Investigating the Role of Artificial Intelligence Voice Assistants for Arabic-Speaking Users within the Context of Knowledge-Based Systems

By

Eslam Najim Badran

A Dissertation Submitted to the Department of Digital, Technology, Innovation and Business

University of Staffordshire

In fulfilment of the requirements for the degree of Doctor of Philosophy

October 2025

Acknowledgement

First and foremost, I would like to express my deepest gratitude to my parents, whose unwavering support, love, and encouragement have been the foundation of all my achievements. Their sacrifices and belief in me have been a constant source of strength throughout my academic journey.

I am also sincerely thankful to my wife for her support throughout this journey, and to my lovely twins for always motivating me. My heartfelt thanks also go to my sisters and brothers, who have always stood by me with their love, motivation, and understanding—especially during the most challenging times of this PhD journey.

I am immensely grateful to my main supervisor, Dr Maryam Shahpasand, for her invaluable guidance, continuous support, and insightful feedback. Her mentorship has been instrumental in shaping both the direction and quality of this research. I truly appreciate her patience, dedication, and belief in my potential.

To all my friends, thank you for being there during the highs and lows. Your companionship, laughter, and encouragement made this journey more manageable and enjoyable.

This work would not have been possible without the contributions and presence of each of you. Thank you all from the bottom of my heart.

Author’s Declaration

I declare that the work presented in this thesis is entirely my own and has not been submitted, either in whole or in part, for the award of a degree at the University of Staffordshire or any other institution of higher education. All sources of information have been duly acknowledged and referenced in accordance with academic conventions.

Where research has been undertaken in collaboration with others, full details of the contribution of collaborators have been provided in the relevant chapters of this thesis. The intellectual contribution, design, analysis, and writing of this work remain my own.

I understand the consequences of plagiarism and confirm that this thesis represents original research carried out and written solely by me, except where explicitly stated.

Candidate’s Name: ESLAM NAJIM AHMAD BADRAN

Signature:

Date: 15/October/2025

ABSTRACT

Artificial Intelligence Voice Assistants (AIVAs) are becoming more and more part of our daily life, but its usage in Arabic-speaking societies is not evenly distributed due to diglossia, code-switching, culturally specific communication patterns, and the lack of its integration with organisational knowledge-based systems (KBS). Weaknesses in dialect recognition, mixed-language interaction and lack of culturally based evaluation frameworks to use in real-world organisations have also limited previous studies on Arabic AIVAs. This paper examined the influence of AI voice assistants on Arabic-speaking users in the United Arab Emirates and looked at the relationships between perceived benefits, challenges, design preferences, effective usage and government communication and the perceived role of AIVAs. The survey design was a two-round cross-sectional survey design. The first round investigated usage patterns, language preferences, cultural issues, technical knowledge, and user expectations, and the second round evaluating an integrated conceptual model with structural equation modelling.

The results showed that the last model had a very high fit on the survey data and accounted a large percentage of variance on the perceptions of respondents about the importance of AIVA. The respondents indicated that they were highly exposed to AI voice assistants but less frequently used in Arabic. The primary perceived barriers related to dialect recognition, code-switching, privacy, transparency, cultural appropriateness and limited integration with organisational systems. The results also indicated that reported benefits, perceived challenges, design preferences, and communication in government significantly correlated with the perceived significance of AI services. The paper adds a culturally and linguistically based list of requirements of Arabic AIVAs, such as dialect-sensitive interaction, code-switching support, privacy-conscious consent, culturally sensitive communication, accessibility, and transparency of the source. It also suggests a conceptual blueprint of AIVA-KBS integration that would enable the Arabic-speaking users to obtain authorised organisational knowledge in a controlled voice interface. This blueprint is a proposed future system, not a technically implemented or tested system. Altogether, the research is a theoretical, empirical, methodological, practical, and policy-related advice on how to create more inclusive and trustworthy Arabic AIVAs in an organisational and public-service environment.

Keywords: AI Voice Assistants, Arabic-speaking community, culture, information systems, and knowledge management.

Table of Contents

List of Tables

List of Figures

LIST OF ACRONYMS

Acronym	Full Form
AI	Artificial Intelligence
AIVA	Artificial Intelligence Voice Assistant
UAE	United Arab Emirates
LLM	Large Language Model
NLP	Natural language processing
MSA	Modern Standard Arabic
ASR	Automatic Speech Recognition
iOS	iPhone Operating System
macOS	Macintosh Operating System
CRM	Customer Relationship Management
Amazon Alexa App	Amazon Alexa Application
IoT	Internet of Things
LDI	language Dialect Identification
sRAM	Service Robot Acceptance Model
UTAUT2	Unified Theory of Acceptance and Use of Technology 2
APCoT	Arabic Personalized Control of Things
ITS	Intelligent Tutoring Systems
TAM	Technology Acceptance Model
QCRI	Qatar Computing Research Institute's
DEWA	Dubai Electricity and Water Authority
DID	dialect identification
MENA	Middle East and North Africa
SPSS	Statistical Package for the Social Sciences
AMOS	Analysis of Moment Structures
ANOVA	Analysis of Variance
SEM	structural equation modelling

RMSEA	Root Mean Square Error of Approximation
CFI	Comparative Fit Index
df	degrees of freedom
CR	Composite Reliability
AVE	Average Variance Extracted
TLI	Tucker & Lewis
AR	Augmented Reality

1. CHAPTER 1: INTRODUCTION

Artificial intelligence voice assistants (AIVAs), such as Siri, Alexa, and Google Assistant, enable users to interact with digital systems through spoken commands (Khan, Rajput & Jeyasekar, 2025). The activities that they support include information retrieval, scheduling, messaging, navigation, and access to online services. Their increased use has provided chances of more natural, more accessible and efficient communication between people and digital media.

In this thesis a Knowledge-Based System (KBS) is an organisational system that captures, organises, stores, updates and retrieves proven knowledge to aid service delivery, problem solving and decision making (Raza, 2022). Some of the knowledge may be contained in policies, procedures, service guidance, frequently asked questions, regulatory information, and institutional records. The AI voice assistant might be used as a conversational interface to a KBS by accepting a spoken command, identifying the intent of the user, searching an authorised source of knowledge, and delivering the answer as spoken (Rawat et al., 2024).

