ISSN: 2716-9189
Journal of Studies in Language, Culture, and Society (JSLCS).
Received: 08/07/2025 Accepted: 05/11/2025

E-ISSN: 2676-1750
(08) 04, 2025. (PP . 279-297)
Published: 31/12/2025

LANGUAGE AND ARTIFICIAL INTELLIGENCE IN ARAB SOCIETIES: A SOCIOLINGUISTIC APPROACH TO IDENTITY AND DIGITAL HEGEMONY

Nadjat Bencherrat [1] **Yamina Baghor** [2]
[1]**University of Abou Bekr Belkaid Tlemcen (Algeria)**
nadjat.bencherrat@univ-tlemcen.dz
[2]**University of Abou Bekr Belkaid Tlemcen (Algeria)**
yamina.baghor@univ-tlemcen.dz

Abstract: This study investigates the complex relationship between language and artificial intelligence (AI) in Arab societies from a sociolinguistic perspective. The research aims to analyse how Modern Standard Arabic (MSA) and local dialects are represented in AI-based applications and to assess the implications of such representations for identity construction and symbolic power in digital environments. A qualitative research design was adopted, combining critical discourse analysis of AI-generated Arabic texts with semi-structured interviews conducted with fifteen participants from six Arab countries. Content monitoring of digital discussions on social media further complemented the data collection. Findings demonstrate that AI systems consistently privilege MSA while marginalizing dialectal diversity, thereby reinforcing pre-existing linguistic hierarchies and reproducing symbolic domination. Participants frequently reported difficulties in using dialects with AI tools and noted that algorithmic interactions often reflect Western-centric frameworks, projecting an elitist image of Arab identity that fails to capture everyday linguistic practices. Such patterns confirm that AI technologies do not operate as neutral instruments but function as sociolinguistic agents that shape identity performance, regulate acceptable discourse, and contribute to what can be described as “digital linguistic hegemony.” The study contributes to emerging debates on digital colonialism and algorithmic bias by situating AI within the broader sociolinguistic realities of the Arab world. It highlights the urgent need for inclusive policies that integrate dialectal diversity into AI training data, the involvement of Arab scholars and institutions in algorithm development, and the promotion of culturally sensitive design. The research concludes by recommending practical steps toward achieving linguistic equity and epistemic sovereignty in the age of intelligent technologies.

Keywords**:** Algorithmic bias, Arabic language, artificial intelligence, digital hegemony, linguistic identity, sociolinguistic, digital colonialism

How to cite the article:

Bencherrat, N., & Baghor, Y. (2025). Language and artificial intelligence in Arab societies: A sociolinguistic approach to identity and digital hegemony. *Journal of Studies in Language, Culture, and Society (JSLCS)8*(4), pp. 279-297.

[1] **Corresponding author** : Nadjat Bencherrat, **Authors’ ORCID ID:** https://orcid.org/0000-0003-0373-6892

1. Introduction

In recent decades, the world has witnessed a profound transformation in patterns of communication and knowledge, driven by the rapid development of artificial intelligence (AI), particularly in the areas of natural language processing and discourse generation. These shifts have reshaped the relationship between language and society: language is no longer the exclusive domain of humans, it is now also produced and utilized by intelligent algorithms that play an increasing role in our daily lives, from machine translation to digital content creation to interactive conversations.

Within the Arab context, fundamental questions emerge regarding how Arabic and its dialects are represented in these intelligent systems, and how such representations impact identity formation and the reproduction of symbolic dominance in digital spaces.

This topic intersects with sociolinguistics, which examines the relationship between language, social context, power, and identity. From this perspective, AI is not merely a technological phenomenon but a new field where linguistic and cultural hegemony is exercised, and where representations of the self and the other are reshaped. Accordingly, there is a pressing need to critically assess the impact of these technologies on Arab societies, particularly in light of the epistemological and linguistic imbalance between the centre (the technologically dominant West) and the periphery (the Arab world, which largely remains a consumer rather than a producer of these technologies).

1.1. Research Objectives:

This study seeks to achieve the following objectives:

1. To analyse the representation of Arabic and its dialects in AI language applications (such as chat-bots, machine translation, and text generation tools).
2. To explore the impact of AI on linguistic identity in Arab societies, especially in light of cultural and linguistic biases embedded in digital technologies.
3. To identify the manifestations of algorithmic linguistic hegemony and assess how they affect the Arab linguistic landscape.
4. To offer a critical sociolinguistic perspective on the relationship between technological power and the reproduction of linguistic inequalities in the Arab digital context.

1.2. Research Questions

This study seeks to answer the following key questions:

1. How are Arabic and its dialects represented in contemporary AI systems?
2. What is the impact of these representations on the linguistic identity of Arab users?
3. To what extent do AI algorithms reinforce or reproduce linguistic and cultural hegemony in Arab societies?
4. What practices and recommendations can promote linguistic justice in AI models operating within the Arab context?

2. Theoretical Background

This study lies at the intersection of three major intellectual domains: sociolinguistics, artificial intelligence studies, and theories of cultural hegemony. Each of these fields has undergone significant development in recent years, particularly with the growing influence of AI in the production and consumption of language.

Sociolinguistics approaches language as more than a tool for communication; it is seen as a means of constructing social reality and reinforcing cultural identities. Pioneering scholars such as Mikhail Bakhtin (1981), Pierre Bourdieu (1991), and Erving Goffman (1959) have shaped a perspective in which language is intrinsically tied to power and social positioning. While numerous studies have examined linguistic inequalities and their role in producing social differentiation, much of this work has been confined to traditional contexts (e.g., education, media, and class structure), without sufficiently extending into the realm of smart technologies.

Over the past decade, AI-driven language applications, such as Large Language Models (LLMs), have emerged with the capacity to generate texts that closely resemble human discourse. However, extensive research including that of Emily Bender and Timnit Gebru has demonstrated that these models are not neutral. Instead, they reflect linguistic and cultural biases inherited from the datasets on which they are trained. According to this literature, languages with strong digital representation (such as English) are disproportionately privileged, while languages from the Global South including Arabic are either marginalized, reduced to stereotypes, or represented only in standardized or formal forms.

