultaneously embedded within exploitative data economies and capable of enhancing early warning systems and post-disaster resilience, underscoring the necessity of governance models grounded in transparency, data sovereignty, and explainability (Sharifi Poor Bgheshmi et al., 2026).

First, explainable AI (XAI) development must become a priority for crisis management applications. Current XAI research focuses primarily on technical interpretability metrics that may not align with crisis managers' decision-making needs or ethical accountability requirements (Pedreschi et al., 2019, p. 9783). What crisis contexts demand is human-centered explainability: explanations that help human operators understand not just which features influenced a model's output but whether the model's

reasoning aligns with domain knowledge, ethical principles, and organizational values. This requires interdisciplinary collaboration between ML researchers, crisis management practitioners, ethicists, and affected communities to define what constitutes adequate explanation in specific crisis contexts.

Second, bias mitigation must extend beyond technical debiasing algorithms to systemic approaches addressing root causes of data inequality. This includes deliberately oversampling underrepresented populations in training data, developing culturally sensitive annotation protocols that respect linguistic and contextual diversity, establishing participatory processes for affected communities to influence system design and evaluation criteria, and implementing continuous fairness auditing with consequences for systems that perpetuate discriminatory patterns (Reddy et al., 2024, p. 4930). Organizations must also acknowledge that perfect fairness may be unattainable and prepare contingency plans for addressing algorithmic failures when they occur.

Third, human-AI collaboration frameworks must recognize that effective crisis management requires combining AI's computational strengths with human capacities for contextual judgment, ethical reasoning, and adaptive improvisation. AI should augment rather than replace human decision-makers, with clear protocols specifying when algorithmic recommendations should be followed, questioned, or overridden (Hossain et al., 2025, p. 159). This requires organizational cultures that value critical questioning of technology, provide training in AI literacy and limitations, and protect individuals who raise concerns about algorithmic recommendations from institutional pressure to defer to automated systems.

Fourth, governance frameworks must establish accountability structures for AI-driven crisis decisions. When AI systems contribute to decisions that affect human lives—resource allocation, evacuation orders, communication prioritization—clear chains of responsibility must exist. This includes transparency about what AI systems are being used, documentation of their training data and known limitations, processes for appealing or contesting algorithmic recommendations, and mechanisms for redress when AI systems cause harm (Yuan et al., 2025, p. 12). Such accountability cannot exist without

transparency, requiring organizations to resist proprietary secrecy about AI systems used in crisis contexts.

Finally, capacity building must ensure that AI's benefits extend beyond technologically advanced organizations to the full range of actors involved in crisis management globally. This requires open-source tool development, training programs accessible to resource-constrained organizations, technical assistance for AI implementation in diverse contexts, and critical examination of how AI deployment may exacerbate existing global inequalities (Komendantova & Erokhin, 2025, p. 98). The goal is not universal adoption of AI but rather equitable access to AI's capabilities for those contexts where it can genuinely improve crisis outcomes.

5. Conclusion

This study utilized a narrative review methodology, underpinned by machine learning (ML) theory, to explore the transformative role of AI in media data analysis during crisis management. The necessity of this framework stems from the unprecedented volume and velocity of media data that render traditional analytical methods obsolete. The findings strongly suggest that AI's operational effectiveness is not merely technical but is rooted in distinct ML paradigms: supervised learning facilitates the immediate classification and detection of misinformation and disinformation, drastically reducing verification time, while unsupervised learning and deep learning enable the identification of emerging patterns, hidden trends, and anomalies essential for building proactive early warning systems. Our theoretical contribution lies in establishing a clear analytical link between these specific ML mechanisms and their resultant operational capabilities in the crisis domain, thereby moving beyond a generic description of AI's functional assistance.

However, the research also highlights critical challenges that prevent the uncritical adoption of these technologies. The lack of interpretability in complex ML models fundamentally challenges the need for accountability and public trust in high-stakes crisis decision-making, where every intervention requires robust justification. Furthermore, data bias remains a significant ethical and operational concern, risking the propagation of

existing inequities in media coverage and potentially leading to the misallocation of aid resources or misrepresentation of vulnerable populations. Methodologically, as a narrative review, this study's findings are theoretical and conceptual, and must be validated through rigorous empirical testing before practical implementation can be fully assured.

To advance the field, future research should focus on developing Explainable AI (XAI) frameworks specifically tailored for crisis media analysis to effectively address the interpretability challenge and enhance human-AI collaboration during emergencies. Further empirical studies are needed to quantitatively test the impact of debiasing techniques on ML models operating under the dynamic and volatile conditions of real-world crises. This work is essential to ensure that future AI applications in crisis management are not only effective in terms of speed and accuracy but also transparent, ethical, and equitable in their overall societal impact.

Conflict of Interest

The authors declare no potential conflicts of interest regarding the publication of this work. In addition, ethical issues including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancy have been fully addressed by the authors.

Author contributions

The authors confirm the study conception and design: Hatef Pourrashidi Alibigloo; data collection: Hatef Pourrashidi Alibigloo; analysis and interpretation of results: Hatef Pourrashidi Alibigloo and Mehran Samadi; draft manuscript preparation: Hatef Pourrashidi Alibigloo. The results were evaluated by all authors, and the final version of the manuscript was approved.

AI Disclosure Statement

Artificial intelligence tools were employed in the initial stages of this manuscript's preparation. Specifically, AI-assisted technologies were used for:

- **Initial Conceptual Framing:** AI tools assisted in organizing preliminary ideas and structuring the initial outline of the manuscript, helping to establish the conceptual framework for exploring AI's role in crisis management.
- **Translation and Language Enhancement:** AI-powered translation tools were utilized to translate the initial draft from Persian to English, followed by language refinement to ensure academic tone, clarity, and grammatical accuracy appropriate for scholarly publication.
- **Literature Organization:** AI tools helped in preliminary organization and categorization of research sources during the early literature review phase.

However, all substantive intellectual content—including the research design, methodology, theoretical framework, critical analysis, interpretation of findings, and conclusions—represents the original scholarly work of the authors. All cited sources were independently verified by the authors, and the final manuscript underwent comprehensive human review and revision to ensure accuracy, academic integrity, and alignment with the study's research objectives. The authors retain full responsibility for all content, arguments, and any errors in this manuscript.

References

- Abed Al Sailawi, A. S., & Kangavari, M. R. (2024). Analyzing the use of social media data to understand long-term crisis management challenges of COVID-19. *Fusion: Practice and Applications (FPA)*, 14(02), 227-243. <https://doi.org/10.54216/FPA.140219>
- Aleessawi, N. A. K., & Alzubi, S. F. (2024). The implications of artificial intelligence (AI) on the quality of media content. *Studies in Media and Communication*, 12(4), 41-51.
- Bălan, C., & Nedelcu, A.-M. (2024). Artificial intelligence and crisis management. *Social Economic Debates*, 13(2), 1-8.
- Barbierato, E., & Gatti, A. (2024). The challenges of machine learning: A critical review. *Electronics*, 13, 416. <https://doi.org/10.3390/electronics13020416>
- Cheng, Y., Lee, J., & Qiao, J. (2025). Crisis communication in the age of AI: Navigating opportunities, challenges, and future horizons. In *Media and Crisis Communication*. Routledge.
- Coombs, W. T. (2014). State of crisis communication: Evidence and the bleeding edge. *Research Journal of the Institute for Public Relations*, 1(1), 1-12.
- Dabas, P. (2023). The role of artificial intelligence in shaping future media landscapes. *Integrated Journal for Research in Arts and Humanities*, 3(5), 328-334. <https://doi.org/10.55544/ijrah.3.5.36>
- Du, K.-L., Zhang, R., Jiang, B., Zeng, J., & Lu, J. (2025). Understanding machine learning principles: Learning, inference, generalization, and computational learning theory. *Mathematics*, 13, 451. <https://doi.org/10.3390/math13030451>
- Eismann, K., Posegga, O., & Fischbach, K. (2021). Opening organizational learning in crisis management: On the affordances of social media. *Journal of Strategic Information Systems*, 30(4), 101692. <https://doi.org/10.1016/j.jsis.2021.101692>
- Farrokhi, A., Shirazi, F., Hajli, N., & Tajvidi, M. (2020). Using artificial intelligence to detect crisis related to events: Decision making in B2B by artificial intelligence. *Industrial Marketing Management*, 91, 257-273. <https://doi.org/10.1016/j.indmarman.2020.09.015>
- Gasana, K. (2024). Crisis communication and reputation management in the age of fake news. *Journal of Public Relations*, 3(1), 28-39.
- Gerlich, M., Elsayed, W., & Sokolovskiy, K. (2023). Artificial intelligence as toolset for analysis of public opinion and social interaction in marketing: Identification of micro and nano influencers. *Frontiers in Communication*, 8, 1075654. <https://doi.org/10.3389/fcomm.2023.1075654>
- Greenhalgh, T. (2016). *Cultural contexts of health: The use of narrative research in the health sector* (Health Evidence Network Synthesis Report No. 49). WHO Regional Office for Europe.
- Hamidovic, H. (2012). *An introduction to crisis management*. ISACA Journal, 5, 1-4.

- Herstein, J. J., et al. (2021). Emergency preparedness: What is the future? *Antimicrobial Stewardship & Healthcare Epidemiology*, 1(1), e29, 1-6. <https://doi.org/10.1017/ash.2021.190>
- Hossain, M. Z., Akter, N., Hasan, L., Bepari, M., & Sultana, S. (2025). Artificial intelligence and machine learning in crisis communication: A management information system perspective. *European Journal of Innovative Studies and Sustainability*, 1(3), 149-163. [https://doi.org/10.59324/ejiss.2025.1\(3\).10](https://doi.org/10.59324/ejiss.2025.1(3).10)
- Jafari, F., Nasrollahi Kasmani, A., Farokhi, A., & Delavar, A. (2021). Media valuation in health news coverage: A case study of the COVID-19 pandemic. *Modiriyat-e Behdasht va Daroeman*, 12(3), 7-21. [In Persian]
- Kapur, S. P., et al. (2024). *The challenges of nuclear security: U.S. and Indian perspectives*. Palgrave Macmillan. <https://doi.org/10.1007/978-3-031-56814-5>
- Komendantova, N., & Erokhin, D. (2025). Artificial intelligence tools in misinformation management during natural disasters. *Public Organization Review*, 25(1), 81-105. <https://doi.org/10.1007/s11115-025-00815-2>
- Liu, C., & Kirubakaran, S. (2022). Deep learning approach for emotion recognition analysis in text streams. *International Journal of Technology and Human Interaction*, 18(2), 1-21. <https://doi.org/10.4018/IJTHI.313927>
- Lundsgaard-Larsen, A., & Gadegaard, M. (2016). *Big data analytics in crisis communication & management: A theoretical evaluation*. Copenhagen Business School.
- Maleki Varnosfaderani, S., & Forouzanfar, M. (2024). The role of AI in hospitals and clinics: Transforming healthcare in the 21st century. *Bioengineering*, 11, 337. <https://doi.org/10.3390/bioengineering11040337>
- Mintoo, A. A., Saimon, A. S. M., & Begum, A. (2024). Transforming global crisis communication through digital twins: Enhancing media response strategies with machine learning. *Innovatech Engineering Journal*, 1(01), 35-52.
- Mishra, A. (2024). A comprehensive review of artificial intelligence and machine learning: Concepts, trends, and applications. *International Journal of Scientific Research in Science and Technology*, 11(5), 126-142. <https://doi.org/10.32628/IJSRST2411587>
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of machine learning*. The MIT Press.
- Nasrollahi Kasmani, A. (2018). Media management of dust hazards: With emphasis on Khuzestan dust storms. *Modiriyat-e Mokhat*, 5(3), 279-294. <https://doi.org/10.22059/jhsci.2018.269355.419> [In Persian]
- Nuortimo, K., Harkonen, J., & Breznik, K. (2024). Exploring corporate reputation and crisis communication. *Journal of Marketing Analytics*. <https://doi.org/10.1057/s41270-024-00353-8>
- Parker, L. D. (2024). Third sector crisis management and resilience: Reflections and directions. *Financial Accountability and Management*, 40(3), 326-343. <https://doi.org/10.1111/faam.12379>

- Pattanayak, S. K., Bhojar, M., & Adimulam, T. (2024). Unsupervised learning for anomaly detection in cybersecurity. *International Journal of Research, Culture, Society*, 8(12), 56-63. <https://doi.org/10.2017/IJRCS/202412011>
- Pedreschi, D., Giannotti, F., Guidotti, R., Monreale, A., Ruggieri, S., & Turini, F. (2019). Meaningful explanations of black box AI decision systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01), 9780-9784. <https://doi.org/10.1609/aaai.v33i01.33019780>
- Perry, D. C., Taylor, M., & Doerfel, M. L. (2003). Internet-based communication in crisis management. *Management Communication Quarterly*, 17(2), 206-232. <https://doi.org/10.1177/0893318903256227>
- Rahmani Maman, A. H., & Dehghani Sanich, M. S. (2024). Innovations of artificial intelligence in crisis management: Examining applications and methods. In *Proceedings of the 3rd International Conference on Architecture, Civil Engineering, Earth Sciences, and Healthy Environment*. Hamadan, Iran. [In Persian]
- Reddy, V. C., Kapoor, S. K., & Mishra, K. S. (2024). Bias and fairness in machine learning models: A critical examination of ethical implications. *International Journal of Multidisciplinary Research in Science, Engineering and Technology*, 7(2), 4927-4931. <https://doi.org/10.15680/IJMRSET.2024.0702033>
- Reynolds, S. A., et al. (2025). The potential for AI to revolutionize conservation: A horizon scan. *Trends in Ecology & Evolution*, 40(2), 191-207. <https://doi.org/10.1016/j.tree.2024.11.013>
- Saghaei, E., Erfani, V., & Khodaveysi, S. (2023). Presenting a framework for provincial crisis response programs in Hamadan Province. *Sci J Rescue Relief*, 15(2), 127-139. <https://doi.org/10.32592/jorar.2023.15.2.6>
- 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.
- Salehi, K., Habib Zadeh Khiyaban, S., & Sabbar, S. (2026). Artificial Intelligence and Crime Detection: A Critical Review. *Journal of Cyberspace Studies*, 181-197.
- Sharifi Poor Bgheshmi, M. S. , Sharajsharifi, M. and Saeidabadi, M. R. (2026). Between exploitation and resilience: Reconciling AI's role in surveillance capitalism and disaster risk management. *Journal of Cyberspace Studies*, 10(1), 109-139. doi: 10.22059/jcss.2025.396045.1165
- Sharma, N., Sharma, R., & Jindal, N. (2021). Machine learning and deep learning applications: A vision. *Global Transitions Proceedings*, 2, 24-28. <https://doi.org/10.1016/j.gltp.2021.01.004>
- Spradley, R. T. (2017). Crisis communication in organizations. In *The International Encyclopedia of Organizational Communication*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118955567.wbieoc050>