The interaction is however complicated in Arabic speaking situations. Arabic has Modern Standard Arabic and a variety of local dialects and many of the Arabic speakers also use Arabic and English interchangeably in the same conversation (Alwash, 2025). These language characteristics can influence speech recognition, intent identification and accuracy of response. Interaction is also affected by culture especially when it relates to politeness, address, religion and the use of formal or informal language. Moreover, privacy, transparency and trust are relevant since voice data can contain personal or sensitive data, and users should be confident that responses are correct, up-to-date and based on reliable sources (McCormack et al., 2025).

The UAE offers a topical environment since the organisations of the public and the private sectors are increasingly adopting artificial intelligence, digital services, automated customer-service, and knowledge-management platforms (Dahabreh 2023). The voice-based access to organisational knowledge may be useful to government departments, healthcare providers, universities, banks, and utility companies. However, to achieve successful adoption, lingo-incusive, culturally suitable, privacy-conscious, and safely linked to credible sources of knowledge are required.

Current studies have centred on English-speaking users or the broad adoption of technology, with little emphasis on Arabic-speaking users in the UAE and the possible integration of AIVAs with organisational KBS systems (Zou et al., 2025). This study thus examined reported usage, perceived benefits, challenges, preferences, and determinants of AIVA importance in two rounds of surveys. It neither created nor technologically tested a workable AIVA or KBS prototype. Rather, the results were applied, together with the literature, to guide an intended conceptual blueprint of future AIVA-KBS integration. In that way, the suggested blueprint can be seen as an empirical informed conceptual input, as opposed to evidence of real technical performance, organisational performance, or real-world application by UAE institutions. The main motive of this study is to explore the impact on AI Voice Assistants emerging trend across Arab world and they're daily going-to-used localization within a country, United Arab Emirates in scope of my project. This paper is mainly about the rising trends of deploying AI voice assistants among Arabic-speaking in UAE, and how it benefits both public & private sectors (Bozeman, 2019).

The study further extends to investigate AI voice assistance about information systems and knowledge management within the firm, where data is retained, managed and analyzed. This research employs a two-round cross-sectional survey. Round 1 mapped usage and perceptions; Round 2 tested a hypothesised structural model using SEM. The results of this research will elaborate the effect of AI voice assistants on Arabs as a group and other institutions that work within it. The initiative will also specify the obstacles that must be overcome so voice assistants are properly leveraged in the Arab world. The paper laid out a couple of sections (Bozeman, 2019).

1.1 Problem Statement

AI voice assistants are gaining significant traction globally, specifically by Amazon's Alexa, Apple's Siri, and Google Assistant, for possible benefits that could enhance access to information, service delivery, and knowledge management across all sectors. While this might be the case, currently, their popularity gained momentum globally, their adoption and utilization in either the Arabic-speaking community in general or the UAE alone remain underexplored. There is a great need for understanding how those technologies are perceived, deployed, and utilized from both the private and public sectors. Their adoption and integration in the Arab world, especially within the UAE, presents quite significant challenges, according to Bozeman, 2019. The challenges include limited linguistic support, dialectical variations, and lack of cultural adaptation, which limit widespread acceptance and usage among Arab users (Zerilli et al., 2022). Another challenge is, of course, how AI voice assistants are integrated into any existing information systems and knowledge management solutions within the area, especially with regard to such important issues as data management, storage, analysis, or decision-making. The current study, therefore, will try to discover and understand the growing usage of AI voice assistants in the Arabic-speaking communities to influence organizational processes, specifically how these technologies can connect complex information systems and knowledge management practices. Consequently, the investigation will add enormous value through profound insights regarding the problems and opportunities regarding AI voice assistant adoption in a specific cultural and organizational context.

1.2 Research Questions and Objectives

Table 1-1: below maps the research questions to their corresponding objectives and the chapters where they are addressed

Objectives	Research Questions
O1: Quantify adoption and usage patterns among Arabic-speaking adults in the UAE (trial vs sustained use; Arabic vs English; usage contexts).	Q1: How are AI voice assistants integrated into the daily activities of Arabic-speaking communities?
O2: Test a structural model of perceived importance/continued use using benefits, challenges, design preferences, and expectations of	Q2: What impact does technical knowledge have on the perceived challenges and benefits associated with AI voice assistant usage in Arab society?

government communication as predictors (SPSS/AMOS).	
O3: Identify linguistically/culturally grounded requirements and an evaluation checklist (dialect coverage, code-switching, forms of address, consent transparency) that address barriers to trust and sustained Arabic use.	Q3: How can AI voice assistants be effectively utilized for communication within this demographic?
O4: Produce an integration blueprint linking Arabic-speaking users interaction needs to organisational knowledge processes in UAE public/private settings, with policy guidance balancing utility and privacy (consent, redress, data residency).	Q4: How can AI voice assistants be designed to better meet the preferences of the Arabic-speaking communities?

1.3 Research Aim

The aim of the study is to investigate the role of AI voice assistants for Arabic-speaking user within the context of knowledge-based systems.

1.4 Purpose of the Study

The purpose of this study is to investigate the impact of AI voice assistants on the daily lives of Arabic-speaking individuals, particularly in the United Arab Emirates (UAE). By examining the usage patterns, benefits, and challenges associated with these technologies, the study provided insights into how AI voice assistants can be better tailored to meet the needs of the Arabic-speaking communities (Grimsley, 2007).

1.5 Significance of the Study

This study shifts from generic, English-centric AIVA evaluation to a UAE-specific analysis that foregrounds Arabic diglossia, code-switching, and socio-pragmatic norms. The primary original contribution of this research is the construction of an empirically-informed, culturally-based knowledge of AI Voice Assistants (AIVAs) among Arabic-speaking users in the context of knowledge-based systems (KBS). Empirically, two large surveys quantify usage, preferences, and constraints; a validated measurement model isolates determinants of AIVA importance. Conceptually, the work links Arabic-speaking users interaction needs to organisational KBS requirements, turning cultural-linguistic insights into explicit design and integration criteria. Practically, it provides (a) an evaluation checklist for dialect/context handling and privacy-forward consent patterns, and (b) an integration blueprint for public- and private-sector systems. Collectively, these contributions can create a platform of designing more inclusive, trustful, and culturally adaptive AIVAs through the integration of empirical evidence of users, linguistic and cultural needs, and organisational knowledge requirements.

A part of this study has been published as a peer-reviewed conference paper—Badran, E. & Seker, H. (2024) ‘Studying Factors Affecting Attitudes on Artificial Intelligence Voice Assistants in the United Arab Emirates’, Proceedings of the ISICS2024 International Conference on Intelligent Computing Systems—which reports the Round-1 findings; the extended analysis and framework are presented in Chapter 4 (Results), Chapter 5 (Discussion) and Chapter 6 (Conclusion and Recommendations).