Recent scholarship has increasingly emphasized that language technologies are never neutral but rather reproduce existing hierarchies. As Blodgett et al. (2020) argue, “language technologies are not neutral; they reflect and reproduce existing power relations, often reinforcing marginalization through biased data and modelling choices” (p. 5455). This resonates with the challenges faced by Arabic NLP, where dialectal varieties remain underrepresented and computationally marginalized In the Arab context, such exclusion reflects not merely technical gaps but deeper sociolinguistic dynamics. As Haouchi (2024) observes, “automated semantic structuring reflects the computer’s logic rather than the author’s intent, highlighting the limits of relying on machine-driven meaning-making” (p. 200). Taken together, these studies highlight that Arabic language use in AI systems is constrained both by structural limitations of computational models and by global hierarchies of linguistic power.

This study draws upon Gramsci's theory of cultural hegemony, which posits that power is not exercised solely through coercion, but also through control over systems of meaning and knowledge. From this perspective, AI systems function as tools for reproducing dominant worldviews, reinforcing specific linguistic and cultural models as the "norm" or "reference standard."

Additionally, the study is grounded in postcolonial theory, particularly through concepts such as "digital colonialism", which suggests that linguistic dominance in AI may replicate historical power relations between the centre (the West) and the periphery (the Arab world). Recent research in internet sociology also highlights how digital spaces have become critical arenas for identity formation. The use of local dialects and cultural symbols has played a central role in reshaping collective identity across communities.

However, the incorporation of AI into these digital spaces introduces a new set of challenges, including:

- Who controls the algorithms that generate language?
- What kind of representation does the Arab user receive?
- Do these tools enrich local identities, or simply reshape them into pre-packaged, imported molds?

3. Conceptual Framework

The key concepts embedded in the title of this study language, artificial intelligence, Arab societies, identity, and digital hegemony serve as operational gateways for understanding the subject and unpacking its social and cultural dimensions. This section provides a precise conceptual delineation of each term within the current sociolinguistic context.

3.1. Language

From this perspective, Pierre Bourdieu regards language as a form of “symbolic capital”, utilized to reproduce power relations and structures of dominance within society. He states: “Linguistic practices are investments in a symbolic marketplace, whose value is determined by the rules of the game.” Bourdieu (1991, p. 37) This view is echoed in Arab intellectual traditions, most notably by Taha Abderrahmane, who considers language a foundational element in shaping civilizational selfhood. For Abderrahmane, language is intrinsically bound to values, identity, and ethics, rather than being a mere technical instrument (Abderrahmane, 2011, p. 91). In this sense, Artificial Intelligence (AI)-a subfield of computer science concerned with simulating human cognitive capacities, including natural language processing-functions today as a cognitive agent reshaping how language is both produced and consumed. However, as(Bender et al., 2021, p. 610) assert, these models “do not generate innocent discourse; they perpetuate patterns of cultural bias inherited from unbalanced training data.”

Several Arabic studies have discussed the linguistic and cultural implications of artificial intelligence in Arab contexts. These studies emphasize that AI technologies often reproduce linguistic hierarchies and cultural biases, underscoring the importance of developing models that account for Arabic’s sociolinguistic diversity.

3.2. Artificial Intelligence (AI)

In the context of Arabic, the integration of artificial intelligence has brought both remarkable progress and persistent challenges. Seyidov (2024) provides an updated overview of the current status of Arabic natural language processing (NLP), emphasizing the dual nature of this evolution: on one hand, rapid advancements in computational linguistics and machine translation; on the other, enduring difficulties related to dialectal variation, orthographic ambiguity, and limited annotated datasets. His study underscores the urgent need for collaborative efforts among Arab researchers and technology developers to create inclusive AI models that reflect the sociolinguistic richness of Arabic rather than perpetuating a monolithic linguistic norm.

3.3. Arab Societies

Abdelfattah Kilito (2009, p. 72) emphasizes that understanding Arabic cannot be divorced from its “historically layered multiplicity of registers and voices,” making digital representation of such societies a complex and nuanced undertaking. In the sociolinguistic context, identity is not a static essence but rather a continuous process of construction, shaped and reshaped through social discourse. Stuart Hall (1996, p. 4) conceptualizes identity as “a site of ongoing negotiation between the personal and the collective, constituted within and through language.” In the digital sphere, identity is enacted through linguistic expression, and algorithmic systems shape what can be said and how it may be said, posing challenges around the performance of identity under subtle forms of digital surveillance and regulation.

The dynamic interaction between users and AI-generated language reflects what Canagarajah (2013) conceptualizes as *translingual practice*-a process in which speakers draw on multiple linguistic resources to negotiate meaning and identity in globalized communication spaces. Within digital contexts, this practice becomes even more complex as AI systems tend to standardize or "normalize" linguistic inputs, often suppressing the fluid, hybrid, and creative ways in which users combine languages. The tension between users' translingual agency and algorithmic standardization highlights a central paradox: while technology enables unprecedented forms of linguistic mobility, it simultaneously enforces new boundaries that constrain authentic multilingual expression.

This paradox forms the basis of what may be termed "digital hegemony," where algorithmic systems not only structure communication but also reproduce symbolic hierarchies that privilege dominant linguistic norms.

3.4. Digital Hegemony

The term "digital hegemony" is derived from Antonio Gramsci's (1971) theory of cultural hegemony, which posits that power is exercised not through coercion but through the imposition of socially accepted norms of meaning and thought. In digital contexts, this hegemony is embedded in the design of algorithms that privilege the languages and cultures of the centre, often at the expense of marginalized or peripheral voices. Shoshana Zuboff (2019, p. 145) warns against this trend, arguing that "digital platforms do not merely collect data; they restructure social reality in accordance with the interests of dominant tech corporations."