- Valivand Zamani, H., & Mortazavi Zadeh, A. R. (2024). Investigating the impact of artificial intelligence use on managers' decision-making process in organizational crisis management. *Motaleāt-e Modiriyat-e Bohrān*, 16(2), 11-26. [In Persian]
- Wodak, R. (2021). Crisis communication and crisis management during COVID-19. *Global Discourse*, 11(3), 329-353. <https://doi.org/10.1332/204378921X16100431230102>
- Yuan, X., Guo, Q., Dadson, Y. A., Goodarzi, M., Jung, J., Dong, Y., Albert, N., Bennett Gayle, D., Sharma, P., Ogunbayo, O. T., & Cherukuru, J. (2025). A review of ethical challenges in AI for emergency management. *Knowledge*, 5(3), 21, 1-18. <https://doi.org/10.3390/knowledge5030021>



Original-Forschungsarbeit

Gefühlsasymmetrien in der KI: Sentiment-bias zwischen Englisch und Persisch in harmonisierten LLM-Pipelines

Michael W. Totaro¹, Leila Gheisi², Ehsan Shahghasemi^{3*}

¹ Professor, Fachbereich Informatik, Universität Louisiana at Lafayette, USA

² Doktorandin in Kommunikation, University of Louisiana at Lafayette, USA

³ Außerordentlicher Professor, Department of Communication, Universität Teheran, Iran

Empfangen: 5. März 2025 Akzeptiert: 8. Juni 2025

Zusammenfassung:

Diese Studie untersucht, wie Sprache die Sentiment-Klassifikation in Ausgaben eines multilingualen großen Sprachmodells (LLM) namens Grok beeinflusst. Basierend auf Langdon Winners Theorie der technologischen Politik, die besagt, dass Technologien inhärent nicht neutral sind und strukturelle Verzerrungen einbetten, wird geprüft, ob Sentiment-Verteilungen auch bei einer vollständig harmonisierten Analysepipeline systematisch zwischen Sprachen variieren. Die Analyse basiert auf einem Korpus von 4.799 Beiträgen (Englisch: n = 2.399; Persisch: n = 2.400), die mit identischen Aufforderungen erzeugt wurden. Sentiment-Ausgaben wurden auf ein gemeinsames dreistufiges Schema (Negativ, Neutral, Positiv) abgebildet, wobei sowohl diskrete Klassenzuweisungen als auch kontinuierliche Wahrscheinlichkeitswerte berücksichtigt wurden. Strukturelle Merkmale wie Satz-, Wort- und Zeichenanzahl wurden berechnet und als Kontrollvariablen einbezogen, um oberflächliche textuelle Unterschiede zu berücksichtigen. Die Ergebnisse zeigen eine deutliche sprachübergreifende Divergenz in Sentiment-Mustern. Englische Ausgaben konzentrieren sich überwiegend auf Neutralität und weisen eine vergleichsweise geringere affektive Intensität auf, während persische Ausgaben eine starke Verschiebung hin zu positivem Sentiment und größere Streuung zeigen. Diese Unterschiede bleiben auch nach Kontrolle struktureller Merkmale statistisch signifikant, was nahelegt, dass die Sprachzugehörigkeit und nicht Textlänge oder Segmentierung der Hauptfaktor für die beobachteten Sentiment-Unterschiede ist. Auf Wahrscheinlichkeitsniveau zeigen englische Verteilungen eine engere Konzentration nahe Neutralität, während persische Verteilungen flacher und stärker positiv verzerrt sind, mit höheren Intensitätswerten. Diese Ergebnisse haben wichtige Implikationen für mehrsprachige Sentiment-Analysen und LLM-Audits. Ohne explizite Modellierung und Kalibrierung von Spracheffekten könnten vergleichende Analysen sprachliche Variation mit affektiver Absicht verwechseln, was zu verzerrten Schlussfolgerungen über Ton, Haltung oder emotionale Valenz führt. Die Studie betont die Bedeutung der Berichterstattung sowohl von Label- als auch Wahrscheinlichkeitsmetriken, die Anwendung sprachspezifischer Kalibrierungsprotokolle und die Berücksichtigung von Sprache als primäre Messdimension in der sprachübergreifenden Inhaltsanalyse.

Schlüsselwörter: sentiment-analyse, mehrsprachige NLP, sprachbias, große sprachmodelle, sprachübergreifender vergleich

* Korrespondierender Autor

✉ shahghasemi@ut.ac.ir

🌐 <https://orcid.org/0000-0002-8716-5806>

Wie dieser Artikel zu zitieren ist:

Totaro, M.W., Gheisi, L., & Shahghasemi, E. (2025). Affective asymmetries in AI: Sentiment bias between English and Persian in harmonized LLM pipelines. *Spektrum Iran*, 38(2), 143-157.

🔗 <https://doi.org/10.22034/spektrum.2026.563602.1052>

© 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

مقاله پژوهشی

عدم تقارن‌های عاطفی در هوش مصنوعی: سوگیری احساسی بین زبان‌های انگلیسی و فارسی در جریان‌های یکپارچه LLM

مایکل و. توتارو^۱، لیلیا غیثی^۲، احسان شاه‌قاسمی^{۳*}

^۱ استاد، دپارتمان علوم کامپیوتر و انفورماتیک، دانشگاه لوئیزیانا در لافایت، ایالات متحده

^۲ دانشجوی دکتری ارتباطات، دانشگاه لوئیزیانا در لافایت، ایالات متحده

^۳ دانشیار، دانشکده ارتباطات، دانشگاه تهران، تهران، ایران

دریافت: ۱۴۰۳/۱۲/۱۵ پذیرش: ۱۴۰۴/۳/۱۸

چکیده:

این مطالعه بررسی می‌کند که چگونه زبان بر دسته‌بندی احساسی در خروجی‌های تولیدشده توسط یک مدل زبانی بزرگ چندزبانه (LLM) به نام Grok تأثیر می‌گذارد. با تکیه بر نظریه سیاست‌های تکنولوژیک لانگدون وینر، که معتقد است فناوری‌ها به‌طور ذاتی خنثی نیستند و سوگیری‌های ساختاری را در خود جای می‌دهند، این پژوهش بررسی می‌کند که آیا توزیع‌های احساسی حتی در یک جریان تحلیلی کاملاً یکپارچه به‌طور سیستماتیک بین زبان‌ها متفاوت است یا خیر. تحلیل بر اساس یک مجموعه داده شامل ۴۰۷۹۹ پست (انگلیسی $n = 2,399$ ؛ فارسی $n = 2,400$) تولیدشده با استفاده از درخواست‌های یکسان انجام شد. خروجی‌های احساسی به یک طرح سه‌رده‌ای مشترک (منفی، خنثی، مثبت) نگاشت شدند و تحلیل‌ها شامل هر دو معیار انتساب دسته و امتیازهای احتمال پیوسته بود. برای کنترل تفاوت‌های سطحی متنی، ویژگی‌های ساختاری شامل تعداد جملات، کلمات و کاراکترها محاسبه و به‌عنوان متغیر کنترل وارد تحلیل شد. نتایج نشان‌دهنده یک تفاوت بین‌زبانی قوی در الگوهای احساسی است. خروجی‌های انگلیسی عمدتاً در رده خنثی متمرکز هستند و شدت عاطفی نسبتاً کمتری دارند، در حالی که خروجی‌های فارسی گرایش قوی به احساس مثبت همراه با پراکندگی بیشتر نشان می‌دهند. مهم این‌که، این تفاوت‌ها حتی پس از کنترل ویژگی‌های ساختاری نیز از نظر آماری معنادار باقی می‌مانند و نشان می‌دهند که وابستگی به زبان، نه طول متن یا بخش‌بندی آن، مهم‌ترین عامل مرتبط با تغییرات احساسی مشاهده‌شده است. در سطح احتمال، توزیع‌های انگلیسی تمرکز بیشتری نزدیک به خنثی دارند، در حالی که توزیع‌های فارسی مسطح‌تر و به سمت مثبت متمایل هستند و شاخص‌های شدت بالاتری دارند. این یافته‌ها پیامدهای مهمی برای تحلیل احساسی چندزبانه و بررسی مدل‌های LLM دارند. در صورتی که اثرات زبانی به‌صراحت مدل‌سازی و کالیبره نشوند، تحلیل‌های مقایسه‌ای ممکن است تفاوت‌های زبانی را با قصد عاطفی اشتباه بگیرند و منجر به برداشت‌های تحریف‌شده درباره لحن، موضع یا ارزش عاطفی شوند. این مطالعه اهمیت گزارش هر دو معیار برچسب و احتمال، اتخاذ پروتکل‌های کالیبراسیون مخصوص زبان و در نظر گرفتن زبان به‌عنوان یک بعد اندازه‌گیری درجه‌اول در تحلیل محتوا میان‌زبانی را تأکید می‌کند.

واژگان کلیدی: تحلیل احساسات، پردازش زبان طبیعی چندزبانه، سوگیری زبانی، مدل‌های بزرگ زبان، مقایسه بین‌زبانی

* نویسنده مسئول

<https://orcid.org/0000-0002-8716-5806>

shahghasemi@ut.ac.ir

<https://doi.org/10.22034/spektrum.2026.563602.1052>



Original Research Paper

Affective asymmetries in AI: Sentiment bias between English and Persian in harmonized LLM pipelines

Michael W. Totaro¹, Leila Gheisi², Ehsan Shahghasemi^{3*}

¹ Department of Computer Science & Informatics, University of Louisiana at Lafayette, USA

² PhD student in Communication, University of Louisiana at Lafayette, USA

³ Associate Professor, Department of Communication, The University of Tehran, Tehran, Iran

Received: Mar. 05, 2025 Accepted: Jun. 08, 2025

Abstract

In the contemporary landscape of crisis management, decision-makers are increasingly overwhelmed by the sheer volume, velocity, and variety of media data generated during emergencies. Traditional manual analytical methods are often insufficient to process this influx effectively, necessitating a paradigm shift toward advanced computational approaches. The primary goal of this study is to bridge the gap between technical data science and practical crisis communication by establishing a clear analytical link between specific machine learning (ML) paradigms and their operational capabilities. This article utilizes a narrative review methodology, underpinned by a theoretical framework grounded in machine learning. The study systematically synthesizes existing literature to categorize and analyze how distinct ML architectures—specifically supervised, unsupervised, and deep learning—are applied within the domain of media data analysis to support decision-making processes during crises. The analysis confirms that artificial intelligence significantly enhances crisis management effectiveness by automating media monitoring and generating actionable real-time insights. The findings delineate specific roles for different algorithms: supervised learning serves as the theoretical foundation for rapid misinformation detection and precise crisis classification. Conversely, unsupervised learning and deep learning are identified as critical tools for detecting data anomalies and recognizing emerging patterns, which are essential for the functionality of proactive early warning systems. While AI offers transformative potential, this study provides a critical reflection on significant implementation challenges. It highlights the “black box” problem—characterized by a lack of algorithmic interpretability—and inherent data biases as major ethical hurdles that can compromise accountability and fairness in crisis response. The present study contributes a structured framework for understanding AI’s role through a theoretical lens. It concludes that future implementation must prioritize explainable AI to balance computational efficiency with ethical responsibility.

Keywords: artificial intelligence, media data analysis, crisis management, crisis communication, machine learning

* Corresponding Author

✉ shahghasemi@ut.ac.ir

🌐 <https://orcid.org/0000-0002-8716-5806>

How to Cite this Article:

Totaro, M.W., Gheisi, L., & Shahghasemi, E. (2025). Affective asymmetries in AI: Sentiment bias between English and Persian in harmonized LLM pipelines. *Spektrum Iran*, 38(2), 143-157.

🔗 <https://doi.org/10.22034/spektrum.2026.563602.1052>

© 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

1. Introduction

Bias in communication has been a central concern across centuries of scholarly inquiry, and the historical debate around this topic reflects persistent anxieties about the accuracy, fairness, and intent behind how messages are transmitted and received. From classical rhetoric to contemporary media studies, scholars have long examined the role of communicative agents, languages, mediums, and institutions in shaping the content and reception of messages. In early rhetorical traditions, figures like Aristotle explored the persuasive functions of ethos, pathos, and logos, which implicitly acknowledge the subjective framing that can influence how information is presented and interpreted. The rise of print culture, the evolution of mass media, and now the proliferation of algorithmically mediated forms of communication have only deepened interest in understanding how structural, cultural, and technological conditions contribute to communication biases. Across disciplines – from sociology and political science to media studies and linguistics – bias in communication is studied as both a deviation from neutrality and as a mechanism through which power and ideology are encoded and reproduced (Entman, 2007; van Dijk, 1993).

Bias is not merely a failure of impartiality; it is often embedded in the architectures of communication itself. As early as the mid-20th century, communication scholars began interrogating how different media channels structure the conditions for discourse, privileging certain temporal or spatial biases depending on their material and institutional configurations. A particularly influential contribution came from Canadian theorist Harold Innis, whose concept of “bias” in communication technologies laid the groundwork for later media ecology frameworks. Innis (1951) proposed that communication media exhibit either a bias toward time or space. Time-biased media, such as stone tablets or oral traditions, facilitate the preservation of culture across generations but are limited in their spatial reach. Space-biased media, such as paper or digital text, enable communication across large distances but tend to be ephemeral and less durable over time. Innis’s framework emphasizes that media are not neutral conduits; they condition what kinds of knowledge are preserved, circulated, and prioritized, thus shaping the trajectory of civilizations. His work highlighted how the very format of a medium predisposes it to certain political and cultural effects.

It should be noted that, For Harold Innis, the concept of "bias" in communication differed significantly from the later, more politicized understandings of bias found in Marxist or critical theory traditions. Innis's use of the term was less concerned with ideological favoritism or systematic discrimination against particular social groups, and more focused on the structural and temporal properties of communication media. His notion of bias referred to the way certain media favor the transmission of knowledge across either time or space, shaping the stability and expansion of civilizations accordingly.