Table 1-2: Significance of the study

Aspect	Prior Studies	This Study
Focus	Adoption of AI voice assistants in global contexts, mainly English-speaking users	Specific investigation into adoption of AI voice assistants in Arabic-speaking regions, e.g., the UAE.
Linguistic Coverage	Limited to major global languages with underrepresentation of Arabic	Discusses Arabic dialectical differences & how it impacts usage of AI voice assistant.

Cultural Sensitivity	Lack of exploration of cultural elements that affects AI adoption in Arabic communities.	Discusses Cultural Differences particular to Arabic-speaking users & their implications for user experience & adoption.
Barriers Identified	General technical challenges like voice recognition & privacy issues	Identification of technical, cultural, & linguistic barriers specific to Arabic users.
Methodology	Broad surveys focusing on global trends.	Mixed-method approach combining quantitative surveys & qualitative feedback tailored to Arabic-speaking respondents.
Regional Context	Majorly Western-centric studies	Focuses on the UAE as a representative case for Arabic-speaking users
Actionable Outcomes	Generic suggestions for enhancing AI systems globally.	Specific recommendations for developers & end users to enhance linguistic & cultural inclusivity in AI systems.

1.6 Scope of the Study

The scope of this study is limited to the UAE, representing a sample of the broader Middle Eastern region. Due to rapid transformations to digital services, UAE now lead the AI to be the main power in the region. The research focuses on the use of AI voice assistants in daily activities within both the public and private sectors and related challenges such as cultural differences and dialectal issues that may affect the proper use of AI voice assistants. The study also examined the relationship between AI voice assistants and information and knowledge management within organizations (Grimsley, 2007).

1.7 Structure of the Thesis

The thesis has been structured as follows:

Chapter 1 – Introduction. States the background and rationale; presents the problem statement, purpose, aims, research questions, significance, scope and an overview of the thesis.

Chapter 2 – Literature Review. Synthesises scholarship on AI voice assistants in Arabic contexts (diglossia, code-switching, cultural pragmatics); identifies theoretical and empirical gaps; and develops the conceptual model and hypotheses underpinning the study.

Chapter 3 – Methodology. Outlines the two-round cross-sectional survey design, instrument development and validation, sampling and data collection, ethics, and analysis methods (SPSS and AMOS SEM).

Chapter 4 – Results. Presents descriptive and inferential statistical analyses, including measurement reliability and validity outcomes, structural-model fit indices, and hypothesis-testing results aligned with the research objectives.

Chapter 5 – Discussion. Interprets findings in relation to the literature and conceptual framework, discusses theoretical and practical implications, and acknowledges study limitations.

Chapter 6 – Contributions and future work. Summarises key contributions, offers stakeholder-specific recommendations, and outlines directions for future research.

2. CHAPTER 2: LITERATURE REVIEW

Artificial intelligence and digital systems are increasingly integrated into different aspects of daily life, influencing information access, decision-making, and communication. AI-based technologies, including AI voice assistants, have revolutionized the way people engage with digital platforms, providing users with instant access to information, automated services, and personalized assistance. While AI research and development have significantly flourished in the last few years, much of the focus has been on Western societies and their needs, often overlooking the societal, linguistic, and cultural variations that impact AI adoption and applicability in non-Western contexts. The role of AI-driven voice technologies in Arab communities, specifically within the framework of knowledge-based systems, remains an underexplored area requiring deeper investigation. Due to the cultural and linguistic specificity of Arabic-speaking societies, it is important to explore how these systems fit user expectations and social norms in the knowledge-based systems.

It is imperative to comprehend linguistic and cultural differences in the development of AI systems that meet several user needs and expectations. Culture is instrumental in shaping human interaction, defining how individuals perceive and interact with technology. The review investigates the existing studies that underscore the importance of culture and language, analysing how linguistic diversity, social structures, and cultural norms influence social behavior, values, and interactions. Subsequent to this, the chapter discusses the significance of technology in reflecting and strengthening cultural identities or biases of different communities by acting as a bridge between cultures. Following that, the literature review looks at the state of literature in regard to AI voice assistants (AIVAs) and how they have progressed from simple natural language processing (NLP) means to sophisticated programs capable of dealing with context-aware interactions and multiple languages.

Advanced capabilities enable AI voice assistants to make the experience more customized and dynamic, but the challenge is significant when utilizing the systems for Arabic-language users. Further, the review section continues discussing the role played by AI voice assistants in boosting facilitation, productivity, and serviceability for knowledge-based systems. Lastly, the chapter concentrates on discussing the extent of current literature on AI voice assistants, particularly with regards to dialectal and cultural flexibility, and recognizing gaps that need to be addressed.

Integrating all these elements, this chapter aims to provide the depth of understanding for the interconnectivity between technology, culture, and AI voice systems for Arabic-speaking populations. It therefore draws focus to culture-sensitive adaptive AI-based speech recognition systems that are designed to accommodate diverse linguistic and social environments while prioritizing ethical perspectives.

In Arabic HCI specifically, three interdependent constraints recur: (i) diglossia (Modern Standard Arabic, MSA, versus regional dialects), (ii) code-switching (Arabic–English/French mixing), and (iii) orthographic ambiguity (absence of diacritics in everyday text and heterogeneity in transliteration). Each directly impacts automatic speech recognition (ASR), natural language understanding (NLU), and dialogue management. Large shared tasks and evaluation campaigns (as MGB-2/MGB-3; MADAR; recent ArabicNLP speech tasks) repeatedly show that dialectal variability measurably degrades recognition and classification performance compared with MSA, even for strong systems, underscoring the need for dialect-aware modeling and evaluation (Khalilia et al., 2024; Ali et al., 2017).

This chapter reviews foundational work on culture and language, surveys technology–culture interactions, tracks the evolution of AIVAs toward context-aware, multi-device systems, and synthesizes what is known about Arabic usage and constraints. It then identifies gaps and motivates why culturally adaptive AIVAs matter for KBS in Arab public-sector, education, and healthcare settings.