The digital mediation of language must also be situated within broader debates on power and colonialism. Couldry and Mejias (2019) argue that "the extraction of human life through data is the latest form of colonial appropriation, turning social existence into raw material for capitalist exploitation" (p. 2). Within this framework, algorithmic preferences for Modern Standard Arabic can be read as a form of digital linguistic colonialism, where local dialects are erased in favor of standardized forms that fit global computational logics. This perspective aligns with Braimoh's (2024) conceptualization of texting language as a "digital symbolic current" (p. 207) that regulates communicative norms reinforcing the idea that digital infrastructures are sites of symbolic domination.

Building on this, Couldry and Mejias (2019) conceptualize data colonialism as the appropriation of human life through extraction and commodification. Milan and Treré (2022) extend this critique by arguing that "data universalism reproduces colonial logics by erasing the epistemologies and lived experiences of the Global South, subordinating them to dominant computational frameworks" (p. 243). This perspective is highly relevant to the Arab sociolinguistic context, where dialectal diversity is rendered invisible within AI systems designed on globalized linguistic standards. By situating the Arab case within this broader critique of digital colonialism, the analysis highlights how algorithmic structures not only marginalize local linguistic practices but also perpetuate asymmetrical power relations in knowledge production.

3.5. Previous Studies

The intersection between language and artificial intelligence remains a nascent topic within sociolinguistic literature. Arabic, in particular, has not yet occupied a central position in international discourse surrounding AI whether in terms of linguistic representation or its broader social impact. Nevertheless, several scholarly contributions, both in Arabic and international academia, offer valuable insights that can inform the critical backdrop of this study.

One of the most influential works is the pioneering paper by Bender et al. (2021) titled "On the Dangers of Stochastic Parrots". This study addresses the ethical and linguistic risks posed by large language models (LLMs), arguing that such systems reinforce existing linguistic and cultural biases rooted in the imbalanced data on which they are trained. Consequently, they contribute to the reproduction of linguistic hierarchies and symbolic domination (Bender, 2021, pp. 610-615) . However, the study does not delve into the cultural and identity-based dimensions of language, particularly in Global South contexts such as the Arab world.

In her seminal book The Age of Surveillance Capitalism, Shoshana Zuboff explores the mechanisms of control exercised by AI-driven platforms. She contends that these systems "do not merely collect data, but rather reconfigure social reality in alignment with the interests of dominant tech corporations" (Zuboff, 2019, p. 145) . While this framework is pivotal for understanding how linguistic behaviour is reshaped within digital spaces, the book does not engage with language or identity from a sociolinguistic perspective.

Similarly, Vincent W. J. van Gerven Oei in his critical study "AI and the Coloniality of Language", highlights how AI perpetuates a colonial model of language, wherein non-Western languages are reduced and framed as "technical problems" rather than living cultural structures (van Gerven Oei, 2020, p. 27) . However, his analysis focuses on sub-Saharan Africa and does not explicitly address the Arab region.

Several Arabic-language studies have also examined the relationship between AI and language from different angles. For instance, (Hanandeh et al., 2024) explored the role of Artificial Intelligence in the Arabic linguistic landscape, emphasizing both the opportunities and challenges it presents. Their findings highlight the scarcity of Arabic linguistic resources compared to English and the need for culturally informed AI models that can reflect the linguistic and social diversity of the Arab world.

In addition to the sociolinguistic and ideological implications of artificial intelligence, several studies have also examined its pedagogical and linguistic applications in Arab educational contexts. Recent research has explored university students' perceptions of integrating AI tools in English language learning. Findings reveal that while students often view AI as an innovative and supportive resource for language practice, they also express concerns regarding accuracy, cultural representation, and over-reliance on machine-generated input. These insights underscore the importance of developing AI applications that are linguistically inclusive and culturally sensitive, ensuring that language technologies contribute to educational equity rather than reinforcing digital hierarchies.

Recent Arabic-language research continues to highlight the challenges of representing Arabic and its dialects within AI-driven systems. These studies emphasize that algorithmic models often reproduce linguistic hierarchies and marginalize regional varieties, underscoring the need for inclusive Arabic datasets and culturally sensitive design approaches.

The discussion on digital identity among Arab youth has highlighted how online spaces serve as arenas for self-expression and social negotiation. These dynamics align with the broader sociolinguistic findings of this study, which demonstrate the interplay between language, technology, and identity in Arab digital contexts.

Despite the richness of these prior works, a significant research gap remains: the absence of a comprehensive sociolinguistic critique that integrates AI, language, and identity within the Arab context. Most previous studies have focused on technical or pedagogical dimensions, neglecting the complex relationship between digital hegemony and the reconfiguration of linguistic selfhood in Arab societies.

This study thus seeks to bridge that gap by offering a critical sociolinguistic framework to examine the social and political ramifications of artificial intelligence on language and identity in the Arab digital sphere.

4. Methodology

This study adopts a qualitative research design, given the nature of the topic, which is inherently concerned with the analysis of meanings, contexts, and linguistic and social representations. It is grounded in a critical sociolinguistic approach, drawing extensively on the tools of Critical Discourse Analysis (CDA)—particularly as developed by Norman Fairclough (1995) to explore the relationship between language and power within AI systems. This methodological framework enables the identification of how AI-generated digital discourses construct specific identity representations and helps unveil the symbolic biases that permeate what often appears to be neutral language.

4.1. Data Collection Tools

To gather data, the following methods were employed:

- **Textual Analysis** of outputs generated by AI language models (such as ChatGPT, Google Translate, and others) related to Arabic and its dialects, based on user-simulation scenarios that reflect natural interaction with the system.
- **Semi-structured Interviews** with a selected group of Arab users (students, bloggers, translators, and digital content creators) to explore their linguistic experiences with AI tools.
- **Content Monitoring and Analysis** of user-generated discussions on social media platforms (Twitter, YouTube, Reddit, etc.) that address language and identity-related interactions with AI technologies.

4.2. Study Population and Sampling

The study targets Arabic-speaking users of AI technologies, including both linguistic professionals and general users who engage with language-based applications. A purposive sample was selected, consisting of 15 participants from six Arab countries (Egypt, Algeria, Lebanon, Saudi Arabia, Sudan, and Tunisia), ensuring diversity in terms of identity, dialect, and gender. The research adheres to ethical principles of qualitative inquiry, including maintaining participant confidentiality and obtaining informed consent prior to conducting interviews. Participants' names were anonymised, and identifiable data were altered to preserve privacy. Furthermore, the study critically interrogated the ethical implications of engaging with AI systems, refusing to treat them as neutral sources of knowledge.