Extending Harold Innis's foundational contributions, subsequent scholars elaborated the analysis of communicative bias across more specific empirical and theoretical domains. Marshall McLuhan, a student of Innis, famously asserted that "the medium is the message," foregrounding the claim that the affordances and constraints of communication technologies generate effects that exceed the semantic content they convey (McLuhan, 1964). Later theorists such as Pierre Bourdieu (1991) shifted attention to the operation of symbolic power through language itself, arguing that linguistic capital and habitus structure not only what can be said, but also who is authorized to speak and how utterances are received. In a complementary vein, Norman Fairclough (1995) developed critical discourse analysis to expose how ideological formations are reproduced through texts that appear ostensibly neutral, while Teun van Dijk (1993) examined the manifestation of systemic racism in media representations and political discourse. Taken together, these lines of inquiry converge on a common premise: communication is never free from bias, and such bias is frequently structurally embedded, culturally sustained, and technologically mediated.

Within this intellectual lineage, Langdon Winner (1980) introduced a critical refinement by explicitly theorizing the politics of artifacts, with particular attention to communication technologies. Winner contended that technologies are not neutral instruments but material instantiations of power relations and social organization. By posing the provocative question "Do artifacts have politics?" he argued affirmatively that certain technologies are intrinsically political, either because they presuppose or reinforce specific social arrangements or because their effects systematically advantage particular groups. Applied to communication, Winner's framework suggests that technologies do more than transmit messages: they actively configure

the conditions of expression, access, and interpretation. This perspective reorients analysis away from content and toward infrastructure, shifting emphasis from discourse to design. Such an approach is especially salient in digital media environments, where algorithms, interfaces, and data architectures increasingly govern what information becomes visible, authoritative, or amplified.

Winner's theory of technological politics dovetails with contemporary concerns about algorithmic bias and automated decision-making. In the era of large language models (LLMs), the processes by which information is classified, ranked, and labeled are often opaque, raising questions about how technological infrastructures encode preferences, assumptions, or systemic inequalities. Scholars such as Noble (2018) have shown how search engines reproduce racial and gendered stereotypes, while Eubanks (2018) documented how algorithmic systems used in public services often penalize marginalized populations. These critiques underscore Winner's insight that technologies, particularly those used in communication, are not apolitical—they are designed and deployed within specific power structures that shape their effects.

Artificial intelligence, particularly in the domain of language modeling and sentiment classification, introduces subtle yet consequential forms of bias into communication (Sabbar & Habib Zadeh Khiyaban, 2023). As evidenced in the present study, large language models (LLMs) like Grok exhibit systematic affective divergence across languages even when analytic pipelines are fully harmonized. These divergences—manifested as higher positivity and greater sentiment intensity in Persian compared to English—persist after controlling for structural variables such as word or sentence count, indicating that the bias is not merely a function of input length or segmentation, but of deeper linguistic, script, or model calibration effects. Such asymmetries highlight how AI systems, through their training data, tokenization mechanisms, and classifier thresholds, do not neutrally interpret or replicate linguistic inputs; instead, they encode and reproduce culturally contingent affective norms that shape interpretive outcomes. This phenomenon aligns with Langdon Winner's (1980) theory of technological politics, which posits that artifacts—including computational models—embody social and political values and can enforce specific regimes of interpretation depending on their embedded assumptions and affordances.

The risk, therefore, is not only technical misclassification but also epistemological distortion: AI-driven communication systems may systematically reshape how tone, sentiment, and stance are understood across linguistic groups (Salehi et al., 2026). For instance, as Noble (2018) and Eubanks (2018) have argued in adjacent contexts, algorithmic systems often reflect and exacerbate structural inequalities by encoding dominant cultural perspectives into seemingly objective technologies. In multilingual sentiment analysis, this dynamic manifests when affective intensity is over- or under-represented based on language-specific priors, potentially skewing comparative analyses in media studies, public opinion research, or international communication scholarship (Hanna et al., 2025; Pessach & Shmueli, 2022; Tejani et al., 2024; Shahghasemi, 2025). Without robust calibration protocols that treat language as a first-order measurement dimension, LLMs risk privileging affective neutrality in one language (e.g., English) while amplifying positivity in another (e.g., Persian), not due to actual differences in tone but due to model artifacts.

The study of sentiment analysis provides a concrete example of how communicative technologies can introduce or exacerbate bias (Venkit & Wilson, 2021; Thelwall, 2018; Díaz et al., 2018; Bhanvadia et al., 2024; Venugopal et al., 2024; Rozado, 2020; Shahghasemi et al., 2025; Radaideh et al., 2025). Automated sentiment classifiers, especially those powered by LLMs, are trained on corpora that reflect existing linguistic, cultural, and ideological biases. As such, these models may encode not only lexical or syntactic patterns but also implicit judgments about tone, emotion, and stance. When applied across languages, such models face the added challenge of cross-linguistic variation in pragmatics, grammar, and script. For instance, a model trained on English may interpret hedging, politeness, or affirmation differently than one trained on Persian, leading to divergent sentiment classifications even when the underlying messages are equivalent in intent or tone. This raises profound implications for comparative media research, public opinion analysis, and digital governance.

Indeed, cross-lingual disparities in sentiment analysis may not simply reflect differences in expression but rather artifacts of the technological pipeline itself. This aligns with Winner's contention that technological systems can enforce or conceal forms of bias depending on how they are structured and operationalized. In multilingual applications of LLMs,

differences in tokenization, script (e.g., Latin vs. Perso-Arabic), and classifier calibration can produce systematic differences in affective outputs, even when inputs are semantically aligned. Thus, communication technologies—particularly those powered by machine learning—do not merely mediate human expression; they participate in shaping it, often in subtle and consequential ways.

Building on Winner's theory, the present study follows this tradition of inquiry by empirically investigating how language-specific characteristics in large language model outputs may introduce bias into sentiment analysis. Specifically, the research examines the outputs of a multilingual LLM (Grok) when prompted to generate content in English (EN) and Persian (FA). By harmonizing the sentiment classification scheme across both languages—reducing native sentiment labels to a common three-class system (Negative/Neutral/Positive)—the study isolates language membership as a potential driver of divergent sentiment outcomes. The design controls for structural features of the text (such as sentence and word counts) to disentangle linguistic from purely formal differences.

This work is guided by a series of research questions that probe the relationship between language and sentiment classification in a harmonized analytic pipeline:

RQ1. Do structural features of Grok's outputs (sentences, words, characters per post) differ between English and Persian?

RQ2. Holding the label space constant (Negative/Neutral/Positive), does the distribution of sentiment classes differ by language?

RQ3. Are the mean sentiment probabilities $P(\text{Negative})$, $P(\text{Neutral})$, $P(\text{Positive})$ systematically different across languages?

RQ4. Does language (EN vs. FA) predict sentiment outcomes under a harmonized pipeline?

RQ5. Do confidence characteristics (e.g., the allocation of probability mass toward neutrality vs. positivity) vary by language in ways that could bias interpretation?

RQ6. Are observed differences substantively meaningful in terms of effect sizes, beyond mere statistical significance, for cross-language comparisons of tone?

RQ7. After accounting for structural text differences, do language-based

differences in sentiment persist, indicating a genuine cross-lingual shift rather than an artifact of message length or segmentation?

2. Methodology

We constructed a time-bounded corpus of Grok outputs in two languages, English (EN) and Persian (FA), generated with identical prompts within a single day. After removing empty and duplicate entries, we retained balanced samples of roughly 2,400 posts per language. Preprocessing served two purposes. First, to characterize text structure, we computed sentence, word, and character counts directly from the raw text so that punctuation and segmentation cues were preserved; only light normalization (e.g., link removal, whitespace cleanup) was applied to avoid altering length or punctuation patterns. Second, to characterize sentiment, both corpora were aligned to a common three-class space (Negative/Neutral/Positive). English posts were scored with a standard three-class transformer; Persian posts were scored with a widely used Persian model and then collapsed from five native categories to the same three classes (Negative = Furious+Angry; Neutral = Neutral; Positive = Happy+Delighted). For both languages, we retained per-class probabilities and the hard label (argmax), and we derived an intensity index defined as $1 - P(\text{Neutral})$.

To ensure cross-lingual comparability, we held the label space, decision rules, and output format constant across EN and FA. All subsequent analyses apply the same procedures to both samples; this enabled a clean assessment of whether language membership (EN vs. FA) is associated with systematic differences in sentiment under a harmonized pipeline.

3. Findings

Using the harmonized three-class scheme (Negative/Neutral/Positive), we summarized sentiment at both the hard-label and probability levels, and tracked a continuous intensity index. The class composition differs markedly by language. In the English corpus, posts are predominantly Neutral, with 1,637 items ($\approx 68.2\%$) assigned to that class; Negative and Positive account for 558 ($\approx 23.3\%$) and 204 ($\approx 8.5\%$) posts, respectively. In the Persian corpus, the distribution shifts toward the Positive pole: 1,006 posts ($\approx 41.9\%$) are labeled Positive, 917 ($\approx 38.2\%$) Neutral, and 477 ($\approx 19.9\%$) Negative.

Table 1. Cross-lingual disparities in sentiment analysis.

Class	EN Count	EN %	FA Count	FA %
Negative	558	23.3	477	19.9
Neutral	1,637	68.2	917	38.2
Positive	204	8.5	1,006	41.9

The analytic corpus comprises two balanced language sets, English (EN; $n = 2,399$) and Persian (FA; $n = 2,400$), processed under a harmonized three-class sentiment scheme (Negative/Neutral/Positive). Descriptive statistics were computed for two families of variables: structural features of the text (number of sentences, words, and characters per post) and sentiment outputs at both the label and probability levels, including a continuous intensity index. All structural measures were derived from the raw message body to preserve punctuation and segmentation cues, and sentiment probabilities are those emitted by the language-specific classifiers after mapping both languages to the common three-class space.

Table 2. Descriptive statistics for structural text features in the English (EN) and Persian (FA) corpora

Metric	EN n	EN M	EN SD	FA n	FA M	FA SD
Sentences per post	2,399	3.04	0.89	2,400	3.49	1.00
Words per post	2,399	42.68	5.62	2,400	50.77	9.21
Characters per post	2,399	268.53	25.46	2,400	259.08	40.98

For structural features, English posts are shorter in sentences and words but slightly longer in characters. English contains on average $M = 3.04$ sentences per post ($SD = 0.89$), whereas Persian averages $M = 3.49$ ($SD = 1.00$); medians in both languages are 3, and the empirical distributions are discrete with narrow interquartile ranges, especially in English. Word counts follow the same pattern: English averages $M = 42.68$ words ($SD = 5.62$), compared with $M = 50.77$ words ($SD = 9.21$) in Persian, with a visibly broader spread in

the Persian sample. Character counts invert this relationship: English averages $M = 268.53$ characters ($SD = 25.46$) versus $M = 259.08$ ($SD = 40.98$) in Persian, with identical medians (276) but heavier tails in Persian. These distributional shapes are consistent with script and tokenization differences: the Persian sample partitions content into more orthographic tokens and sentences without increasing character length proportionally, while English concentrates more tightly around a stable character count.

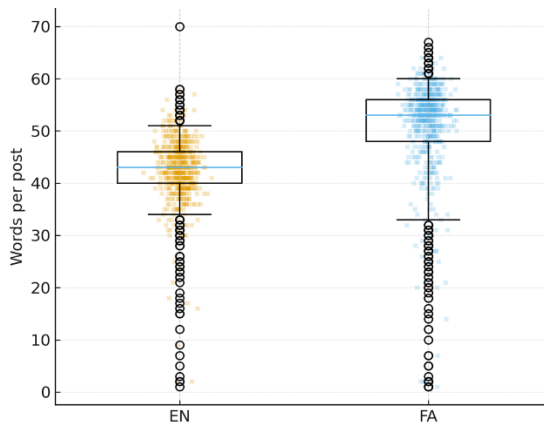


Figure 1. Distribution of words per post in the English (EN) and Persian (FA) corpora, shown as boxplots with overlaid jittered observations

At the label level, the composition of sentiment classes differs markedly between the two languages. In English, Neutral dominates the distribution (1,637 of 2,399; $\approx 68.2\%$), with smaller shares for Negative (558; $\approx 23.3\%$) and Positive (204; $\approx 8.5\%$). In Persian, the mass shifts toward Positive (1,006 of 2,400; $\approx 41.9\%$), with Neutral reduced (917; $\approx 38.2\%$) and Negative somewhat lower (477; $\approx 19.9\%$). These shares are reflected in the underlying probability profiles. The mean neutrality probability is substantially higher in English than in Persian ($M = .613$, $SD = .259$ vs. $M = .384$, $SD = .394$), indicating that English outputs concentrate probability near the neutral class and do so with comparatively low dispersion. The positive dimension shows the converse: Persian assigns more probability to positivity and exhibits greater spread ($M = .414$, $SD = .409$) relative to English ($M = .138$, $SD = .208$). On the negative dimension, English carries a modestly higher mean than Persian ($M = .249$,

$SD = .281$ vs. $M = .201$, $SD = .325$), while the Persian distribution again shows wider dispersion. Visual diagnostics—including overlaid histograms, boxplots, and violin plots—reinforce the same pattern: English curves are steeper around neutrality, whereas Persian curves are flatter and longer-tailed toward positivity, with more mass away from the neutral center.

The intensity index condenses these probability patterns into a single continuous measure of affective strength. English posts exhibit lower and less variable intensity ($M = .387$, $SD = .259$), while Persian posts are both higher on average and more dispersed ($M = .616$, $SD = .394$). Empirical cumulative distribution functions for intensity and for the maximum class probability (a proxy for classification confidence) display corresponding differences, with English curves rising more steeply (concentration near moderate intensity and higher neutrality) and Persian curves rising more gradually (greater heterogeneity, with more high-intensity, positive-leaning posts).

Table 3. Sentiment probability outputs and the continuous sentiment intensity index in the English (EN) and Persian (FA) corpora.