2.1 Culture and Linguistics: Role in Societal Structures

Language and culture are intrinsically related, serving as the basis for social structures and determining human interactions. Culture involves the norms, customs, beliefs, and values of a particular region, while linguistics examines language as a tool for communication and social expression. The interaction of these factors determines the functioning, development, and solidarity of societies. Language is not just a tool for communication but a vehicle for cultural heritage. It maintains and conveys cultural knowledge from one generation to the next, reaffirming social identity and community values. The research by Muminovna (2024) underscores the significance of linguistic culture in modern society, discussing how language structures impact identity construction and social functions. While Madhi (2025) emphasises the role of language in identity construction and social functioning, later work on Arabic human–computer interaction extends this argument by showing that cultural meaning must also be reflected in the design of digital interfaces. Nonetheless, overall discussions of linguistic culture fail to describe how these cultural expectations must be realised in AI voice assistants or organisational knowledge systems (Susanto & Hendarman, 2025). This brings about a necessity to study particular interactional aspects, e.g. register, politeness, forms of address and culturally relevant response.

AIVAs that operate as interfaces to KBS must therefore perform socio-pragmatic alignment: selecting pronouns of address, politeness markers, and register (formal/colloquial) that fit the user and the situation. In Arabic, this includes appropriate use of religious salutations, avoidance of informal vocatives in official contexts, and sensitivity to kinship terminology when scheduling or relaying family-related information. Failure to align pragmatics can lower perceived trust and competence even if the propositional content is correct, a pattern noted in service encounters literature and increasingly reported in voice UX research for non-English markets. Additionally, this stance contrasts with completely technical methods that assess AIVAs primarily by their recognition accuracy or task completion. The socio-pragmatic research claims that a technologically correct answer can be evaluated as negative situation when the assistant employs an unsuitable register or address. Thus, linguistic precision and cultural appropriateness should be considered as two similar yet different aspects of AIVA efficiency.

Policy frameworks now encourage culturally grounded AI. UNESCO's 2021 Recommendation on the Ethics of AI calls for context-sensitive design, human oversight, fairness, and cultural diversity protection across education, science, culture, and communication principles directly relevant to public AIVAs in Arabic-speaking societies. National strategies and procurement can translate these principles into design requirements as like dialect evaluation, content provenance (Vinti, 2021). UNESCO

(2021) gives general ethical guidelines, and Vinti (2021) is more focused on the practical implementation of these guidelines into the requirements of governance and design. Both contributions, however, are mostly normative and do not define what cultural, linguistic, or privacy-related demands Arabic-speaking users deem as the most important in practice. The latter weakness justifies the empirical research of perceived problems and design preferences of users.

May (2013) examines the relationship between social power, linguistic diversity, and national identity. In this regard, conceptual linguistic relativity posits that the language structure affects the cognition and worldview of its speakers, shaping societal behavior and decision-making processes. In addition, societies with diverse linguistic communities tend to grapple with issues of integration and representation, which require policies that encourage multilingualism and cultural inclusivity in all aspects of life (Roman, 2024).

A further strand of scholarship concerns speech acts and ritual language, which are central to coordinating social life. In many Arabic-speaking settings, directives, requests, and offers are routinely embedded in indirect forms that protect face and signal deference (like law samaḥt “if you permit,” mumkin…? “Would it be possible…?”). Classic speech-act theory (Searle, 1979) and later cross-cultural pragmatics (Blum-Kulka, House, & Kasper, 1989) show that felicity in requests, refusals, and apologies depends on culturally conventionalized mitigators and formulae. For AIVAs interfacing with knowledge-based systems (KBS), intent recognition must therefore be coupled with politeness-aware generation, so confirmations, clarifications, and error repairs conform to local expectations especially in high-stakes, public-service contexts. Systems that default to bald imperatives (“repeat your phone number”) can be experienced as brusque; those that use mitigated prompts (“law samaḥt, could you repeat…”) tend to be rated as more respectful and trustworthy. A second, equally practical dimension is registering management under diglossia. Ferguson’s (1959) model, elaborated by Albirini (2016), explains how the High variety (MSA) indexes formality, education, and institutional authority, while Low varieties index intimacy and locality. AIVA outputs that mismatch the activity frame for instance, responding in a colloquial register to a ministry query risk undermining institutional persona. Conversely, MSA-only output in domestic contexts can sound stiff or distant. A workable design pattern is task-sensitive register switching: default

MSA for official answers with optional plain-Arabic paraphrases; default vernacular for personal reminders with the ability to "elevate" formality when the user's goal (like submitting a complaint) calls for it.

Searle (1979) provides a general explanation of how speech acts function, while Blum-Kulka, House, and Kasper (1989) demonstrate that their appropriate realisation varies across cultures. In the same vein, Ferguson (1959) sets the difference between high and low Arabic varieties, and Albirini (2016) demonstrates that this difference can be applied to the modern social and institutional communication. These points of view are complementary to each other, although they were not designed specifically to AI-mediated interaction. Thus, their usage in AIVAs is conceptually convincing but needs to be supported by users in relation to the preferred register, politeness, and style of communication.

Third, addressivity and kinship merit explicit modelling as an Arabic kinship terms (ʿamm, khāl, bint ʿamm) often surface in calendaring and messaging ("remind me to call khāli"). Kinship is not merely lexical; it encodes obligations, respect gradients, and lineage (Antoun, 1968). Entity resolution that treats such terms as first-class referents, plus disambiguation prompts that preserve deference ("Do you mean your maternal uncle Khālid?"), improves both accuracy and social fit. Similarly, honorifics and titles—Ustādh/Ustādha, Duktūr/Duktūra—signal roles that AIVAs should neither over- nor under-apply; defaulting to honorifics in initial turns and mirroring user preferences in subsequent turns aligns with accommodation theory (Giles, Coupland, & Coupland, 1991). Fourth, bilingualism and code-switching are normative in many Arab cities. Users routinely mix Arabic with English (and, in the Maghreb, French) within the same turn. Research on Arabic NLP confirms that token-level switching and borrowed phonology degrade ASR/NLU if systems assume monolingual input (Bouamor, Habash, & Salameh, 2018; Ali et al., 2021). For KBS interfaces, robust handling requires: (i) automatic language-ID at the sub-utterance level, (ii) dialect identification (DID) prior to parsing, and (iii) graceful repair strategies ("I understood the English medication name; would you like the Arabic instructions?"). Public-sector procurement can require dialect-stratified WER/CER reporting and MADAR-style DID benchmarks to prevent systematic exclusion of dialect communities.