4.3. Justification of Sampling Strategy

This study employed purposive sampling due to its relevance in exploratory qualitative research. Participants were selected based on their active engagement with Arabic AI applications and their potential to offer diverse, context-rich perspectives on the intersection between language and digital identity. The aim was not statistical generalization, but rather a deeper sociolinguistic insight into specific user experiences across different Arab countries.

4.4. Study Limitations

While this study seeks to offer an in-depth sociolinguistic analysis of the relationship between language and artificial intelligence in the Arab context, it is not without methodological and contextual limitations, which must be acknowledged:

- **Geographical Scope.** The study was limited to a purposive sample from six Arab countries, which does not allow for full representation of the geographic and cultural breadth of the Arab world (e.g., the Arabian Gulf or the Horn of Africa regions were not comprehensively included).
- **Linguistic Scope.** The research focused on representations of Modern Standard Arabic (MSA) and selected dialects (e.g., Egyptian, Algerian, and Levantine), which limits the generalizability of the findings to all Arabic dialects or to other parallel languages spoken by minority communities within the studied countries (such as Amazigh or Kurdish).
- **Technical Limitations.** The analysis was based on content generated by widely accessible open-source AI tools. This may not reflect the performance of more advanced commercial or proprietary systems, particularly paid models that are less commonly used by the general public, potentially affecting the study's findings.
- **Temporal Constraints.** Data collection and interviews were conducted within a specific timeframe (from October 2024 to March 2025). As artificial intelligence technologies are evolving rapidly, this fixed period may not capture emerging developments in real-time.
- **Generalizability.** Given its qualitative nature and reliance on semi-structured interviews, this study does not aim for statistical generalization but rather seeks to offer a deep understanding of the sociolinguistic phenomena under investigation.

5. Results

It should be noted that the percentages reported in this section are employed for illustrative purposes only, with the aim of highlighting general tendencies in participants' responses. Given the qualitative orientation of the study, these figures should not be interpreted as statistical measurements or as representative indicators of the wider population. Their primary function is to support the interpretive analysis and to provide greater clarity regarding the patterns that emerged from the interviews and observations. This approach is consistent with the logic of qualitative inquiry, which privileges in-depth understanding of experiences and discourses over statistical generalization or numerical representativeness.

Table 1

General information of participants

Participant No.	Gender	Country	Category	Level of AI Interaction	Main Dialect Used
1	Male	Algeria	University student	Medium	Algerian Darja
2	Female	Lebanon	Digital blogger	High	Lebanese Dialect
3	Male	Egypt	Freelance translator	High	Egyptian Arabic
4	Female	Saudi Arabia	Content creator	Medium	Gulf Arabic
5	Male	Sudan	University student	Low	Sudanese Dialect
6	Female	Tunisia	Translator	Medium	Tunisian + French
7	Male	Algeria	Tech blogger	High	Algerian Darja
8	Female	Lebanon	Graduate student	High	MSA + Lebanese
9	Male	Egypt	Video content creator	Medium	Egyptian Arabic
10	Female	Saudi Arabia	Linguist researcher	High	MSA + Gulf Arabic
11	Male	Sudan	University student	Low	Sudanese Dialect
12	Female	Tunisia	YouTube content creator	Medium	Tunisian Dialect
13	Male	Algeria	Software developer	Medium	Algerian Darja
14	Female	Egypt	Independent writer	High	Egyptian Arabic
15	Male	Saudi Arabia	Cultural blogger	High	Gulf Arabic

This table illustrates the academic level, cultural background, and gender distribution of participants from various Arab countries.The data show that the sample included three participants from Algeria, three from Egypt, three from Saudi Arabia, two from Lebanon, two from Sudan, and two from Tunisia, making a total of 15 participants. In terms of gender, eight participants were male (53.3%) and seven were female (46.7%).

Participants represented a variety of roles: four university students, three tech or cultural bloggers, three content creators, two translators, one independent writer, one software developer, and one linguistic researcher.

Regarding interaction with AI, 46.7% reported a high level of engagement, 40% indicated a moderate level, and the remaining 13.3% reported low engagement.

The results also showed that female participants were more likely to engage with AI at a high level compared to males. Moreover, all bloggers, researchers, and writers demonstrated 100% high engagement with AI tools. University students were predominantly in the low-engagement category, while content creators were split between moderate and high levels of interaction.

Table 2

Distribution by country and linguistic identity (Dialects)

Country	Dialect(s) Used in Interaction	Number of Participants
Egypt	Egyptian Arabic	3
Algeria	Algerian Darja + Modern Standard Arabic	3
Lebanon	Lebanese + Modern Standard Arabic	2
Saudi Arabia	Gulf Arabic + Modern Standard Arabic	3
Sudan	Sudanese Dialect + Modern Standard Arabic	2
Tunisia	Tunisian + French / Modern Standard Arabic	2

The data reflect genuine diversity within the participant sample; however, it also revealed the presence of intermediary languages-such as French in Tunisia and Algeria, and English in Lebanon which influenced how users interacted with AI systems. All participants used Modern Standard Arabic (MSA) to some extent, though the frequency varied depending on the context and purpose.

In summary, local cultural and linguistic contexts significantly shape how AI language tools are used. Therefore, such cultural and linguistic plurality must be considered when designing and training language models intended for Arab societies.

Table 3

Use Cases of AI Language Tools: Use Cases of AI Language Tools

Use Case	Number of Participants	Percentage
Translation	12	80%
Grammar/Language Correction	9	60%
Text/Content Generation	10	67%
Semantic/Text Analysis	4	27%
Entertainment	6	40%

The results show that translation (80%) is the most common use, reflecting the ongoing need for language-conversion tools. This suggests that most participants operate in multilingual environments (such as Algeria, Tunisia, and Lebanon). It also indicates that AI is viewed primarily as a functional complement to human linguistic skills not a complete substitute. Translation is perceived more as a tool for digital integration than as a purely linguistic task.