Metric	EN n	EN M	EN SD	FA n	FA M	FA SD
P(Negative)	2,399	0.249	0.281	2,400	0.201	0.325
P(Neutral)	2,399	0.613	0.259	2,400	0.384	0.394
P(Positive)	2,399	0.138	0.208	2,400	0.414	0.409
Sentiment intensity	2,399	0.387	0.259	2,400	0.616	0.394

Table 4. Between-language differences in structural and sentiment-related measures, reporting mean differences ($\Delta M = EN - FA$), 95% confidence intervals, permutation-test p -values, and effect sizes (Cliff's δ)

Metric	ΔM (EN-FA)	95% CI ΔM	p (perm, mean)	Cliff's δ
Sentences per post	-0.4500	[-0.5000, -0.3900]	< .001	0.000
Words per post	-8.1000	[-8.5400, -7.6600]	< .001	0.007
Characters per post	+9.4600	[+7.5400, +11.4300]	< .001	0.045

Metric	ΔM (EN-FA)	95% CI ΔM	p (perm, mean)	Cliff's δ
P(Negative)	+0.0477	[+0.0306, +0.0650]	< .001	0.302
P(Neutral)	+0.2287	[+0.2098, +0.2475]	< .001	0.332
P(Positive)	-0.2764	[-0.2972, -0.2547]	< .001	-0.299
Sentiment intensity	-0.2287	[-0.2475, -0.2098]	< .001	-0.332

4. Conclusion

Evidence from descriptive profiles and harmonized outputs points to a clear pattern: language systematically shapes Grok's expressed affect. English responses concentrate probability mass around neutrality and, to a lesser extent, negativity; Persian responses shift mass toward positivity and exhibit higher intensity (i.e., lower neutrality probabilities). These differences persist when sentiment is summarized as both hard labels and probability distributions, indicating that the phenomenon is not an artifact of a particular decision rule. In practical terms, this means that direct, cross-language comparisons of tone can be misleading unless the measurement pipeline explicitly accounts for language effects.

Regarding RQ1 (structural variation), English-language posts are marginally shorter in terms of sentences and words but slightly longer when measured by character count, whereas Persian-language posts exhibit greater length and variability at the token level. Importantly, these structural differences are relatively small and do not correspond to the observed direction of sentiment divergence. Persian texts, despite containing more words and sentences, are associated with higher levels of positivity and affective intensity, while English texts—characterized by more compact character distributions—tend toward greater neutrality. This decoupling between structural properties and affective outcomes indicates that superficial measures of text length are unlikely to account for the cross-linguistic sentiment gap.

Addressing RQ2-RQ4 (differences in class distributions and probabilities; language as a predictor), the distributions diverge in straightforward ways: English outputs cluster in Neutral, Persian in Positive. Because the label space and decision rules were held constant across languages, the most

plausible explanations are (a) cross-lingual differences in pragmatics (how affirmation, hedging, and stance are expressed); (b) script and tokenization effects that shape model attention and priors; and (c) model calibration differences across the English and Persian classifiers. The intensity profiles reinforce this interpretation: Persian's lower neutrality and broader spreads are consistent with a classifier (and discourse register) that distributes affect more decisively away from the center, while English appears more tightly centered around neutrality.

Regarding RQ5 (confidence characteristics), English exhibits steeper concentration near the neutral probability region, whereas Persian displays broader, more heterogeneous probability mass, especially on the positive pole. This asymmetry has interpretive consequences: confidence thresholds calibrated to English neutrality distributions will overstate "decisiveness" in English and understate it in Persian, unless calibrated separately.

Turning to RQ6 (substantive meaning of differences), the size and stability of the gaps visible at both the label and probability levels are large enough to matter for communication research. In audits of LLM behavior, newsroom analytics, or comparative political communication, a neutrality-skew in English versus a positivity-skew in Persian can produce different storylines about tone even when the underlying prompts are identical. Substantively, the safest reading is that language membership is not just a nuisance covariate; it is a consequential measurement dimension in its own right.

For RQ7 (persistence after accounting for structure), the persistence of the affective gap alongside only small structural differences argues against a length-based explanation. Even when one controls analytically for sentences/words/characters by examining distributions conditional on length bands or by focusing on probability space rather than hard labels, the neutrality-versus-positivity contrast remains the defining axis across languages. This points to linguistic and modeling factors rather than superficial length as the primary sources of divergence.

Implications. For multilingual content analysis, three practices are advisable. First, report both hard labels and probability summaries, and predefine language-specific calibration checks (e.g., reliability curves, threshold sensitivity). Second, adopt harmonized label spaces and decision rules, but allow post-hoc calibration by language (temperature/Dirichlet

scaling), so neutrality is not implicitly privileged in one language. Third, incorporate measurement-invariance diagnostics, e.g., replicate key contrasts using a multilingual model or a translate-to-pivot approach with human adjudication on a stratified subsample to distinguish linguistic pragmatics from model artifacts.

Taking the research questions together, structural differences between English and Persian are modest and do not account for the affective gap; the distribution of sentiment classes and the underlying probability profiles differ systematically by language; and language membership functions as a meaningful predictor of sentiment outcomes under a harmonized pipeline. For scholars and practitioners, the practical upshot is straightforward: in multilingual analyses of LLM outputs, language must be modeled, calibrated, and reported as a first-order measurement dimension, not treated as noise to be averaged away.

References

- Bhanvadia, S., Radha Saseendrakumar, B., Guo, J., Spadafore, M., Daniel, M., Lander, L., & Baxter, S. L. (2024). Evaluation of bias and gender/racial concordance based on sentiment analysis of narrative evaluations of clinical clerkships using natural language processing. *BMC medical education*, 24(1), 295. <https://doi.org/10.1186/s12909-024-05271-y>
- Bourdieu, P. (1991). *Language and symbolic power* (J. B. Thompson, Ed.; G. Raymond & M. Adamson, Trans.). Harvard University Press.
- Díaz, M., Johnson, I., Lazar, A., Piper, A. M., & Gergle, D. (2018, April). Addressing age-related bias in sentiment analysis. In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 1-14). <https://doi.org/10.1145/3173574.3173986>
- Entman, R. M. (2007). Framing bias: Media in the distribution of power. *Journal of Communication*, 57(1), 163-173. <https://doi.org/10.1111/j.1460-2466.2006.00336.x>
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Fairclough, N. (1995). *Critical discourse analysis: The critical study of language*. Longman.
- Hanna, M. G., Pantanowitz, L., Jackson, B., Palmer, O., Visweswaran, S., Pantanowitz, J., ... & Rashidi, H. H. (2025). Ethical and bias considerations in artificial intelligence/machine learning. *Modern Pathology*, 38(3), 100686. <https://doi.org/10.1016/j.modpat.2024.100686>
- Innis, H. A. (1951). *The bias of communication*. University of Toronto Press.
- McLuhan, M. (1964). *Understanding media: The extensions of man*. McGraw-Hill.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
- Pessach, D., & Shmueli, E. (2022). A review on fairness in machine learning. *ACM Computing Surveys (CSUR)*, 55(3), 1-44. <https://doi.org/10.1145/3494672>
- Radaideh, M. I., Kwon, O. H., & Radaideh, M. I. (2025). Fairness and social bias quantification in Large Language Models for sentiment analysis. *Knowledge-Based Systems*, 113569. <https://doi.org/10.1016/j.knosys.2025.113569>
- Rozado, D. (2020). Wide range screening of algorithmic bias in word embedding models using large sentiment lexicons reveals underreported bias types. *PLoS one*, 15(4), e0231189. <https://doi.org/10.1371/journal.pone.0231189>
- Sabbar, S., & Habib Zadeh Khiyaban, S. (2023). Algorithms of Displacement: Emotional and Rhetorical Responses to AI-Driven Job Loss in Digital Public Discourse. *International Journal of Advanced Multidisciplinary Research and Studies*, 3(4), 1324-1331. <https://doi.org/10.62225/2583049X.2023.3.4.5012>
- Salehi, K., Habib Zadeh Khiyaban, S., Sabbar, S. (2026). Artificial Intelligence and Crime Detection: A Critical Review. *Cyberspace Studies*. 10(1): 181-197. <https://doi.org/10.22059/jcss.2025.402206.1179>

- Shahghasemi, E. (2025). AI; A Human Future. *Journal of Cyberspace Studies*, 9(1), 145-173. doi: 10.22059/jcss.2025.389027.1123
- Shahghasemi, E., Gholami, F. & Alikhani, Z. (2025). Global patterns of social media use and political sentiment. *Discover Global Society*, 3, 36. <https://doi.org/10.1007/s44282-025-00171-y>
- Tejani, A. S., Ng, Y. S., Xi, Y., & Rayan, J. C. (2024). Understanding and mitigating bias in imaging artificial intelligence. *Radiographics*, 44(5), e230067. <https://doi.org/10.1148/rg.230067>
- Thelwall, M. (2018). Gender bias in sentiment analysis. *Online Information Review*, 42(1), 45-57. <https://doi.org/10.1108/OIR-05-2017-0139>
- van Dijk, T. A. (1993). Principles of critical discourse analysis. *Discourse & Society*, 4(2), 249-283. <https://doi.org/10.1177/0957926593004002006>
- Venkit, P. N., & Wilson, S. (2021). Identification of bias against people with disabilities in sentiment analysis and toxicity detection models. *arXiv preprint arXiv:2111.13259*. <https://doi.org/10.48550/arXiv.2111.13259>
- Venugopal, J. P., Subramanian, A. A. V., Sundaram, G., Rivera, M., & Wheeler, P. (2024). A Comprehensive Approach to Bias Mitigation for Sentiment Analysis of Social Media Data. *Applied Sciences*, 14(23), 11471. <https://doi.org/10.3390/app142311471>
- Winner, L. (1980). Do artifacts have politics? *Daedalus*, 109(1), 121-136. <http://www.jstor.org/stable/20024652>



Original-Forschungsarbeit

Zukunft des öffentlichen Vertrauens in Medien im Zeitalter der künstlichen Intelligenz: Szenarienplanung für den Iran 2036

Amir Garousi^{1*}, Mahmood Jamali², Einollah Keshavarz Turk³

¹ Doktorand im Medienmanagement, Universität Tehran, Teheran, Iran

² Doktorand im Medienmanagement, Universität Tehran, Teheran, Iran

³ Außerordentlicher Professor für Zukunftsstudien, Imam-Khomeini-International-Universität, Tehran, Iran

Empfangen: 1. April 2025 **Akzeptiert:** 11. Juni 2025

Zusammenfassung:

Das öffentliche Vertrauen in Medien ist ein zentraler Bestandteil des Sozialkapitals und der kommunikativen Legitimität, wird jedoch zunehmend durch die schnelle Integration von Künstlicher Intelligenz und synthetischen Medien in die Nachrichtenproduktion und -verbreitung herausgefordert. Diese Studie untersucht alternative Zukünfte des öffentlichen Vertrauens in Medien im Zeitalter der KI und entwickelt szenariobasierte Erkenntnisse für den Iran bis zum Jahr 2036. Unter Verwendung eines Zukunftsstudien-Ansatzes kombiniert die Forschung Environmental Scanning und eine systematische Überprüfung akademischer und politischer Quellen (2018–2025) mit einer zweirundigen Delphi-Befragung von fünfzehn Expert:innen aus Medien, KI und Governance. Eine Strukturanalyse mittels der MICMAC-Methode untersuchte Einfluss-Abhängigkeits-Beziehungen zwischen Schlüsselvariablen und identifizierte Medien-Transparenz und die Qualität der KI-Regulierung als zwei kritische Unsicherheiten, die die zukünftigen Vertrauensverläufe prägen. Auf Basis dieser Achsen wurden vier alternative Szenarien entwickelt: Smart Trust, Total Distrust, Islands of Trust und Imposed Trust, die jeweils unterschiedliche Konfigurationen von Governance-Entscheidungen, Technologieeinsatz und Reaktionen des Publikums darstellen. Die Ergebnisse zeigen, dass zukünftige Muster des öffentlichen Vertrauens nicht technologisch deterministisch sind, sondern hauptsächlich durch institutionelle Transparenz, regulatorische Maßnahmen und Governance-Entscheidungen bestimmt werden. Die Studie schließt mit der Empfehlung, verantwortliche KI-Governance zu stärken, Medien-Transparenz zu erhöhen und in Medienkompetenz der Bevölkerung zu investieren, um das Mediensystem des Iran in eine nachhaltige und vertrauensbasierte Zukunft zu steuern.

Schlüsselwörter: zukunftsstudien, öffentliches vertrauen, medien, künstliche intelligenz, szenarienplanung

* Korrespondierender Autor

✉ garousi@ut.ac.ir

🌐 <https://orcid.org/0009-0003-5431-6705>

Wie dieser Artikel zu zitieren ist:

Garousi, A., Jamali, M., & Keshavarz Turk, E. (2025). Futures of public trust in media in the age of artificial intelligence: Scenario planning for Iran 2036. *Spektrum Iran*, 38(2), 159-186.

🔗 <https://doi.org/10.22034/spektrum.2026.566873.1056>

© 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



مقاله پژوهشی

آینده اعتماد عمومی به رسانه‌ها در عصر هوش مصنوعی: برنامه‌ریزی سناریویی برای ایران ۲۰۳۶

امیر گروسی^{۱*}، محمود جمالی^۲، عین‌الله کشاورز ترک^۳

۱ دانشجوی دکتری مدیریت رسانه، دانشگاه تهران، تهران، ایران

۲ دانشجوی دکتری مدیریت رسانه، دانشگاه تهران، تهران، ایران

۳ دانشیار مطالعات آینده، دانشگاه بین‌المللی امام خمینی، تهران، ایران

دریافت: ۱۴۰۴/۱/۱۲؛ پذیرش: ۱۴۰۴/۳/۲۱

چکیده:

اعتماد عمومی به رسانه‌ها بخش اساسی سرمایه اجتماعی و مشروعیت ارتباطی است، اما با یکپارچگی سریع هوش مصنوعی و رسانه‌های مصنوعی در فرایند تولید و توزیع اخبار، چالش‌های فزاینده‌ای را تجربه می‌کند. این مطالعه آینده‌های جایگزین اعتماد عمومی به رسانه‌ها در عصر هوش مصنوعی را بررسی کرده و دیدگاه‌های مبتنی بر سناریو برای ایران تا افق ۲۰۳۶ ارائه می‌دهد. با استفاده از رویکرد مطالعات آینده، پژوهش از طراحی روش‌های ترکیبی بهره می‌برد که اسکن محیطی و مرور نظام‌مند منابع علمی و سیاست‌گذاری (۲۰۱۸-۲۰۲۵) را با دو دور مشاوره دلفی شامل پانزده کارشناس در زمینه رسانه، هوش مصنوعی و حاکمیت ترکیب می‌کند. تحلیل ساختاری با استفاده از روش MICMAC برای بررسی روابط تأثیر-وابستگی بین متغیرهای کلیدی انجام شد و شفافیت رسانه و کیفیت مقررات هوش مصنوعی به‌عنوان دو عدم قطعیت حیاتی که مسیرهای اعتماد آینده را شکل می‌دهند، شناسایی شدند. بر اساس این محورها، چهار سناریوی جایگزین توسعه یافتند: اعتماد هوشمند، بی‌اعتمادی کامل، جزایر اعتماد و اعتماد تحمیلی، که هر کدام ترکیب متفاوتی از انتخاب‌های حاکمیتی، کاربرد فناوری و پاسخ مخاطب را نشان می‌دهند. یافته‌ها نشان می‌دهند که الگوهای آینده اعتماد عمومی فناوری‌ها تعیین شده نیستند، بلکه عمدتاً توسط شفافیت نهادی، سازوکارهای قانونی و تصمیمات حاکمیتی هدایت می‌شوند. مطالعه نتیجه می‌گیرد که تقویت حاکمیت پاسخگو در هوش مصنوعی، افزایش شفافیت رسانه‌ای و سرمایه‌گذاری در سواد رسانه‌ای مخاطبان برای هدایت اکوسیستم رسانه‌ای ایران به سمت آینده‌ای پایدار و مبتنی بر اعتماد ضروری است.