Fifth, gender, age, and accessibility shape interactional expectations. Sociolinguistic work documents gendered variation in forms of address, hedges, and evaluative particles across Arabic varieties (Huehnergard, 2017). Older adults may prefer slower speech rates, explicit confirmations, and larger prosodic phrase boundaries; persons with speech impairments may rely on constrained grammars or alternative modalities. Inclusive voice design therefore spans acoustic, lexical, and pragmatic accommodation, aligning with universal design principles and accessibility research (Czaja et al., 2006). For Arabic AIVAs, this includes clear Hijri/Gregorian disambiguation, diacritic-aware TTS for proper names, and opt-in repetition/summary modes that reduce memory load. Sixth, institutional trust practices in the region emphasize provenance and recency. Knowledge claims are evaluated not only by content but by source identity (ministry portals, official gazettes) and update timestamps. International frameworks UNESCO (2021) and NIST AI RMF 1.0 (2023) converge on transparency and documentation. Concretely, AIVAs should announce sources ("From the Ministry of Health, updated 12 Shaʿbān 1446 / 21 Feb 2025"), expose links for verification, and signal confidence and applicability conditions (like emirate-specific rules). Such practices convert abstract "AI ethics" into operational trust signals native to KBS workflows.

The evaluation must include culture-linguistic metrics, not just generic accuracy. Teams should submit dialectal coverage, code-switch robustness, pragmatic appropriateness scores, and pronunciation accuracy for Arabic proper names in addition to ASR WER and NLU F1. User studies can measure repair success following misrecognition, assess register preferences by task, and elicit Likert-rated politeness/appropriateness. Participatory design with Arabic-speaking users, particularly those from underserved dialect groups, lowers the risk of culturally incongruent deployments and is in line with best practices in human-centered AI (Amershi et al., 2019). Furthermore, policies convert ideals into legally binding procedures. In order to comply with local data protection laws, national AI and digital government strategies can incorporate (a) dialect-aware procurement benchmarks; (b) minimum provenance and timestamp requirements for public responses; (c) privacy-by-design defaults; and (d) appeal and human-override mechanisms when decisions have significant ramifications. These guardrails respect linguistic diversity, protect cultural expression, and ensure that

AIVAs act as trustworthy mediators of knowledge, not merely transactional query handlers.

There should be a sharp difference in technical evaluation and user-perception research. Actual system performance is measured by use of WER, NLU F1, pronunciation accuracy and code-switching benchmarks, and survey measures users' perceptions of language, privacy, cultural sensitivity, and trust. Despite the technical and ethical advice proposed in the literature, there is little evidence of Arabic-speaking users in the UAE. This leaves three gaps: incomplete knowledge of user perception of dialect and code-switching issues; lack of empirical validation of politeness, privacy, accessibility, transparency, and cultural appropriateness in a single framework; and the weak relationship between user needs and organisational KBS functions. These loopholes inform the research questions and justify a tailored framework that connects benefits, challenges, design preferences, government communication, perceived importance, and the proposed AIVA-KBS integration blueprint.

2.1.1. LLM-based models and implications for voice assistants

Voice assistants are increasingly powered by large language models (LLMs), which are the foundation of today's conversational systems. Technical studies and evaluations demonstrate multimodal understanding for text-image-audio inputs (Team et al., 2023), strong performance from open models tailored for dialogue (Touvron et al., 2023), and broad increases in reasoning, instruction following, and tool use (Bubeck et al., 2023; Achiam et al., 2023). Nonetheless, the studies focus on various aspects of LLM capability. Team et al. (2023) tested multimodal understanding, but Touvron et al. (2023) tested open models that were created to facilitate dialogue. Bubeck et al. (2023) and Achium et al. (2023) in their turn analyzed more general reasoning improvements, post-instructional improvement, and the use of tools. Despite the overall evidence of rapid progress, these findings do not prove that such capabilities can be as effective across languages and dialects and culturally specific voice interactions.

In interactive contexts, prompting and alignment techniques RLHF and Constitutional AI enhance usefulness and safety (Ouyang et al., 2022; Bai et al., 2022). Additionally, Ouyang et al. (2022) tested the reinforcement learning based on human feedback, whereas Bai et al. (2022) tested Constitutional AI as a different method of alignment. Both techniques are intended to enhance safety and usefulness, yet the usefulness of these techniques lies in the representativeness of the feedback, rules and training data in the culture and language involved. It is especially of interest to Arabic-speaking users, the dialectal and socio-pragmatic expectations of which might not be well captured in largely English-language alignment processes. Particularly pertinent to knowledge-based systems in government, health, and education, retrieval-augmented generation (RAG) lowers hallucination for knowledge tasks by firmly establishing responses in referenced sources (Lewis et al., 2020).

If Arabic variations and code-switching are explicitly included in training and assessment, surveys and model reports for Arabic use indicate that coverage lags behind English; performance improves with targeted pre-training/fine-tuning and dialect-aware evaluation (Antoun et al., 2023; Bubeck et al., 2023). Antoun et al. (2023) present more direct evidence on the limitations of the Arabic-language models and the usefulness of the specific Arabic training, whereas Bubeck et al. (2023) speak about the general capabilities of the LLM which are not directly related to the Arabic voice interaction. This comparison suggests that overall progress in the performance of LLM may not necessarily be projected to Arabic dialects, mixed-language speech, or culturally aware organisational environments. These results support the study's emphasis on dialect-sensitive evaluation for Arabic voice assistants and provenance-aware KBS integration.

Regardless of these developments, there are three gaps. To begin with, current research primarily assesses the model potential, fit, or the performance of retrieval and not the perceptions of Arabic-speaking users of the voice assistants powered by LLMs. Second, there is a dearth of research on the interaction of dialect variation, code-switching, cultural expectations, privacy, and trust within Arabic conversational systems. Third, RAG and other KBS-based methods describe how the responses can be based on the knowledge of the organisation, however, they give little evidence on how the Arabic-speaking users should access and assess that knowledge using voice interfaces. These gaps enable the research questions about the use of AIVA, its perceived advantages and challenges, effective communication and user-centred design, as well as influencing the adapted framework and the suggested conceptual blueprint of AIVA-KBS integration.

2.2 Key Cultural Values in Arabic-Speaking Communities

A study by Mirsagatova (2023) shows that Arabic-speaking societies are deeply rooted in cultural values and customs such as emphasizing hierarchy, honour, and respect, which are reflected in their linguistic expressions. The use of respectful language, including honorific titles, religious phrases, and formal address forms, underscores the importance of politeness, social status, and deference in communication. Family is the cornerstone of social structure, with strong intergenerational ties and a collective sense of responsibility shaping personal and communal interactions (Ballout, 2024). Honor, both personal and familial, plays a crucial role in decision-making and social conduct, influencing relationships, reputation, and societal expectations.