Additionally, 60% of participants reported using AI for language correction, a practice that aligns with what Goffman calls "impression management." Users aim to fine-tune their written messages to conform to acceptable linguistic and cultural standards an effort to shape their digital self-presentation. This highlights how AI tools have become integral to everyday editorial practices, with all their strengths and limitations.

About 40% of respondents reported using AI for entertainment, indicating a tendency among younger, digitally native users to explore new possibilities without professional pressure. This recreational use reflects not just technical curiosity, but a new form of cultural coexistence between humans and machines in daily digital spaces where “play” becomes a means of learning and interaction, as described by Castells (2010) in his concept of the “network society.”

Furthermore, 67% stated they use AI for text or content generation, especially among content creators and bloggers. This signals a shift in the nature of writing from a solitary skill to a collaborative process with AI. This phenomenon reflects Bourdieu's concept of "redistribution of cultural capital," where access to digital tools becomes a new factor in shaping one's social efficacy. It highlights the transformation of users from mere content consumers to active producers.

Only 27% reported using AI for semantic or text analysis, suggesting this function remains underutilized, likely due to its complexity or limited accessibility for general users. This type of usage requires deeper computational linguistic knowledge. The figure supports Manuel Castells' notion of the "digital knowledge gap," where basic AI functions are accessible, but more advanced analytical tools remain out of reach for the broader public. This indicates unequal levels of digital empowerment even within the same social group.

In summary, AI is currently used in a practical and straightforward manner, but it has yet to become a widespread tool for in-depth linguistic analysis among most users.

Table 4

AI systems' interaction with Arabic and local dialects: AI systems' interaction with Arabic and its dialects

Observed Interaction Level	**Number of Participants**	**Percentage**
Good with Modern Standard Arabic (MSA)	13	87%
Weak with Local Dialects	10	67%
Biased toward MSA, lacking flexibility	8	53%
Able to understand local dialect	3	20%

Approximately 87% of participants stated that AI tools generally perform well when interacting with Modern Standard Arabic (MSA). This indicates a digital reinforcement of the "formal linguistic norm," where inherited language hierarchies (MSA vs. dialects) are reproduced within digital structures. The result is the promotion of a linguistic model that views MSA as the sole authentic representation of Arab identity while marginalizing the everyday, lived linguistic realities of the users.

67% of respondents found that AI interaction with dialects was relatively weak and subpar. This reflects the "invisibility" of dialects within global digital systems, constituting a form of implicit exclusion of subnational linguistic identities. This aligns with Bourdieu's concept of symbolic domination, where dialects—despite being the language of daily life are denied a place in formal platforms, thus perpetuating a disconnect between the language people live by and the one used digitally.

Over 53% observed that AI tools lack openness in interacting with diverse language forms. This reveals the inability of current algorithms to grasp linguistic diversity and social context. AI here is not neutral; it reproduces a fixed idea of a "model language" written MSA. This bias is not just reflected in outputs but also manifests in the user-tool relationship, forcing users to conform to MSA norms in order to be understood.

Only 20% of participants said that AI tools could actually understand their local dialects. This low figure indicates that AI is still far from engaging with Arab users in their everyday spoken language. In a region marked by dialectal richness, this shortcoming is not merely technical, but deeply cultural. AI systems are operating based on a linguistic centralism that fails to account for users' social and cultural environments.

In conclusion, AI tends to treat Arabic as a monolithic, standardized language (MSA), revealing significant limitations in understanding dialectal variation, which restricts its accessibility and inclusiveness, especially for average Arab users who rely on local dialects in everyday communication.

Table 5

Participants' perceptions of symbolic representation of Arab identity in AI

Perceived Representation	**Number of Participants**	**Percentage**
Culturally neutral	6	40%
Biased toward Western models	7	47%
Reproduces elite/idealized user identity	9	60%
Supports local Arab cultural diversity	2	13%

60% of respondents stated that the language produced by AI reflects an elitist model, revealing a clear gap between the discourse of the tool and the discourse of everyday reality. Participants noted that AI systems tend to address an "ideal user": someone who is well-educated, speaks fluent Modern Standard Arabic (MSA), and adheres to formal speech norms while neglecting broader societal segments such as working-class, informal, or illiterate users.

47% of participants indicated that the epistemological foundation of these models stems from non-Arab cultural references-including conceptual frameworks, content descriptors, and topic preferences. As a result, the Arab user is positioned as the "other" within the digital environment.

A significant 40% perceived AI as ideologically neutral, reflecting a widespread but flawed belief that technology operates outside of culture. According to Darin Barney and Shoshana Zuboff, however, no AI system is fully neutral. Every model implicitly carries the perspectives and assumptions of those who designed and trained it. If it seems neutral, it is only because it is unaware of your cultural context.

Only 13% of respondents believed that AI tools support Arab cultural diversity. This limited support suggests the dominance of a monolithic identity framework embedded in algorithms-reproducing a "standardized identity" imposed from Western perspectives, rather than embracing the horizontal diversity of Arab identities.

In conclusion, current AI tools and platforms tend to reproduce a unified yet unrealistic cultural image of the Arabic language, potentially leading to a form of symbolic stereotyping of Arab users.

Table 6

Manifestations of algorithmic dominance and standardized language use in AI

Observed Phenomenon	**Number of Responses**	**Percentage**
Enforces a unified formal language style	10	67%
Marginalizes non-Gulf/Levantine dialects	5	33%
Reinforces English centrality as a primary reference	12	80%
Omits local cultural backgrounds during interaction	9	60%

The table shows that 80% of respondents observed that the AI systems they use rely heavily on English-language sources, pointing to a global epistemological centralism imposed by algorithmic structures that are non-local in nature.

Additionally, 60% of participants confirmed the marginalization of dialects and the absence of clear cultural context in AI interactions. According to 67% of the sample, the enforcement of a unified formal linguistic style stands out as one of the most prominent algorithmic patterns.