واژگان کلیدی: مطالعات آینده، اعتماد عمومی، رسانه، هوش مصنوعی، برنامه‌ریزی سناریویی

* نویسنده مسئول

<https://orcid.org/0009-0003-5431-6705>

garousi@ut.ac.ir

<https://doi.org/10.22034/spektrum.2026.566873.1056>



Original Research Paper

Futures of public trust in media in the age of artificial intelligence: Scenario planning for Iran 2036

Amir Garousi^{1*}, Mahmood Jamali², Einollah Keshavarz Turk³

¹ PhD Student in Media Management, University of Tehran, Tehran, Iran

² PhD Student in Media Management, University of Tehran, Tehran, Iran

³ Associate Professor of Futures Studies, Imam Khomeini International University, Tehran, Iran

Received: Apr. 01, 2025 Accepted: Jun. 11, 2025

Abstract

Public trust in media constitutes a core component of social capital and communicative legitimacy, yet it is increasingly challenged by the rapid integration of artificial intelligence and synthetic media into news production and distribution processes. This study explores alternative futures of public trust in media in the age of artificial intelligence and develops scenario-based insights for Iran toward the horizon of 2036. Adopting a futures-studies approach, the research employs a mixed-methods design that combines environmental scanning and a systematic review of academic and policy sources (2018–2025) with a two-round Delphi consultation involving fifteen experts in media, artificial intelligence, and governance. Structural analysis using the MICMAC method was applied to examine influence–dependence relationships among key variables, leading to the identification of media transparency and the quality of AI regulation as the two critical uncertainties shaping future trajectories of public trust. Based on these axes, four alternative scenarios were developed—Smart Trust, Total Distrust, Islands of Trust, and Imposed Trust—each illustrating a distinct configuration of governance choices, technological use, and audience responses. The findings demonstrate that future patterns of public trust are not technologically deterministic but are primarily driven by institutional transparency, regulatory arrangements, and governance decisions. The study concludes that strengthening accountable AI governance, enhancing media transparency, and investing in media literacy among audiences are essential for steering Iran’s media ecosystem toward a sustainable and trust-based future.

Keywords: futures studies, public trust, media, artificial intelligence, scenario planning

* Corresponding Author

✉ garousi@ut.ac.ir

🌐 <https://orcid.org/0009-0003-5431-6705>

How to Cite this Article:

Garousi, A., Jamali, M., & Keshavarz Turk, E. (2025). Futures of public trust in media in the age of artificial intelligence: Scenario planning for Iran 2036. *Spektrum Iran*, 38(2), 159-186.

📄 <https://doi.org/10.22034/spektrum.2026.566873.1056>

1. Introduction

Public trust in the media has always been one of the main pillars of the stability of the political system and social cohesion. In modern societies, media serve not only as tools for transmitting news but also as key actors in shaping public opinion, strengthening social capital, and creating collective solidarity (Coleman, 1990). In fact, the level of people's trust in the media is considered an important criterion for assessing the health of social communications and the legitimacy of political institutions. However, over the past two decades, the decline in public trust in the media has been widely noted by researchers and international institutions. The annual report of the Reuters Institute shows that the average public trust in news media globally decreased to about 40% in 2023; moreover, many countries have faced a downward trend in this index in recent years, which indicates a crisis of trust in the media arena (Newman et al., 2023).

The emergence of new technologies, especially artificial intelligence, has added further complexity to this issue. Artificial intelligence algorithms are now used at all stages of the news cycle, from production and editing to distribution and content personalization. This transformation, on the one hand, provides opportunities for improving content quality, increasing speed, and responding to individual audience needs; on the other hand, it brings serious threats, including the spread of fake news, rapid dissemination of rumors, and the emergence of phenomena such as deepfakes that can blur the boundary between reality and imagination in the public sphere (Floridi & Chiriatti, 2020). Such conditions have turned the issue of trust in the media into one of the critical topics in the era of artificial intelligence.

The significance of this topic is heightened in Iran. Iranian society, with generational and cultural diversity in media consumption patterns, has a distinct encounter with new technologies. Younger generations have turned more to social media and digital platforms, while older generations are still somewhat dependent on traditional media. This generational gap, along with varying levels of media literacy among social groups, causes public trust in Iran to be neither uniform nor homogeneous, but multilayered and pluralistic. In addition, the role of regulatory policies in managing or restricting the media can have a direct impact on public trust. As a result, the future of trust in the media in Iran in the context of artificial intelligence will

not only be a reflection of technological developments but also the product of complex interactions among society, policy, and technology.

Therefore, the present research, utilizing a futures studies approach and especially the scenario writing method, seeks to draw a perspective of the future of public trust in the media in Iran up to the horizon of year 1415 SH. In this regard, the study presents alternative scenarios for possible futures of public trust by identifying macro trends, main drivers, and critical uncertainties. The ultimate goal of the research is to help policymakers, media managers, and researchers design effective strategies to strengthen public trust in the face of rapid transformations in artificial intelligence by clarifying probable paths.

2. Literature Review

Domestic research has generally focused on three main axes: trust in national media, the role of social networks in public opinion, and forward-looking applications of artificial intelligence in media. Safavi et al. (1402 SH), by designing a model for science journalism in the era of new media, emphasized components such as communication between the scientific community and the public, empowering journalists, and strengthening infrastructures. Although this research highlights the importance of trust in media in the scientific domain, its focus on science journalism has caused broader dimensions of public trust in media, especially in political and social areas, to be overlooked. Ajlali and Khatibi (1403 SH), by examining the role of social networks and algorithms in cognitive warfare, showed that algorithms can intensify fake news and create filter bubbles. This study has effectively revealed the threats of artificial intelligence in the arena of public opinion, but it has limitations in the field of futures studies and scenario writing, and has merely limited to case analysis. Mohammadnejad and Shahmohammadi (1400 SH) focused on strategies to enhance trust in national media news and used SWOT and QSPM methods, providing suggestions such as strengthening adherence to legal standards and structural revision. The strength of this research is addressing operational policy-making, but its limitation is its sole focus on national media and a lack of attention to digital and social media. In the field of futures studies, two studies by Taghipour et al. (1404 SH) and Al-Mohammad and Asadi (1403 SH) are prominent.

Taghipour, using Delphi and scenario writing, showed that artificial intelligence can improve content quality, but challenges such as privacy also exist. Al-Mohammad and Asadi also focused on constants, trends, and uncertainties of future journalism with the Pilkkan five-factor model. Both studies have provided a valuable perspective on future media transformations, but they have paid less attention to "public trust" as a key variable. Overall, domestic studies, although emphasizing the importance of artificial intelligence, trust in national media, and future transformations in journalism, often lack a comprehensive framework for combining these three dimensions in the form of futures studies on public trust.

At the international level, research has more broadly addressed dimensions of trust, fake news, and the impact of artificial intelligence on journalism. Vaccari and Chadwick (2020) showed that deepfakes more often cause doubt and cynicism than deception, and thus intensify the crisis of trust. Opdahl et al. (2023), in contrast, have a more optimistic approach and introduce artificial intelligence as a tool for strengthening trust through improving journalism quality. These two studies together show that trust can be both weakened and strengthened. Peña Fernández et al. (2023), emphasizing the social dimension of generative artificial intelligence, addressed the risks of media dependency on technological platforms and the threat to journalists' independence. This view has great importance in clarifying social consequences, but it has paid less attention to forward-looking perspectives. Toff and Simon (2024) examined the issue of disclosing the use of artificial intelligence in news production and showed that labeling can reduce trust unless the sources used are transparently clarified.

This research is one of the first efforts to empirically test the relationship between "technological transparency" and "public trust." In other research, Neyazi et al. (2025) examined the role of information seeking in shaping trust in artificial intelligence and showed contradictory results in traditional and social media. Karaaslan et al. (2024) analyzed public attitudes toward automated news and stated that people's awareness of artificial intelligence journalism is relative, and concerns still persist. Kim et al. (2023) also technologically addressed the use of artificial intelligence and blockchain in detecting fake news, while Jungherr and Schroeder (2023) and Robles and Mallinson (2025) focused on structural and governance dimensions of public

trust in the arena of artificial intelligence. Overall, international studies compared to domestic ones have less thematic fragmentation and a more comprehensive view of the link between artificial intelligence and public trust, but scenario writing in futures studies in this area is still very limited. A structured overview of prior domestic and international studies is provided in Appendix A.

Overall, domestic studies tend to address public trust through normative or institutional lenses, often focusing on national media and short-term policy concerns, while international scholarship emphasizes empirical assessment of AI-driven media practices with limited contextualization for non-Western governance environments. What remains underdeveloped in both streams is an integrated futures-oriented framework that conceptualizes public trust as an outcome of interacting technological, regulatory, and socio-cultural dynamics over time. This study addresses this gap by combining futures studies methodologies with a governance-sensitive understanding of AI-mediated media trust in the Iranian context.

3. Theoretical Foundations

3.1. Artificial Intelligence and the Transformation of Media Trust

In this study, public trust in media is defined as a relational and dynamic expectation through which audiences assess the credibility, integrity, and accountability of media institutions based on their perceived transparency, governance arrangements, and performance over time. Public trust is operationalized not as a fixed attitude toward individual media messages, but as a systemic outcome emerging from repeated interactions among media organizations, technological infrastructures—particularly artificial intelligence systems—and regulatory frameworks. From this perspective, trust reflects audiences' confidence that media actors will use technologies responsibly, disclose relevant information about content production processes, and remain accountable to societal norms and public interests under conditions of uncertainty.

Public trust in the media, from the perspective of classical communication theories, is a fundamental component of social capital and a mechanism for reducing uncertainty in collective actions. In this framework, media act as

institutions that strengthen norms of cooperation and consolidate social cohesion by providing access to credible information (Coleman, 1990). However, the rise of new technologies—and especially artificial intelligence—has qualitatively transformed the dynamics of trust formation and reproduction. In recent decades, a range of media innovations from satellites and the internet to social networks and digital platforms have continuously shifted the boundaries of content production and distribution. However, the turning point is the entry of artificial intelligence, which not only provides new tools to media actors but redefines the essence of production, editing, distribution, and content consumption processes and expands the scope of action from a single human organization to a dynamic network of journalists, algorithms, audiences, and technological infrastructures (Floridi & Chiriatti, 2020). In this perspective, public trust in the era of artificial intelligence is the result of a multilevel interaction among three intertwined domains: media as communication institutions, artificial intelligence as a transformative agent, and audiences as active receivers and interpreters of messages. Empirical research on digital public discourse further indicates that audiences interpret AI-driven transformations through emotionally structured narratives that shape perceptions of institutional credibility and legitimacy in online environments (Sabbar & Khiyaban, 2023). This reinforces the understanding of trust as an interactional and discursively mediated process rather than a purely technical or procedural outcome.

Findings from recent research show that artificial intelligence has found an effective presence in most links of the news production cycle: from automated production of text and images, estimation and analysis of big data, to personalized distribution and audience exposure to content. In this regard, empirical findings in the local context confirm that effective implementation of artificial intelligence in media is not realized solely by relying on "technology"; rather, it requires the alignment of three interrelated domains: causal factors such as data infrastructure and computational capacity, financial investment, and precise knowledge of audience characteristics; contextual factors such as specialized training of journalists, media richness, skilled human capital, and organizational knowledge maturity; and intervening factors such as organizational and social culture, regulatory environment, institutional collaborations, and managers' attitudes. In other words, success in utilizing artificial intelligence is the result of aligning structural, cultural, and

managerial components and rather than a purely technological achievement (Soltanpour et al., 2024). This conclusion is consistent with foundational definitions of artificial intelligence; a definition that considers it "the study and construction of agents that do the right thing," referring both to machines' capability to perform complex tasks and to the capacity for goal-setting, reasoning, and adaptation in uncertain environments. The current successes of artificial intelligence are owed to leaps in machine learning and especially deep learning that have provided computational sufficiency for perceptual and linguistic tasks (Jungheer & Schroeder, 2023).

From a policy and organizational perspective, responsible implementation of artificial intelligence in media requires complementary strategies: general and specialized education to enhance journalists' digital competencies and increase audiences' media literacy; strengthening the link between media practitioners and technology experts; forecasting sustainable financing mechanisms for innovative projects; and finally establishing intelligent monitoring systems for algorithms to control errors, biases, and unintended side effects. The natural consequence of such strategies— if properly designed and implemented—is increasing data processing speed, improving verification accuracy, enabling automated routine production, and allocating human time to more analytical and value-added tasks; which can simultaneously improve organizational efficiency and the quality of news experiences (Soltanpour et al., 2024). However, just as artificial intelligence has the capacity to facilitate the realization of shared social values, it can also create conditions for weakening them. For this reason, "public trust" must be the axis of designing artificial intelligence governance frameworks (Schiff et al., 2021). Empirical research on large-scale social media discourse further indicates that public perceptions of artificial intelligence are shaped by contested geopolitical narratives, technological rivalry, and affective responses such as fear and skepticism, all of which influence how the legitimacy of AI governance is evaluated in the public sphere (Salehi et al., 2025). This reinforces the view that trust in AI-mediated systems emerges within broader discursive and political contexts rather than being determined solely by technical performance. A critical review of artificial intelligence governance literature shows that the dimension of public trust has been insufficiently conceptualized as a foundational component, and this absence has created "blind spots" in decision-making in

policy-making for emerging technologies; a situation in which social distrust turns into real policy challenges, but governance mechanisms are unable to anticipate and respond to it in a timely manner (Robles & Mallinson, 2025; Veen et al., 2011). At the newsroom level, increasing dependence on technological platforms and generative services can threaten editorial independence, weaken the symbolic position of journalists, and ultimately lead to the erosion of professional social capital (Peña Fernández et al., 2023). Therefore, balancing technological efficiency and normative values of journalism – including honesty, transparency, and social responsibility – is a necessary condition for rebuilding and sustaining public trust; neglecting this balance increases the risk of turning artificial intelligence from an "opportunity" to a "threat" for media legitimacy (Floridi & Chiriatti, 2020; Peña Fernández et al., 2023).