Religion and language are also fundamental to Arabic cultural identity. Islam, which guides moral and ethical values promoting peace, justice, and understanding in our diverse and interconnected world. (Mahmudulhassan, 2024). The Arabic language serves as a unifying factor across diverse regions, preserving historical and literary traditions. Respect for elders, communal solidarity, and a deep appreciation for poetry and storytelling further highlight the rich cultural heritage of Arabic-speaking societies. These values collectively foster a sense of belonging and continuity within communities.

From a systems perspective, three design consequences follow:

1. **Privacy and discretion by default.** In collectivist contexts with dense family networks, unintended disclosure (such as reading messages aloud in shared spaces) can be reputationally costly. Conservative defaults (screen-first for sensitive items, masked playback, opt-in for hands-free summaries) align with cultural expectations of discretion.
2. **Religio-cultural awareness.** Functionality that is calendar- and ritual-aware Hijri date support, prayer times, fasting-aware reminders improves perceived usefulness and cultural fit. Partnerships between content authorities and platforms (for instance, Arabic Language Centre with Alexa) illustrate institutional approaches to culturally grounded content integration.
3. **Register control.** Switching between MSA for official matters and dialect for casual queries increases acceptability. Allowing users to select a preferred dialectal persona (for example, Gulf vs. Levantine) can reduce friction.

2.3 Linguistic Diversity in Arabic-Speaking Societies

Linguistic diversity in Arabic-speaking societies is a result of historical, social, and geographical factors that have shaped the region's dialectal variations. Arabic, as a macro-language, is spoken by some 300 million people in an area spanning roughly from northwest Africa to the Persian Gulf. Traditional Arabic dialectology has dealt predominantly with geographical variation (Horesh & Cotter, 2016). A vast variety of Arabic dialects has emerged due to geographical and socioeconomic influences, thus leading to grammatical, pronunciation, and vocabulary differences between the speech communities. For instance, Mahmoud Alhilal (2024) states that standard Arabic is different from colloquial Arabic, such as Egyptian Arabic, in terms of pronunciation and vocabulary. Lebanon Arabic is spoken predominantly in Syria, Lebanon, Palestine, and Jordan. It has distinct syntax rules and delicate speech. On the other hand, Maghrebi Arabic is common in North African countries, including Tunisia, Morocco, and Algeria. It has different sentence structures and phonetics because of historical influences, as noted by Harrat et al. (2018). The research consensus is that dialect-aware modeling, domain/genre adaptation, and robust diarization are essential for production-grade performance.

Arabic's linguistic landscape is best understood as a constellation of related varieties rather than a single uniform code. Contemporary scholarship shows that historical migrations, urbanization, education policy, and media ecologies have produced layered repertoires in which Modern Standard Arabic (MSA) coexists with dozens of regional dialects that differ in phonology, morphology, lexicon, and pragmatics. Recent NLP surveys emphasize that these varieties are not merely “accents” but partially divergent systems with their own grammars, increasingly written online and in messaging, which complicates both human literacy planning and machine processing across genres (news, social, conversational) and modalities (text, speech). In practical terms, the same citizen may prefer MSA for official tasks, a regional vernacular for family coordination, and code-switch into English or French for domains saturated with loan terminology (finance, medicine, technology). Any knowledge-based interface that aspires to broad coverage in the Arab world must therefore be dialect-aware, register-sensitive, and code-switch-robust (see recent overviews 2019–2024).

Contemporary NLP surveys go beyond the traditional dialectology by not considering the varieties of Arabic as regional accents but as the systems that have to be computationally different. This point of view is closer to AIVAs as it connects sociolinguistic variation to speech recognition, intent detection and knowledge retrieval. However, a lot of this work is benchmark-focused and lacks the necessary description of how the choice of a register, use of mixed languages, or dialect mismatch influences the trust, satisfaction, or continued use of voice assistants by the users.

Over the last five years, the research community has operationalized this diversity through shared tasks that define measurable targets and public datasets. The MADAR and NADI series move beyond coarse “MSA vs. dialect” distinctions to fine-grained dialect identification at country and even city/province level, using naturally occurring social media text. NADI 2020 expanded the agenda to 21 nations and 100 provinces, standardized evaluation for sub-country variants. MADAR (ACL 2019) incorporated travel-domain and Twitter subtasks with big label sets across Arab cities. Dialectal sentiment analysis was incorporated in later rounds (NADI, 2022), representing a shift toward downstream tasks where user experience and retrieval quality are significantly impacted by dialect understanding. These benchmarks are important for deployment because they allow vendors and purchasers to compare models based on genre and dialect instead of using aggregate scores that conceal disparities.

The arc of speech technology is parallel. The problem of open-domain audio with overlapping speakers and noise was highlighted by the MGB-5 competition (2019), which expanded dialect recognition to 17 classes and moved Arabic ASR toward multi-genre YouTube audio with a Moroccan focus. Simultaneously, self-supervised and weakly supervised learning have improved low-resource recognition. XLSR-53 (wav2vec 2.0) demonstrated that multilingual pre-training on raw waveforms transfers across languages, reducing error rates when fine-tuned on modest labelled data highly relevant to under-resourced Arabic varieties. Whisper (2022) scaled weakly supervised training to ~680k hours, yielding strong zero-shot performance across many languages and accents; however, audits caution that large weakly supervised models can hallucinate content, so sensitive KBS deployments require task-appropriate safeguards and human verification.

The procurement implication is clear: insist on dialect-stratified WER/CER reporting and domain-specific fine-tuning.

Practical deployment must manage code-switching (Arabic with English or French terms in finance, medicine, and technology) and named-entity variability (for instance, foreign proper names with Arabic phonotactics). Knowledge bases that index entities across scripts (Arabic/Latin) and map common transliterations (like, Google) improve retrieval and reduce user repair cycles. For on-device dictation and short commands, mixed-language language models with pronunciation lexicons that cover loanwords can materially reduce word error rates (WER).

New shared tasks in 2025 explicitly target multidialectal ASR, spoken dialect identification, and diacritization restoration, signalling maturation of evaluation practice. KBS procurement in Arab administrations can leverage these tasks to require dialect-stratified performance reporting before adoption. The growth of common work however, is a sign of improvement in technical measurement, as opposed to successful organisational execution. Such standards can show the level of performance in the systems under specific test conditions but do not show how well Arabic-speaking users can make sense of the resulting interactions, whether they are culturally compatible, personal, or confidential, and whether they can be trusted.