In conclusion, current AI tools reproduce the logic of linguistic hegemony by privileging both Modern Standard Arabic (MSA) and Western knowledge sources, while largely ignoring the cultural and linguistic diversity of Arab societies.

Table 7

Future outlook and participant recommendations

Recommendation	Number of Responses	Percentage
Integrate local Arabic dialects into model training	11	73%
Involve Arab experts in system and algorithm design	9	60%
Develop AI systems tailored to diverse Arab communities	7	47%
Enhance tools' ability to detect Arab cultural contexts	10	67%

The highest percentage of the sample (73%) called for the integration of local Arabic dialects into AI models and databases, reflecting a general sentiment of linguistic marginalization in current AI technologies. Participants expressed that MSA alone is not sufficient to represent their daily realities, and that supportive models are needed. This points to a growing awareness of local linguistic identity, where dialects are seen as an essential component of cultural identity. It also highlights a demand for "digital linguistic justice", aligning with postcolonial theories that view language as both a tool of hegemony and a vehicle for liberation.

A consensus among 67% of participants supported the need to enhance cultural contextual understanding within AI systems. This reflects a clear recognition that AI does not function in a vacuum; rather, it must be able to grasp the social and cultural contexts in which it operates. This concern reveals anxiety over algorithmic bias the idea that current systems may misinterpret local Arabic expressions and concepts due to their limited cultural awareness. Hence, participants advocate for the development of a "socially aware AI", not just a linguistically capable one.

More than half of the respondents (60%) also emphasized the importance of involving local experts in AI model design and training to improve cultural sensitivity and support dialects. This reflects an awareness of class and power dynamics in the digital field, as users recognize that those who build the tools also shape their function and boundaries. It echoes the call for "epistemic sovereignty," as discussed by Edward Said suggesting that technology should not merely be consumed, but must be created from within the cultural environment to serve its users authentically.

Additionally, 47% of the proposals supported the Arabization of AI and the development of localized versions tailored to different Arab communities. There is a clear shift from viewing Arab identity as a single, unified block, to recognizing it as a network of layered sub-identities (e.g., Maghrebi, Gulf, Levantine). This resonates with Anthony Giddens' concept of "multi-layered identity" in the context of digital globalization.

To conclude, the data show that Arab users are not rejecting technology per se, but rather demanding a genuine and practical localization of AI. This includes not just superficial Arabization, but active involvement of Arab experts in designing and developing culturally aligned AI systems.

6. Discussion

This section is dedicated to interpreting and analysing the research findings in light of the theoretical framework adopted by the study, which draws from sociolinguistics, theories of representation, and symbolic domination. It explores how artificial intelligence technologies are reshaping the linguistic and identity landscapes in Arab societies within a global symbolic structure centered on a dominant language and cultural authority.

6.1. Algorithmic Reproduction of "Legitimate Language":

The study's findings reveal that AI systems replicate what Pierre Bourdieu (1991) termed "legitimate language," granting Modern Standard Arabic (MSA) full centrality and symbolic legitimacy in AI environments while excluding other dialects as "unintelligible" or "unsuitable" for processing. This bias reflects what Bourdieu (1991) called the *"logic of the symbolic market,"* where the value of linguistic forms is contingent upon institutional recognition. As Bourdieu (1991, p. 55) states , "In the symbolic marketplace, it is not words alone that are valued, but the social and cultural capital they embody." Thus, AI systems not only represent language but also restructure the social positioning of Arab users based on their linguistic choices.

6.2. Conditional Identity Performance in AI Spaces:

The results further show that Arab users often adapt their digital discourse to align with what algorithms can "understand" or support. This aligns with Stuart Hall's notion that identity is not fixed but is "a continuous narrative negotiated within the boundaries of language, culture, and technology." In AI platforms, identity performance becomes conditional on one's ability to match the system's acceptable discourse model, leading to a hybrid identity shaped by algorithmic representation. As Hall suggests, *"the digital self is not what you choose, but what the platform allows you to perform."* This logic reinforces the idea that technology is not neutral it reproduces power by controlling the terms of expression and identity. In this light, AI becomes an agent of linguistic and identity normalization, as influential as traditional institutions such as media and education.

The sense of exclusion reported by participants mirrors findings from both global and regional studies. Birhane (2021) underscores that "algorithmic systems embed injustice by codifying existing inequalities, producing harms that disproportionately affect marginalized communities" (p. 3). This explains why Arab users perceive dialectal invisibility in AI interactions as a form of symbolic erasure. The issue is further compounded in translation systems, where "dialectal Arabic remains a major bottleneck…performing significantly worse compared to Modern Standard Arabic" (Elmadany et al., 2023, p. 8). These patterns resonate with the experiences of social media users documented in JSLCS (Benarab, 2024), where algorithmic amplification creates feelings of invisibility when content does not conform to dominant norms (p. 225). At the same time, as noted in that study, users develop strategies of resilience and resistance-paralleling how Arab users attempt to reinsert dialectal features despite algorithmic bias. Such dynamics confirm that AI-mediated interaction is not merely about communication with machines, but about reconfiguring the sociolinguistic conditions of human-to-human interaction.

6.3. Soft Power and Linguistic Coloniality

Building on Gramsci's notion of "soft hegemony," AI can be understood as a modern language of power that operates not through coercion, but by constructing a normative system of meaning that excludes incompatible alternatives. The privileging of English, the marginalization of dialects, and the algorithmic redefinition of acceptable identity reflect what Shoshana Zuboff describes as *"behavioral colonialism"*. This concept can be extended to what we might term *"digital linguistic colonialism."* The study's findings, which reveal the erasure of Arabic content and marginalization of dialects, resonate with Eubanks (Eubanks, 2017, p. 44), who linked automation to class-based submission-here, however, the submission is linguistic and cultural in nature. As Eubanks notes (2017, p. 44),"Algorithms do not impose opinions; they exclude dissent by not retrieving it." In this sense, dominance in the digital age no longer requires explicit control, but is perpetuated silently through the algorithmic reinforcement of symbolic hierarchies, masked as "technical efficiency" or "design constraints."