3.2. Transparency, Labeling, and Multidimensional Levels of Trust

At the micro level of audience encounter with content, one of the most controversial policy interventions is "disclosure/labeling," of the use of artificial intelligence in news production. Empirical evidence shows that merely labeling "produced with artificial intelligence" does not necessarily lead to increased trust and can even have a negative effect on the evaluation of news credibility – even when the accuracy and fairness of labeled content do not meaningfully differ from human-produced text. Moreover, sensitivity to this label is more intense among audiences who previously had higher trust in media or better knowledge of journalism processes. However, when media, in addition to the label, transparently disclose the list of sources, data processing methods, and the level of human intervention, the negative effect of labeling decreases considerably. These findings confirm that "multilayered transparency" – not merely attaching a label – is the key to rebuilding trust and must be formulated in the form of a comprehensive communication policy with a clear explanation of content production mechanisms (Toff & Simon, 2024). It becomes even more significant when we consider when we consider the rapid growth of synthetic media in digital ecosystems: analyses show that in the period from December 2022 to October 2023, the prevalence of synthetic media on platform X has had a noticeable jump, and the limited sample examined has gained over 1.5 billion views – a jump that intensified with the release of Midjourney and shows that monitoring, measuring, and

understanding the dynamics of synthetic media proliferation is an increasing necessity for protecting public trust (Corsi et al., 2024).

Of course, public trust is not merely the outcome of regulation and institutional intervention; audience characteristics, lived experiences, and their media/data literacy also play a determining role in trust evaluation. Changes in news tastes, increasing need for speed and accuracy, and direct encounter with fake news on social networks have increased audiences' sensitivity to content authenticity and credibility. For groups that have experienced fake news encounters, simple labeling is not sufficient, and they demand more reassuring mechanisms for verification. In contrast, audiences less familiar with content production processes and technological boundaries may perceive the "produced with artificial intelligence" label as a threat to news authenticity, but more literate audiences see it as a sign of transparency and accountability (Brewer et al., 2022). This pattern is consistent with the literature on the formation of public attitudes toward emerging technologies: public evaluations often form before encountering official information and under the influence of cultural and media framings; hence, post hoc communication interventions, if lacking subtlety, may lead to strengthening "pre-existing pessimistic stereotypes" instead of repairing trust (Druckman & Bolsen, 2011; Brewer et al., 2022). As a result, media literacy is not just a side educational policy but the axis of trust strategy. Recent integrative research on AI literacy emphasizes that effective literacy extends beyond technical skills to include ethical awareness, critical evaluation of algorithmic systems, and informed engagement with AI's societal implications (Khodabin et al., 2022). This supports the view that media literacy must function as a multidimensional capacity enabling audiences to interpret transparency signals and assess AI-mediated content. It must enhance audiences' understanding of algorithmic mechanisms, technical limitations, and verification methods through layered programs of general and specialized education to strengthen the ability to distinguish between transparency signs and "fake trust signals" (Soltanpour et al., 2024; Toff & Simon, 2024).

For operational formulation of trust in the era of artificial intelligence, one can distinguish four interconnected levels from a systemic perspective. The first level is "infrastructure and investment": access to up-to-date hardware/software, high-quality data, and sustainable financing is a necessary condition for effective utilization of AI capacities; weaknesses at

this level not only brings technical inefficiency and errors but translates into trust erosion in the audience experience (Soltanpour et al., 2024). This proposition is consistent with broader evidence: lack of information transparency and public participation opportunities in relation to emerging technologies reduces social acceptance and sows the seeds of policy distrust (Veen et al., 2011). The second level is "institutional capabilities": without technological and ethical competencies of journalists and managers, and without an explicit strategy for responsible use of artificial intelligence, technology application is reduced to superficial and non-transparent interventions, and trust is damaged (Peña Fernández et al., 2023; Soltanpour et al., 2024). The third level is "audience characteristics": media/data literacy, experiences of encountering fake news, and content consumption preferences shape the the stability of trust/distrust (Brewer et al., 2022; Druckman & Bolsen, 2011). And the fourth level is "policy-making and regulation": transparent laws, support for information freedom, and ethical frameworks for AI application can strengthen trust; in contrast, monopolistic control or regulatory ambiguity intensifies trust gaps (Schiff et al., 2021; Robles & Mallinson, 2025). The sum of these four levels determines the probable paths of trust over the mid-term horizons: if multilayered transparency is accompanied by effective governance and institutional/social empowerment, artificial intelligence becomes a tool for intelligent trust rebuilding; but in the absence of this balance, the same technology can become an engine for producing doubt, polarization, and legitimacy erosion (Floridi & Chiriatti, 2020; Toff & Simon, 2024; Corsi et al., 2024).

Based on this, the theoretical foundations of media trust in the era of artificial intelligence can be formulated in a concise summary: trust is not a linear output of content quality, but the consequence of dynamic interaction among infrastructure, institutional capacity, audience characteristics, and governance; artificial intelligence has a dual capacity that, with appropriate institutional and policy design, can be directed toward "intelligent trust"; and "meaningful transparency"—relying on disclosure of sources, process, and human role—along with enhancing media literacy, is the key to neutralizing side effects of labeling and curbing risks of synthetic media (Soltanpour et al., 2024; Schiff et al., 2021; Robles & Mallinson, 2025; Veen et al., 2011; Peña Fernández et al., 2023; Toff & Simon, 2024; Corsi et al., 2024; Brewer et al., 2022; Druckman & Bolsen, 2011; Floridi & Chiriatti, 2020; Jungherr & Schroeder, 2023).

3.3. Conceptual Framework of the Article

Public trust in the media in the era of artificial intelligence is a multidimensional phenomenon shaped by the interaction of three main domains: trust as social capital, media as communication and information institutions, and new technologies, especially artificial intelligence, as a transformative agent in content production and distribution. In this framework, drivers such as technological transformation in the domain of algorithms, blockchain, and intelligent content production, media policy-making and regulation, citizens' level of media and data literacy, and the degree of media transparency and accountability act as key drivers. However, the future of public trust is not merely a function of these drivers but is heavily influenced by two critical uncertainties: first, the level of media transparency and accountability to audiences; second, the degree of intervention and regulation of artificial intelligence in media processes. These two factors can change the future path in the direction of strengthening or weakening public trust. Based on this, the conceptual framework of the article is based on the assumption that the combination of drivers and uncertainties will lead to the formation of alternative scenarios for public trust in Iran. These scenarios include intelligent trust (when media are transparent and AI acts responsibly), total distrust (in the absence of transparency and technology abandonment), islands of trust (trust of part of society in specific media), and imposed trust (superficial trust resulting from control and monopoly of information flow). Thus, the present conceptual framework seeks to provide an integrated image of the link between public trust, media, and artificial intelligence and, with a futures studies approach, explain the probable paths of this link in the horizon of year 1415 SH for Iran.

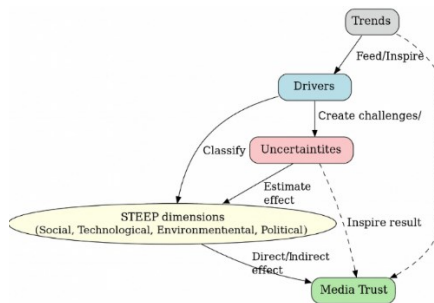


Figure 1. Research Conceptual Model

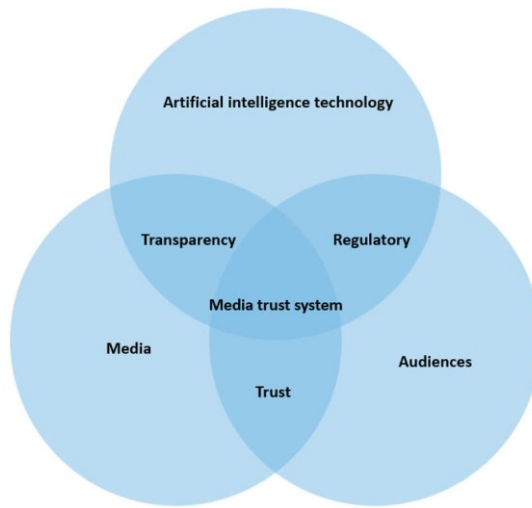


Figure 2. Research Conceptual Model

4. Methodology

This research was conducted using a scenario-based futures studies approach in the form of a mixed exploratory-explanatory study. Its goal was to draw alternative futures of public trust in media in the era of artificial intelligence up to the horizon of year 1415 SH in Iran. The research design has been carried out in several consecutive steps.

In the first step, through environmental scanning and systematic review of sources (2018–2025), including scientific articles, policy documents, and professional reports, a set of trends, drivers, and uncertainties related to media trust and artificial intelligence applications were identified. Then, using the STEEP framework (social, technological, economic, environmental/infrastructural, and political-governance), these findings were classified and clustered to form the initial portfolio of drivers and uncertainties. The search strategy was designed based on combining Persian/English keywords including ["media trust", "AI", "synthetic media", "algorithmic transparency", "fact-checking", "Iran"] and their Persian equivalents. Databases Google Scholar, Scopus, Web of Science, and professional institutes' reports were searched during the period 2018–2025. Entry criteria: (1) direct focus on media trust/AI application in news; (2)

specified methodology; (3) access to full text. Exit criteria: (1) non-peer-reviewed articles/press notes; (2) content overlap; (3) unrelated subject scope. In total, 571 records were identified; after removing duplicates (129) and screening titles/abstracts (330), 112 texts remained for full review, and finally 34 sources entered the analysis.

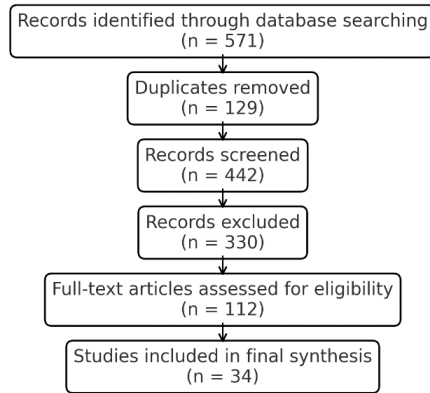


Diagram 1. Research PRISMA

In the next stage, an expert panel consisting of 15 specialists in media, information technology, policy-making, and civil society was formed. Using the Delphi technique, the list of key factors was reviewed and completed, and based on experts' scoring, the most important drivers and uncertainties were identified.

After that, using structural analysis (MICMAC), relationships among variables were examined to clarify the degree of influence and dependence of each factor. The results of this stage enabled the selection of two critical uncertainties—"media transparency level" and "intensity and quality of artificial intelligence regulation"—as the main axes for scenario writing.

In the scenario writing step, based on combining these two axes, four alternative scenarios were drawn: intelligent trust, total distrust, islands of trust, and imposed trust. For each scenario, an analytical narrative, implications, and signposts were developed. Finally, using stakeholder workshops, proposed policies and strategies for each scenario were evaluated to provide a set of resilient policy options for managing the future trajectories of public trust in media.

This integrative approach, while relying on secondary data and expert opinions, seeks to provide a comprehensive and reliable image of the future of media trust in the era of artificial intelligence for Iran. Details regarding the composition of the expert panel are provided in Appendix B.

Experts were selected through purposive and snowball method to cover diversity in media/editorial, AI technology, governance, and civil society domains. Entry criteria: at least 10 years of related professional/research experience, at least three published works or policy projects related to the topic, familiarity with media trust/AI topic, and readiness to participate in 2 Delphi rounds. In total, invitations were sent to 17 people; 15 agreed and 15 experts completed the first round questionnaire. All participation was voluntarily, without financial compensation and based on informed consent.

Table 1. Drivers and Uncertainties Scoring Matrix

Factor	Importance 1-9	Uncertainty 1-9	Influence (MICMAC) 0-3	Dependence 0-3
Transparency and accountability of the media	9	8	3	2
Intensity and quality of AI regulation in the media	9	9	3	3
Audience media and data literacy	8	6	2	2
Editorial independence and newsroom governance	8	7	3	2
Maturity of data infrastructure and fact-checking	7	5	2	1
Media business model (advertising dependency)	7	7	2	3

Delphi was conducted in two rounds. In each round, experts evaluated the identified factors along the dimensions of importance and uncertainty using a nine-point Likert scale. Consensus criteria were defined a priori as a median score of at least 7 and an interquartile range (IQR) of 1 or less, which are commonly accepted thresholds in Delphi-based futures research. Across the two rounds, these criteria were met for the core drivers and uncertainties retained for further analysis.

The overall level of agreement among experts was assessed using Kendall's coefficient of concordance, which indicated a satisfactory degree of consensus ($W = 0.74$, $p < 0.001$). Internal consistency of the Delphi questionnaire in the first round was evaluated using Cronbach's alpha, yielding an acceptable reliability coefficient ($\alpha = 0.79$). After each round,

anonymized group feedback (median values and dispersion indicators) was provided to participants to facilitate reflection and convergence in subsequent assessments.

To enhance methodological transparency and replicability, this study reports key reliability and consensus indicators for the Delphi process, including interquartile ranges, Kendall's coefficient of concordance, and internal consistency measures. Expert selection followed clearly defined criteria regarding experience, domain diversity, and familiarity with AI-mediated media governance, ensuring both epistemic robustness and contextual relevance.