A distinctive challenge in Arabic is diacritization restoring short vowels and marks typically omitted in everyday orthography. Diacritics are essential for TTS naturalness, morphological disambiguation, and ASR–NLP alignment. Since 2019, research has progressed from sequence-to-sequence baselines to Transformer-based approaches (including Arabic-BERT variants), with steady gains on classical and modern corpora; specialized efforts have emerged for Gulf Arabic, providing guidelines and new datasets that make evaluation more representative beyond MSA. In production systems, the practical lesson is to pick domain-matched diacritizers (news vs. conversational speech) and prefer context-aware models rather than purely lexical heuristics.

Code-switching within or across utterances has become normative in many Arab cities, especially in professional and youth registers. Empirical studies show that Arabic–English (and Arabic–French in the Maghreb) switching degrades both ASR and NLU if models assume monolingual input. Recent work explores contextual embeddings, hybrid decoding, and token-level language identification to mitigate the impact, with transformer ASR outperforming WFST pipelines for intra-sentential code-switching while classic decoders can remain competitive for inter-sentential switching depending on lexical coverage. For KBS retrieval, practical steps include loanword-augmented pronunciation lexicons and dual-script entity indices so that queries for “metformin” converge to the same knowledge node.

Resource building remains uneven across regions as to counter data imbalance, researchers have proposed integrated dialect datasets that merge multiple corpora to balance countries and domains for Arabic Dialect Identification (ADI) and related tasks, yet coverage is still thinner for Yemeni, Sudanese, and parts of the Maghreb compared to Egyptian/Levantine/Gulf. In order to stabilize training and benchmarking, the IADD dataset is an effort to unify multiple sources (135k+ texts) across five regions/nine nations. This is a step toward more generalizable models and more equitable evaluation. Recent surveys on the LLM side from 2023 to 2024 show that Arabic-centric and Arabic-capable models are growing quickly, but they also point out that without focused pre-training and task-specific fine-tuning, dialectal coverage, tokenization, and mixed-script handling are still fragile (Djanibekov et al., 2024). These reviews facilitate due-diligence checklists for purchasers and policy makers, including which languages were used during pre-training, how many hours or tokens each dialect requires, and what dialect-specific evaluations the vendor can provide.

Named-entity variability is another pain point where linguistic diversity meets operational need. Transliteration and cross-script mapping create fragile edges for names of people, places, and organizations. Recent multilingual transliteration work (2020–2022) shows that transformer models trained on Wikidata-scale name pairs improve cross-lingual list search and reduce out-of-vocabulary errors critical when Arabic queries use variant Latin spellings or when Latin queries require Arabic results. Complementary resources (such as, English–Arabic name transliteration/classification datasets) help normalize frequent variants, essential for KBS that must reconcile “Mohammed/Mohamad/Muhammad” or “Doha/Ad-Dawhah” with a single entity graph.

The growing assessment ecosystem makes it possible to incorporate minimal performance guarantees by task and dialect from the perspective of procurement and governance. "Models must meet or exceed baseline on NADI country-level ID" or "WER on Gulf Arabic broadcast ≤ X% at Y SNR" are examples of pre-made benchmarks that authorities can use in tenders. Other shared tasks from 2020–2022 (NADI) and continuing community efforts from 2024–2025 (dialect ID, diacritization, multidialect ASR) provide these benchmarks. In particular, for knowledge services that are accessible to the public and where misrecognitions may have service consequences, requiring dialect-stratified reporting, code-switch stress tests, and proper-name pronunciation accuracy links technical metrics with lived user demands.

However, the benchmarks of procurement mostly outline minimum technical requirements and fail to reflect a complete picture of socio-pragmatic suitability, privacy anticipation, openness, and trust in organisational data by the users. A system can achieve WER or dialect-identification targets whilst giving outdated or culturally inappropriate or poorly exposed responses. In this respect, technical acceptance criteria are to be integrated with user-centred and a governance-based assessment.

Subsequently, there are clear design implications for the sociolinguistic reality MSA for formal domains, dialects for daily life, and code-switching for specialized conversation. Dialect selection and register control are treated as first-class settings in successful deployments, provenance and update timestamps are presented to conform to institutional trust norms, and learners and older adults are given accessible options (slower TTS, diacritized output, multilingual summaries). The evaluation scaffolding (MADAR, NADI, MGB-5) and algorithms (self-supervised speech, dialect ID, diacritization, and transliteration) required to carry out this vision responsibly are provided by the last five years of Arabic NLP/ASR research, as long as deployments pledge to dialect equity, open reporting, and human oversight where mistakes carry risk.

Overall, current literature concurs that Arabic AIVAs need dialect-sensitive modelling, code-switching, diacritisation, entity resolution, domain adaptation and transparent evaluation. Nevertheless, their datasets, dialect coverage, modality, tasks, and deployment contexts are different, and hence, their results cannot be the same evidence of full functionality. Three gaps exist: there is a lack of evidence on how Arabic speaking users in the UAE view linguistic challenges; a lack of analysis of how language performance interacts with culture, privacy, trust, accessibility and continued use; and a lack of integration between speech processing and organisational KBS functions. The gaps used in the research questions and to aid the adapted AIVA-KBS framework to be used in the study in future.

2.4 Technology and Cultural Interactions

Advancements in the field of digital technology have increased cultural interconnectedness by establishing continuous connectivity and bridging geographical distances all over the world. Online social platforms like social media, online video chat, and instant messaging have enabled individuals from all backgrounds to communicate and exchange thoughts instantly. This has led to the phenomenon called diffusion of cultures, where the world has become a connected global village because of blending of customs, lifestyles, and languages (Huang, 2024). In addition, AI and virtual reality innovations have also enhanced the level of cross-culture experience by offering chances for individuals to experience different culture setups virtually, avoiding the need for physical travel (Litt et al., 2020).

Huang (2024) underlines the importance of digital connectivity to speed up cultural diffusion, but Litt et al. (2020) pay more attention to immersive technologies and online cultural experiences. Even though both studies conclude that technology can broaden intercultural interactions, they analyze various types of interactions and fail to determine whether digitally mediated experiences can maintain the depth of cultural meaning as face-to-face communication.

However, while technology promotes cultural exchange, it also presents challenges such as cultural homogenization and the dominance of certain cultural narratives over others. For instance, the global reach of the influence of the Western media has caused the degradation of local languages and cultural concerns. Additionally, digital interactions may sometimes lack the depth of in-person communication, impacting the authenticity of cultural exchanges. Nevertheless, when implemented cautiously, the potential of technology is great for the conservation of culture alongside the exchange of ideas and thoughts across diverse communities (Salehan, Kim, & Lee, 2018).