6.4. AI as a Sociolinguistic Agen:

This study positions AI not merely as a tool, but as a *sociolinguistic agent* that exerts structural influence on linguistic behaviour and symbolic representation within digital platforms. Interview data suggest that users do not passively engage with AI they negotiate with it, comply at times, and resist at others. This dynamic makes AI an integral part of users' daily identity construction. The findings contribute to an expanded understanding of sociolinguistics by shifting the analysis from physical interaction to algorithmic space, where language is no longer confined to human interlocutors but is produced and reshaped by "intelligent" systems imbued with power.

The results of this study are consistent with emerging debates on the sociolinguistic implications of digital technologies. This aligns with findings discussed in recent JSLCS studies on digital identity and algorithmic representation (Benarab, 2024; Braimoh, 2024). Recent research as in Braimoh (2024, pp. 200–209) has emphasized that digital communication does not simply transmit language but restructures symbolic power and identity. For instance, Braimoh (2024) demonstrates how texting language acts as a "digital symbolic current" that enables intercultural communication while simultaneously regulating pragmatic competence (pp. 205–207). This parallels the present findings that AI systems privilege Modern Standard Arabic while marginalizing dialectal diversity, thereby creating selective pathways of participation in digital discourse.

Similarly, Haouchi (2024) highlights the structural limits of computational semantics, noting that "automated semantic structuring reflects the computer's logic rather than the author's intent, highlighting the limits of relying on machine-driven meaning-making" (p. 200). This insight underscores the observation that AI systems are not neutral mediators but impose algorithmic logics that reshape meaning and exclude culturally embedded forms of expression such as dialects.

Finally, studies on digital identity formation also support the claim that algorithmic infrastructures exert normative pressures on users. A recent *JSLCS* article on Instagram shows that "the platform's algorithm amplifies the visibility of highly engaging content, perpetuating performance pressure" Braimoh (2024, pp. 200–209). This resonates with participants' testimonies in the present study, where the invisibility of dialects in AI-mediated interactions mirrors the feelings of exclusion reported by social media users. Yet, as the Instagram study also notes, some users resist these pressures by asserting unique forms of expression *(Braimoh, 2024, p. 207)* a pattern also evident among Arab users who attempt to reintroduce dialectal features despite algorithmic bias.

Taken together, these findings suggest that AI-mediated communication is best understood not as a neutral tool but as part of a wider ecology of digital technologies that simultaneously empower and constrain linguistic identity, reinforcing digital hegemony while also opening spaces for resistance.

6.5. ***Recommendations***

In light of the qualitative findings from this exploratory study, the following recommendations are proposed. While these suggestions emerge from a limited sample, they reflect recurring patterns in participants' experiences that may serve as a basis for broader reflection and future inquiry.

Expand Sociolinguistic AI Research. Encourage interdisciplinary research that treats AI not merely as a technical tool but as a sociolinguistic actor influencing language practice, identity performance, and symbolic representation, especially in digitally peripheral contexts such as the Arab world.

Avoid Linguistic Centralism in AI Systems. AI developers and language model designers should critically assess the overrepresentation of Modern Standard Arabic (MSA) and work toward integrating diverse Arabic dialects, based on real linguistic use among different social groups.

Include Arab Experts and Local Institutions. Encourage the **active involvement of Arab scholars, sociolinguists, and developers** in AI system design to ensure contextual and cultural sensitivity. This can also mitigate the replication of Western-centric linguistic norms.

Support Open, Culturally Diverse Arabic Data Sets. Promote **localized and inclusive Arabic language corpora**, especially for underrepresented dialects, to enhance AI's ability to reflect the region's linguistic diversity and cultural plurality.

Regulate for Linguistic Equity in AI. Initiate public policy discussions and academic-industry collaborations that develop ethical guidelines ensuring **linguistic equity and cultural inclusion** in AI technologies used in the Arab world.

Empower Users through Digital Awareness. Encourage critical digital literacy programs that help users understand the **sociotechnical implications of AI language tools**, including how they may reinforce or challenge existing symbolic hierarchies.

These recommendations are not prescriptive for all Arab societies but reflect the views and concerns expressed by the study's participants. Broader implementation should be informed by larger-scale, comparative studies across different social, linguistic, and regional contexts.

6.6. Future Research Directions

This study opens the door to a range of questions not fully explored, such as:

- How do multilingual AI models represent dialects compared to native languages?
- What is the impact of smart model interaction on the linguistic and identity formation of Arab children?

These questions offer fertile ground for further field-based and cumulative research aimed at understanding the rapid transformations imposed by algorithms on Arab linguistic, representational, and cultural existence.

7. Conclusion

This study explored how artificial intelligence (AI) systems interact with Arabic language varieties and how such interactions influence users' perceptions of identity and cultural representation. Through a sociolinguistic lens grounded in theories of symbolic power and cultural hegemony, the findings highlighted a clear tendency of AI to prioritize Modern Standard Arabic (MSA) while marginalizing local dialects, reinforcing a normative linguistic hierarchy within digital environments.

Importantly, these insights are derived from a qualitative, purposive sample of 15 participants from six Arab countries. As such, the findings do not claim statistical generalizability but rather aim to offer a context-rich, interpretive understanding of how linguistic identity is shaped within AI-mediated spaces.

The participants' experiences suggest that current AI tools function not only as communicative agents but also as symbolic gatekeepers, promoting certain language forms while excluding others. This asymmetry reproduces existing power dynamics between global and local knowledge systems and reflects broader sociological patterns of digital stratification and cultural centralism.

By drawing on Bourdieu's theory of symbolic capital, Hall's identity theory, and Gramsci and Zuboff's perspectives on hegemony, the study unpacks the mechanisms of what might be called "soft algorithmic colonization," which is perpetuated through the illusion of technical neutrality.

While the results are specific to the sample studied, they raise broader questions about linguistic justice, epistemic sovereignty, and the role of AI as a cultural actor in Arab societies. These issues warrant further sociological investigation using larger, more diverse samples and interdisciplinary frameworks.