5. Findings

The results of implementing futures studies stages and analyzing qualitative and quantitative data showed that, in Iran, public trust in media in the era of artificial intelligence by the horizon year 1415 SH is influenced by a set of key drivers and uncertainties. The first set of findings concerns the Delphi analysis and driver identification. Experts reached consensus that transparency in content production process, the level of audience media literacy, and the quality of artificial intelligence regulation are three central factors shaping the future of public trust. Alongside these factors, editorial independence and the maturity of technological infrastructure (such as intelligent fact-checking systems and open databases) were also identified as influential variables, though with lower priority than critical uncertainties. The extended STEEP analysis informing the identification of key drivers and uncertainties is reported in Appendix C.

Findings from the MICMAC analysis showed that two variables – “media transparency and accountability” and “intensity and quality of artificial intelligence regulation” – exhibit the highest degree of influence and uncertainty. These two variables were selected as axes for scenario design. In contrast, factors such as media business models, societal news consumption patterns, and journalists' independence level were introduced as sensitive and dependent variables whose changes are subject to transformations in the main variables.

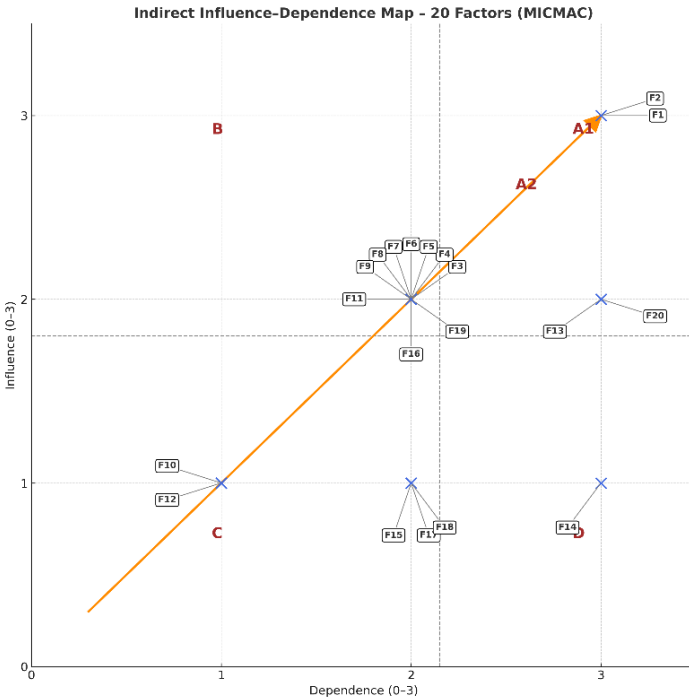


Figure 7. Indirect Influence-Dependence Map of 20 Key Drivers of Public Trust in Media (MICMAC Analysis)

This figure visualizes the distribution of twenty key drivers derived from the Delphi and MICMAC structural analysis. The X-axis represents the level of dependence, and the Y-axis represents the degree of influence on a 0-3 scale. The dashed lines mark the mean values and divide the plane into four standard quadrants (A1, A2, B, C).

The results indicate that F1 (Transparency & Accountability) and F2 (AI Regulation) exhibit the highest influence and dependence values, positioning them as the main scenario axes shaping the future of public trust in media. Additional variables such as F13 (Platform Governance in Elections) and F20 (Data Governance & Interoperability) also appear in the upper quadrants, highlighting the institutional and technological dynamics driving the evolution of trust in AI-mediated journalism.

Table 2. Key Factors Identified in the MICMAC Analysis and Their Influence–Dependence Scores

Code	Factor	Influence (0-3)	Dependence (0-3)
F1	Transparency & Accountability	3	3
F2	AI Regulation (Scope & Enforcement)	3	3
F3	Audience Literacy (Media/Data)	2	2
F4	Personalization & Recommenders	2	2
F5	Synthetic Media (Text/Audio/Video)	2	2
F6	Authenticity Tooling (C2PA/Watermark)	2	2
F7	Automated News Agents (Newsbots)	2	2
F8	Automated Fact-checking	2	2
F9	Data Infrastructure & Access	2	2
F10	Cybersecurity & XAI	1	1
F11	Privacy Concerns & Data Rights	2	2
F12	Influencers & Citizen Journalism	1	1
F13	Platform Governance in Elections	2	3
F14	Business Model (Ad Dependence)	1	3
F15	Revenue Diversification (Subs/Micro)	1	2
F16	Cost Pressure & Newsroom Automation	2	2
F17	Migration Across Platforms (Audiences)	1	2
F18	News Fatigue & Avoidance	1	2
F19	Deepfake Risk	2	2
F20	Data Governance & Interoperability	2	3

This table presents the twenty key STEEP drivers derived from the Delphi rounds and MICMAC structural analysis. Influence and dependence values (scale 0–3) indicate each variable’s systemic role. The variables F1 (Transparency & Accountability) and F2 (AI Regulation) show the highest scores and were selected as the main scenario axes.

Direct Influence Graph (20 Factors, MICMAC)

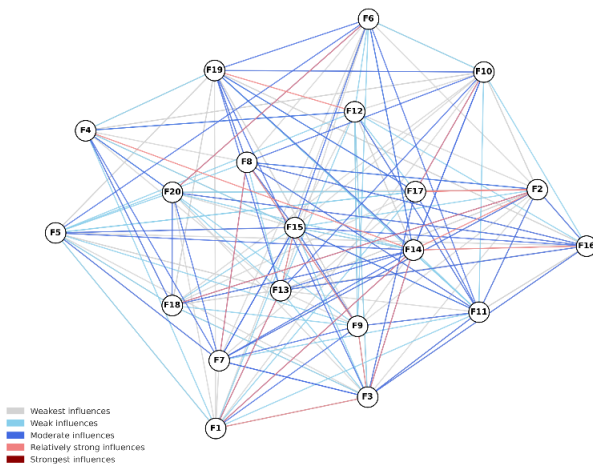


Figure 4. Direct Influence Graph

This graph illustrates the network of direct relationships among the twenty key factors identified in the MICMAC analysis. Each node (F1–F20) represents a variable, and the arrows and colored lines indicate the direction and strength of their direct influences. The color gradient represents five levels of influence intensity: light gray – weakest, sky blue – weak, royal blue – moderate, light coral – relatively strong, and dark red – strongest.

The dense central cluster indicates strong interdependence among core variables, while F1 (Transparency & Accountability) and F2 (AI Regulation) exert the strongest influence over other factors, consistent with the indirect influence–dependence analysis.

Table 3. Trends Analysis

Dimension	International Trends	Local/Iranian Trends	Comparative Notes
Social	Declining public trust in traditional news media; growth of news consumption on social media (especially TikTok and Instagram); the emergence of “news fatigue” and active news avoidance; increased sensitivity to AI-generated content labeling.	Generational trust gap (younger audiences gravitate more toward social media and domestic/foreign messaging apps); reliance on national television as the primary news source for older generations; rapid growth of Telegram channels and Instagram pages as alternatives to official media.	Globally, platforms dominate; in Iran, the duality between national television and social media is more pronounced.
Technological	The development of generative language models (ChatGPT, Gemini, etc.); the use of AI in news production and distribution; content authentication tools (C2PA); the threat of deepfakes.	Initial focus on domestic Persian language-processing tools; challenges in accessing computing infrastructure and GPUs; expansion of news bots on Telegram; sensitivity to fake news on WhatsApp and Instagram; the absence of national standards for content verification.	Global: rapid innovation and platform competition. Iran: focus on the Persian language, limited access to advanced technologies, and weaknesses in official fact-checking systems.
Economic	Crisis in media revenue models; reliance on targeted advertising; development of subscriptions and micropayments; venture capital investment in AI startups.	official media’s heavy reliance on government funding; lack of sustainable revenue models in private media; limitations on targeted digital advertising (due to regulations and filtering); growth of informal advertising on Instagram and Telegram.	International: financial crisis accompanied by revenue innovation. Iran: lack of financial diversity and reliance on government funding or informal advertising markets.

Dimension	International Trends	Local/Iranian Trends	Comparative Notes
Environmental/Infrastructural	Expansion of cloud and 5G infrastructure; international content authentication standards (C2PA); growing concerns about energy consumption and the sustainability of large models.	Inequality in access to high-speed internet; broadband limitations and filtering; absence of a national system for content authenticity labeling; challenges in digital archiving of media; reliance on foreign infrastructure for AI tools.	Global: focus on sustainability and data security. Iran: focus on access and infrastructural limitations.
Political/Governmental	Adoption of AI regulations (e.g., the EU AI Act); policies on algorithmic transparency and accountability; professional reports (AP, BBC, EBU) on AI use in newsrooms; emphasis on audience rights.	Emphasis on “trust-building” in national media and top-down policymaking; absence of clear regulations on AI-generated content; security-oriented approach to media and social platforms; efforts to legislate against fake news and control content.	International: regulation for transparency and audience rights. Iran: regulation for content control and management with a focus on national security.

This table was designed to compare international and local/Iranian trends in the domain of media trust and artificial intelligence applications. At the global level, transformations revolve more around technological innovation, the trust crisis, and new media financial models. Media worldwide face phenomena such as declining public trust, rapid growth of artificial intelligence content production, the emergence of deepfakes, and pressure for transparency in AI use. In Iran, however, conditions are very different, and many trends are tied to structural, economic, and governance limitations. Media trust in the country is shaped by the duality between national media and social networks, and the generational gap in news consumption patterns is very prominent. From a technological perspective, although movements have been made in Persian language processing, limited access to computational infrastructures and the absence of national standards for content validation have hindered progress. Additionally, severe dependence on the government budget and weakness of sustainable revenue models for media, along with policymaking that is more security-oriented than transparency-oriented, distinguish the future of media trust in Iran from the global path. In the following, alternative scenarios for public trust were designed.

5.1. Future Scenarios of Public Trust in Media in the Era of Artificial Intelligence

Based on the analysis of drivers, uncertainties, and results obtained from the expert panel, two critical axes, namely the level of media transparency and accountability and the intensity and quality of artificial intelligence regulation, were selected as the main dimensions of the scenario writing matrix. The combination of these two axes outlines four alternative scenarios for the future of public trust in Iran with a time horizon of year 1415 SH, each providing a different image of the relationship between media, technology, and society.

Scenario One: Intelligent Trust (High Transparency and Efficient Regulation)

In this scenario, media actively utilize artificial intelligence technologies to enhance quality, speed, and accuracy in content production, but at the same time commit to disclosing processes and sources used. Laws and regulations are also designed in a way that, on the one hand, prevent monopoly and data misuse, and on the other hand, do not limit media innovation and creativity. In such conditions, public trust is not only preserved but strengthened due to the combination of "technological efficiency" and "institutional transparency." Indicators such as an increased rate of news consumption from credible domestic media, enhancement of media literacy, and reduction of fake news impact are signs of realizing this future.

Scenario Two: Total Distrust (Low Transparency and Weak Regulation)

This scenario represents the most critical possible situation. In conditions where transparency does not exist, and artificial intelligence regulation is inefficient or minimal, media and platforms move toward excessive use of technology for mass content production without accountability. The result is the spread of fake news, prevalence of deepfakes, and formation of a media legitimacy crisis. Public trust sharply decreases, and part of society even becomes deeply distrustful of the entire media system. Signs of this scenario include increased reliance on unofficial and foreign sources, growth of anti-media movements, and weakening of social capital.

Scenario Three: Islands of Trust (Relative Transparency and Insufficient Regulation)

In this scenario, some media organizations or platforms implement partial transparency and accountability policies, but the absence of comprehensive regulation prevents public trust from being rebuilt uniformly. As a result, society moves toward fragmentation of trust; one group becomes loyal to specific domestic or foreign media, while other groups remain in distrust or indifference. This scenario leads to the emergence of "islands of trust," which, although preventing total collapse of trust, intensifies social and informational gaps. Indicators of this situation include increased audience polarization and high diversity of news consumption sources without a common credible reference.

Scenario Four: Imposed Trust (Strict Regulation and Minimal Transparency)

This scenario describes a situation in which government or supervisory institutions control information flow through monopolistic and restrictive policies. Transparency remains at a minimal level, and audiences, due to limited alternatives or restricted access, are compelled toward apparent trust in official media. This type of trust is superficial and fragile and is based more on structural control than on media legitimacy. Signs of this scenario include reduced media diversity, power concentration in monopolistic media, and audiences' forced dependency on limited sources.

Overall, these four scenarios draw an image of possible futures of public trust in media in Iran. Intelligent trust is the desirable and sustainable scenario achieved through the combination of transparency and effective regulation. In contrast, total distrust is a serious threat to social capital and national cohesion. The islands of trust and imposed trust scenarios are also intermediate scenarios that, although preventing complete collapse, have consequences such as social gaps or superficial legitimacy. Therefore, choosing the future path depends on today's decisions in the domains of policymaking, public education, and the media's approach to new technologies. Supplementary findings from the validation workshop showed that experts evaluated the realization of the first scenario as the most desirable and the second scenario as the most probable in Iran's current conditions. This indicates that the gap between the current situation and the desirable future, especially in the domain of transparency policymaking and enhancing media literacy, is substantial. Overall, the research results showed

that the future of public trust in Iran depends more than anything on how technology and media governance interact, and any neglect of these domains can cause the deepening of the trust crisis in the coming decade.

Scenarios Policy Roadmap

The policy roadmap of this research, based on the four scenarios, shows that moving from undesirable scenarios toward the desirable situation, namely "intelligent trust," requires phased, coordinated, and multilevel actions. In the desirable scenario, media must move toward responsible integration of artificial intelligence along with maximum transparency and efficient regulation. This is achieved in the short term by developing a content production transparency charter, creating joint fact-checking units, and intensive training of journalists and editors in ethics and algorithmic bias. In the medium term, national labeling standards and supportive funds for small and local media must be institutionalized, and in the long term, the development of open news data infrastructure and annual algorithm audits by independent institutions can create sustainable trust. In contrast, the "total distrust" scenario is the most critical situation that requires urgent actions to contain the crisis. In the short term, the use of high-risk artificial intelligence modules, especially in sensitive news, should be suspended, and fake news prevention teams must be deployed in key events. Immediate transparency campaigns and regular reporting on errors can provide a basis for rebuilding credibility. In the medium term, reengineering of editorial processes and mandatory human review is necessary, and in the long term, creating specialized arbitration boards and national algorithmic risk assessment standards will contribute to policymaking stability.

In the "islands of trust" scenario, which indicates fragmentation of trust and part of society's loyalty to specific media, the main strategy must focus on converging scattered trusts. In the short term, developing an inter-media cooperation code and launching joint verification hubs is necessary; in the medium term, moving toward label standardization and mutual audits, and in the long term, forming a national media trust index coalition and developing a data transparency ecosystem can cause trust rebuilding at the macro level.