Acknowledgements

The author extends sincere gratitude to all participants who generously contributed their time and insights to this study, providing valuable qualitative data that enriched the analysis.

The author also acknowledges the limited use of Artificial Intelligence (AI) tools during the preparation of this manuscript. Such tools were employed exclusively for technical tasks, including proofreading, paraphrasing, and reference formatting, under the author's full supervision. All critical analysis, interpretation of findings, and scholarly contributions presented in this article are entirely the author's original work

References

Abderrahmane, T. (2011). *Al-hadatha wa-l-muqawama [Modernity and resistance]*. Beirut: Ma'arif al-Hikmiyya Institute for Religious and Philosophical Studies.

Bakhtin, M. M. (1981). *The dialogic imagination: Four essays,* (M. Holquist, Ed.; C. Emerson & M. Holquist, Trans.). University of Texas Press.

Barney, D. (2004). *The network society.* Polity Press.

Benarab, I. H. (2024). The promise of emancipation of digital technologies and the risks of alienation in the self-construction: The case of Instagram among Algerian youth aged 15 to 25. *Journal of Studies in Language, Culture and Society (JSLCS), 7*(3), 214–230.

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

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns, 2*(2), 100205. https://doi.org/10.1016/j.patter.2021.100205

Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5454–5476). ACL. https://doi.org/10.18653/v1/2020.acl-main.485

Bourdieu, P. (1991). *Language and symbolic power,* (J. B. Thompson, Ed.; G. Raymond & M. Adamson, Trans.). Polity Press.

Braimoh, J. J. (2024). Texting language as a digital symbolic current: Implications for pragmatics and intercultural communication in the digital age. *Journal of Studies in Language, Culture, and Society, 7*(2), 200–209.

Canagarajah, S. (2013). *Translingual practice: Global Englishes and cosmopolitan relations.* Routledge.

Castells, M. (2010). *The rise of the network society* (2nd ed.). Wiley-Blackwell.

Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism.* Stanford University Press.

Elmadany, A., Nagoudi, E. M. B., & Abdul-Mageed, M. (2023). ORCA: A challenging benchmark for Arabic language understanding. *arXiv preprint, arXiv:2212.10758.* https://arxiv.org/abs/2212.10758

Eubanks, V. (2017). *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.

Goffman, E. (1959). *The presentation of self in everyday life.* Anchor Books.

Gramsci, A. (1971). *Selections from the prison notebooks* (Q. Hoare & G. N. Smith, Eds. & Trans.). International Publishers.

Hall, S. (1990). Cultural identity and diaspora. In J. Rutherford (Ed.), *Identity: Community, culture, difference* (pp. 222–237). Lawrence & Wishart.

Hall, S. (1996). Who needs identity? In S. Hall & P. du Gay (Eds.), *Questions of Cultural Identity* (pp. 1–17). Sage.

Hanandeh, A., Ayasrah, S., Kofahi, I., & Qudah, S. (2024). Artificial intelligence in Arabic linguistic landscape: Opportunities, challenges, and future directions. *TEM Journal, 13*(4), 3137–3145. https://doi.org/10.18421/TEM134-48

Haouchi, A. (2024). Systematic construction of semantic structure computationally: Digital communication systems as a model. *Journal of Studies in Language, Culture, and Society (JSLCS), 7*(3), 197–213.

Kilito, A. (2009). *Lughat al-Akhar: Mudhakkirat min al-Tarjama* [The language of the other: Notes on translation]. Dar Toubkal.

Milan, S., & Treré, E. (2022). Big data from the South(s): Beyond data universalism. *Television & New Media, 23*(2), 240–255.

Said, E. W. (1978). *Orientalism.* Pantheon Books.

Seyidov, R. (2024). Arabic language processing: Current status and future prospects of artificial intelligence. *Journal of Namibian Studies: History, Politics, Culture, 41,* 224–240. https://doi.org/10.59670/jwszy037

van Gerven Oei, V. W. J. (2020). AI and the coloniality of language. *Critical Algorithm Studies Journal, 4.*

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power.* PublicAffairs.

Appendix

Semi-Structured Interview: Linguistic Interaction with Artificial Intelligence in the Arab Context

General Information of the Interviewee:

- Age:
- Gender:
- Country:
- Academic or Professional Background:
- Level of interaction with AI language tools:
  ☐ High ☐ Medium ☐ Low

Section 1: General Experience with AI Language Tools

1. When did you first use a language-based AI tool (such as ChatGPT, Google Translate, etc.)? And why?
2. What types of tasks do you most commonly use these tools for? (e.g., translation, paraphrasing, content generation, grammar correction, etc.)
3. Do you notice a difference in performance between Arabic and other languages? ☐ Yes ☐ No — If yes, please describe the difference.

Section 2: Linguistic Interaction and Dialects

4. How would you describe the AI tool's interaction with your local dialect? Do you use these tools in your dialect or in Modern Standard Arabic (MSA)?
5. Have you encountered any cases of misinterpretation or distortion when using dialectal or Arabic expressions? Please provide examples.
6. Do you feel the tool "understands you" when you use culturally or locally specific expressions?

Section 3: Identity and Symbolic Representations

7. Do you think these tools reflect a specific image of "Arab identity"? Or do they tend to follow globalized or Western linguistic models?
8. How do AI-generated responses influence your sense of linguistic or cultural belonging?
9. Have you ever felt that the tool reproduces stereotypes about Arabs or the Arabic language? How so?

Section 4: Trust, Hegemony, and Reproduction of Power

10. To what extent do you trust AI-generated results when it comes to Arabic language? Why?
11. Do you think these tools impose a specific linguistic style or a globally standardized register?
12. Have you noticed that certain Arabic dialects or expressive forms are overlooked or excluded by these systems?

Section 5: Future Outlook and Recommendations

13. What developments or improvements would you like to see in AI language tools to better serve Arab communities?
14. In your opinion, who should be responsible for developing culturally and linguistically relevant AI for the Arab world?
    ☐ Tech companies ☐ Universities ☐ Civil society
15. Do you believe these systems could support Arab linguistic diversity rather than marginalize it?