The "imposed trust" scenario also indicates a situation in which monopolistic media control leads to superficial trust. To transform this

apparent trust into sustainable trust, in the short term, a minimum level of transparency, such as periodic news sources reports, and creating independent complaint handling committees must be provided. In the medium term, gradual reduction of distribution monopoly and creating algorithmic accountability frameworks is necessary, and in the long term, safeguarding information freedom, sustainable diversification of media actors, and independent trust rankings can change the path of this scenario toward intelligent trust.

The key recommendation of this roadmap is that transitioning from undesirable scenarios toward intelligent trust requires linking actions in four domains: governance through adopting transparency standards and independent audits, infrastructure by supporting small media and developing fact-check hubs, human capacity by training and empowering newsrooms in ethics and technology, and audience level by layered media literacy enhancement and creating public transparency dashboards. Only by combining these actions can the path to a sustainable and trust-based future for media in Iran be guaranteed.

6. Conclusion

The findings of this research showed that public trust in media in the era of artificial intelligence is a multidimensional phenomenon influenced by the interaction among technology, policymaking, and audience behavior. Futures analysis based on two critical uncertainties—namely, the level of media transparency and the quality of artificial intelligence regulation—drew four alternative scenarios for the horizon year of 1415 SH: "intelligent trust," "total distrust," "islands of trust," and "imposed trust." Among these scenarios, intelligent trust is considered the most desirable scenario in which media, while utilizing artificial intelligence capacities to enhance content production quality and speed, operate with complete transparency in processes and sources and within an effective regulatory framework. In contrast, total distrust is the most critical possible future which, in the absence of transparency and efficient regulation, leads to the erosion of media legitimacy and weakening of social capital. The islands and imposed trust scenarios, although preventing complete trust collapse, will lead to strengthening social gaps or creating superficial trust.

Based on this, the proposed policy roadmap of the research emphasizes that moving toward the desirable scenario requires a set of coordinated actions at four levels of governance: infrastructural, institutional, and audience. At the governance level, developing national standards for content labeling and annual algorithm audits with independent institutional supervision is among the fundamental necessities. At the infrastructural level, creating supportive funds for small media and launching joint verification hubs can help reduce costs and increase transparency. In the institutional dimension, empowering newsrooms through educational programs in technological skills and professional ethics is necessary so that journalists can redefine their role in the era of artificial intelligence. Finally, at the audience level, enhancing media literacy in schools, universities, and general public, along with creating transparency dashboards, can rebuild trust based on participation and awareness.

From a stakeholder perspective, the proposed scenarios imply differentiated policy priorities: regulators are primarily responsible for establishing transparent and enforceable AI governance frameworks; public service media must invest in institutional transparency, professional training, and accountable AI adoption; and private platforms are expected to operationalize algorithmic accountability, content provenance, and user-facing transparency tools. Aligning these actor-specific responsibilities is essential for steering the media ecosystem toward the intelligent trust scenario.

Overall, this research emphasizes that the future of public trust in media is not the product of a linear trend but the result of today's decisions and policies. The choice between a future based on intelligent trust or trapped in a state of total distrust depends on the commitment of media organizations and policymakers in pursuing transparency, accountability, and responsible regulation. Only through these coordinated actions can the path to a sustainable and legitimate future for media in Iran be secured.

Appendix A: Review of Previous Studies

Researcher(s)	Year	Research Field	Research Method	Research Findings	Research Limits
Safavi and colleagues	2023/2024	Science journalism and new media	Combined method (IS-M, PLS)	Model of science journalism with emphasis on trust and infrastructure	Limited to the field of science; lacks a macro-level perspective on public trust
Ejlali and Khatibi	2024/2025	Cognitive warfare and algorithms	Case-analytical	The role of algorithms in filter bubbles and the threat to democracy	Lacks futures studies and scenario writing
Mohammadnezhad and Shahmohammadi	2021	Trust in the National Media	Qualitative-Quantitative, SWOT, QSPM	Strategies for Enhancing Trust in IRIB News	Limited to the National Media; Lack of Attention to social media
Taghipour et al.	2025	Artificial Intelligence in Media	Futures Studies (Delphi Method)	Scenarios of Media Transformation and Challenges	Focus on Technology; Neglect of Public Trust
Almohammad and Asadi	2025	The Future of Journalism and Artificial Intelligence	Futures Studies (Pillkan Model)	Constants, Trends, and Uncertainties	Lack of Analysis of Trust in the Media
Wakari and Chadwick	2020	Deepfakes and Trust	Experimental Study	Cause Cynicism and a Crisis of Trust	Focus on the West; Limited Generalizability to Iran
Opdahl et al.	2023	Trustworthy Journalism with AI	Conceptual Article	AI Enhances Quality and Trust	No Future Scenario Building Included
Pena Fernández et al.	2023	Generative AI and the Social Dimension of Media	Systematic Review	Threat to Journalistic Independence	Lack of Futures Studies Approach
Tuff and Simon	2024	AI Labeling in News	Survey-Experiment	Labeling Reduces Trust; Source Transparency Has Positive Effects	Limited to the United States
Niazi et al.	2023	Trust in AI in Asia	Multi-country Survey	Information Seeking and Variable Trust	Cultural-Regional Limitations
Karaslan et al.	2024	Public Attitudes toward AI-Generated News	Survey	Moderate Awareness and Concerns about Trust	No Policy Analysis Included
Kim et al.	2024	AI and Blockchain for Fake News	Algorithm and Experiment	Improved Fake News Detection	Purely Technological Perspective
Jungherr and Schroeder	2023	AI and the Public Arena	Conceptual Analysis	Changes in the Function of the Public Arena	Lack of Empirical Data
	2023	AI, Public Trust, and Governance	Survey	Role of Trust in AI Governance	Focus on the United States

Public trust in media in the age of AI

Appendix B: Composition of the Expert Panel

Field	Count	key selection criteria
Media managers / editors	6	At least 10 years of newsroom/editorial leadership experience, familiarity with content policy-making
Universities (media, communication, futures studies)	5	Track record of publications and projects on media trust assessment and media policy
Digital rights	1	Experience in media literacy initiatives and error-monitoring projects
AI specialists	3	Expertise in AI-driven approaches to media analysis and decision-making

Appendix C: STEEP Framework

	Social	Technological	Economic	Environmental/Infrastructural	Political/Governmental
Drivers	The shift in news consumption patterns toward mobile devices and social media; the rise of messaging apps and closed digital spaces; news fatigue and active avoidance of news; generational gaps in trust; demands for transparency and disclosure of AI use; the role of influencers and citizen journalism; heightened public sensitivity to privacy and data; the expansion of media and data literacy; personalization and recommender systems; audience migration across platforms.	Generative and multimodal language models; synthetic content production (text/audio/image/video); authenticity verification tools (C2PA/watermarking); ranking and recommender algorithms; automated news agents and newsroom robots; fact-checking automation; fact-checking platforms; cybersecurity and XAI tools; data governance (federated/privacy-preserving).	Cost-cutting pressures in newsrooms; productivity gains from automation; revenue diversification (subscriptions, micropayments, events/B2B services); the market for AI tools in the media sector; changes in the advertising and targeting market; venture capital investment in generative technologies; the costs of fact-checking and verification; opportunities for product innovation.	Cloud capacity and data-center scalability; advances in GPUs/TPUs and the semiconductor supply chain; broadband/5G expansion; content authentication standards (C2PA); digital archiving and training data; resilience and business continuity; energy optimization of models; secure content distribution and CDNs.	AI regulations (the EU AI Act and similar frameworks), the DSA/DMA, and the GDPR; newsroom AI policies (AP/BBC/EBU); transparency requirements and labeling of generative content; election integrity policies; content authenticity initiatives (CAI/C2PA); the role of media regulatory bodies; accountability and responsibility frameworks.
Uncertainties	Audience responses to "AI-generated" labeling; the impact of AI disclosure on trust (increase or decrease); levels of acceptance of automation across different groups; the persistence or intensification of polarization and filter bubbles; the effectiveness of media-literacy programs; sudden audience shifts across platforms; attribution and responsibility norms; changes in trust across generations and cultures.	The outcome of the "arms race" between deepfake generation and detection; the robustness of watermarking and content provenance systems; error/hallucination and quality assurance in LLMs; the degree of dependence on a few major providers (concentration); the status of intellectual property in training data; the future of open-source versus closed models; data provenance traceability; the safety of autonomous agents and the limits of their autonomy.	The durability of the subscription economy and direct payment models; the revenue share of publishers versus platforms; the impact of privacy regulations on advertising; energy and computing costs for small newsrooms; the return on investment of AI projects; employment implications and emerging skill requirements; the financial sustainability of local and investigative media; legal risks related to offenses/defamation stemming from AI errors.	Energy/carbon constraints and environmental policies; disruptions in the semiconductor supply chain and sanctions; the computing-access gap between countries and organizations; interoperability of authentication standards; the vulnerability of infrastructures to attacks and disruptions; the sustainability of long-term storage; ownership and access to media archives; the balance between edge and cloud processing.	The approach and strictness of regulation enforcement (scope/exemptions); legal liability assignment for generative content; international coordination and conflicts of law; the impact of governance on innovation and freedom of expression; researcher access to platform data; platform policies during election periods; data governance and cross-sector sharing; operational algorithmic auditing and transparency frameworks.

References

- Ajlali, M. M., & Khatibi, A. (1403 SH). The role of social networks in cognitive warfare: The impact of algorithms on public opinion. *Journal of Political Studies in Cognition*, 1(3), 121–139.
- Al-Mohammad, H., & Asadi, A. (1403 SH). The future of journalism and the impact of artificial intelligence on media content. *Journal of Virtual Space Studies and Social Media*, 1(2), 73–103.
- Brewer, P. R., Bingaman, J., Dawson, W., Paintsil, A., & Wilson, D. C. (2022). Eyes on the streets: Media use and public opinion about facial recognition technology. *Bulletin of Science, Technology & Society*, 42(4), 133–143.
- Coleman, J. S. (1990). *Foundations of social theory*. Harvard University Press.
- Corsi, G., Marino, B., & Wong, W. (2024). The spread of synthetic media on X. *Harvard Kennedy School Misinformation Review*, 5(3), 1–19.
- Druckman, J. N., & Bolsen, T. (2011). Framing, motivated reasoning, and opinions about emergent technologies. *Journal of Communication*, 61(4), 659–688.
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681–694.
- Jungherr, A., & Schroeder, R. (2023). Artificial intelligence and the public arena. *Communication Theory*, 33(2–3), 164–173.
- Karaaslan, İ. A., Baha Ahmet, Y., & Yağmur, K. (2024). Investigation of the awareness of automated news in terms of public opinion: Artificial intelligence journalism. *Evolutionary Studies in Imaginative Culture*, 8(7), 1658–1671.
- Khodabin, M., Sharifi Poor Bgheshmi, M. S., Piriyaee, F. and Zibaei, F. (2022). Mapping the Landscape of AI Literacy: An Integrative Review. *Socio-Spatial Studies*, 6(1), 51-61. doi: 10.22034/soc.2022.223715
- Kim, S. K., Huh, J. H., & Kim, B. G. (2024). Artificial intelligence blockchain-based fake news discrimination. *IEEE Access*, 12, 53838–53854.
- Mohammadnejad, A., & Shahmohammadi, M. (1400 SH). Strategies to enhance public trust in news and political programs of national media. *Journal of Political Studies*, 14(53), 21–49.
- Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Nielsen, R. K. (2023). *Reuters Institute digital news report 2023*. Reuters Institute for the Study of Journalism.
- Neyazi, T. A., Khai Ee, T., Nadaf, A., & Schroeder, R. (2025). The effect of information-seeking behaviour on trust in AI in Asia: The moderating role of misinformation concern. *New Media & Society*, 27(4), 2414–2433.
- Opdahl, A. L., Tessem, B., Dang-Nguyen, D. T., Motta, E., Setty, V., Throndsen, E., ... Trattner, C. (2023). Trustworthy journalism through AI. *Data & Knowledge Engineering*, 146, 102182.

- Peña Fernández, S., Meso Ayerdi, K., Larrondo Ureta, A., & Díaz Noci, J. (2023). Without journalists, there is no journalism: The social dimension of generative artificial intelligence in the media. *Profesional de la Información*, 32(2).
- Robles, P., & Mallinson, D. J. (2025). Artificial intelligence technology, public trust, and effective governance. *Review of Policy Research*, 42(1), 11–28.
- Sabbar, S., & Habib Zadeh Khiyaban, S. (2023). Algorithms of displacement: Emotional and rhetorical responses to ai-driven job loss in digital public discourse. *International Journal of Advanced Multidisciplinary Research and Studies*, 3(4), 1324–1331.
- Safavi, B., Tajik Esmaeili, S., Ghadimi, A., & Niroomand, L. (1402 SH). Design and validation of science journalism model in the era of new media. *Journal of Interdisciplinary Studies in Communication and Media*, 5(4), 135–164.
- Salehi, K., Habib Zadeh Khiyaban, S. and Sabbar, S. (2025). Artificial Intelligence and the Future of International Law and Power. *Journal of World Sociopolitical Studies*, 9(4), 923–958. <https://doi.org/10.22059/wsps.2025.401951.1552>
- Schiff, D. S., Schiff, K. J., & Pierson, P. (2021). Assessing public value failure in government adoption of artificial intelligence. *Public Administration*, 100(3), 653–673.
- Soltanpour, A, Esmaeilzadeh Ghandehari , M. Fahim Devin, H. (2024). Conceptual Framework of the Application of Modern Technology in Media (Case Study of Artificial Intelligence in Sports Journalism), *Communication Management in Sports Media*, 12(45), 117–134.
- Taghipour, F., Yektashi, V., & Janghorban, A. (1404 SH). Futures studies of artificial intelligence applications in media industries. *Journal of Futures Studies of the Islamic Revolution*, 6(1), 249–285.
- Toff, B., & Simon, F. M. (2024). “Or they could just not use it?”: The dilemma of AI disclosure for audience trust in news. *The International Journal of Press/Politics*. Advance online publication.
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1), 2056305120903408
- Veen, M., Gremmen, B., te Molder, H., & van Woerkum, C. (2011). Emergent technologies against the background of everyday life: Discursive psychology as a technology assessment tool. *Public Understanding of Science*, 20(6), 810–825.