Digital platforms both bridge and flatten cultures as they enable cross-cultural exchange, yet can promote homogenization and marginalize local narratives (Smairi, 2023). For Arabic, under-representation in the digital sphere remains a barrier to inclusive

AI, a point emphasized by regional AI institutes driving Arabic-first NLP and dataset creation. Digital technologies have transformed how cultures circulate, interact, and change, creating transnational publics where symbols and norms move quickly across languages and borders (Chen & Wei, 2024). Extended reality (XR) now supports embodied encounters with heritage and cultural practices, a trend documented in HCI as a driver of intercultural empathy and learning when accessibility and context are designed (Noe, 2021).

Salehan, Kim, and Lee (2018) present technology as capable of supporting both cultural conservation and exchange when implemented carefully, while Smairi (2023) places greater emphasis on the risks of homogenization and the marginalization of local narratives. Chen and Wei (2024) further highlight the rapid circulation of cultural symbols through digital platforms, whereas Noe (2021) identifies more positive outcomes from immersive cultural experiences. These contrasting perspectives indicate that technology can strengthen or weaken cultural representation depending on platform design, content diversity, and the degree of contextual adaptation.

Platform logics can also flatten difference: engagement-optimized ranking promotes conventional aesthetics and majority languages, contributing to a known “digital language gap” where Arabic’s share of web content trails well behind its speaker base. Lack of Arabic (and many dialects) results in inferior training coverage and benchmarks, which feeds back into a feedback cycle of decreased accuracy and usage in Arabic communities and this is because AI learns from vast amounts of online data (AlAfnan, 2025). Additionally, common assessments for Arabic (such as MADAR and MGB-3) consistently demonstrate that dialectal variability lowers ASR/NLU performance in comparison to MSA, particularly when genre change is present (like YouTube discussion vs. broadcast news). This can be combated by institutional and policy initiatives, such as demanding dialect-stratified reporting in bids, licensing authoritative content to assistants, and procuring Arabic corpora. The Abu Dhabi Arabic Language Center's collaboration with Alexa is an example of this content-provenance approach.

Overall, the literature shows that digital technologies can promote cultural exchange, accessibility, and intercultural learning, but may also reinforce majority languages, Western-oriented narratives, and cultural homogenization. A significant gap remains because existing studies provide limited empirical evidence on how Arabic-speaking users in the UAE perceive these cultural effects in relation to AIVAs. The literature also does not sufficiently explain how cultural representation, Arabic under-representation, dialect variability, and institutional content governance jointly influence trust, continued use, and access to organisational knowledge. This gap informs the research questions concerning daily AIVA use, perceived benefits and challenges, effective communication, and culturally appropriate design. It also supports the adapted framework by linking cultural suitability, design preferences, institutional communication, and perceived AIVA importance to the proposed AIVA–KBS integration blueprint.

2.4.1. Role of AI and Digital Technology in Mediating Cultural Differences

AI and computer technology hold a solid position in mediating culture disparities through the capacity for communication, greater understanding of various cultures, and filling linguistic gaps. Dwivedi, Sahu, & Srivastava (2024) state that AI-based language tools, for example, Google Translate, enable individuals from different linguistic backgrounds to communicate effortlessly. Similarly, AI recommendation platforms personalize content across different cultural backgrounds; thus, the audience gets engaged by relevant cultural and preferential information. In addition, according to Chen and Wei (2024), Virtual reality (VR) and augmented reality (AR) platforms also make exchange possible where the audience gets to experience and perceive different traditions, histories, and societal norms. AI language tools, recommender systems, and XR experiences can tailor content across cultures. In Arabic contexts, high-impact use cases include government service triage, education, health information access, and media accessibility provided systems are dialect-aware, bias-audited, and subject to human oversight in consequential decisions, as called for by UNESCO's Recommendation.

Dwivedi, Sahu, and Srivastava (2024) emphasise the role of AI language tools in reducing linguistic barriers, whereas Chen and Wei (2024) focus on immersive technologies that expose users to cultural traditions and social practices. Although both studies present technology as a means of supporting intercultural understanding, they examine different forms of mediation. Language tools prioritise communication accuracy, while VR and AR emphasise experiential engagement. However, neither perspective fully explains whether these technologies preserve culturally specific meanings when adapted to Arabic-speaking users or organisational service environments.

Furthermore, social networking platforms combined with AI tools serve as one of the main elements for building global interactions through the facilitation of communication and prevention against cultural misunderstandings. AI models suggest suitable discussion topics for users after examining their cultural dynamics, ethnic practices, as well as preferences, minimizing the biases, and optimizing cross-cultural interactions (Xia et al., 2024). Besides, the research study by Monroy-Osorio (2024) evaluated that AI-assisted sentiment analysis is applied by business corporations and organizations for the understanding of cultural diversity and the formation of business models for different marketplaces. Though AI and digital tools bring significant advantages for filling the cultural gaps, ethical perspectives, including AI models being biased, and data security must be taken into consideration for respectful cultural representation.

Xia et al. (2024) consider how AI can personalise interactions and reduce cultural misunderstandings, whereas Monroy-Osorio (2024) evaluates the organisational use of sentiment analysis to interpret cultural diversity and support market decisions. The former focuses more on interpersonal communication, while the latter examines commercial application. Nevertheless, both approaches depend on culturally representative data, and neither eliminates the risk that biased models may misinterpret minority dialects, values, or communication styles. Their findings therefore support cultural adaptation but also demonstrate the need for stronger privacy, bias, and governance safeguards.

In principle, these capabilities advance participation and pluralism; in practice, benefits hinge on three conditions: (i) linguistic adequacy (dialect and code-switching support), (ii) bias auditing and evaluation, and (iii) human oversight for high-stakes decisions, aligning with UNESCO's Recommendation on the Ethics of AI (2021) and the NIST AI Risk Management Framework (Leopard, 2025; AI, 2023). Vendor production narratives for Arabic illustrate the complexities of cultural fit: decisions about whether assistants should speak MSA, a Gulf vernacular, or a hybrid; how to localize humor, politeness strategies, and forms of address; and how to accommodate real-world code-switching without eroding clarity. Public-sector deployments in the Gulf (for example, bilingual voice triage in utilities and city services) show voice becoming a legitimate interface for civic knowledge, while raising governance questions about provenance, auditability, and equitable service across dialect communities (McCracken & Hogan, 2021). In education, Arabic-first assistants can align with national curricula, incorporate Hijri calendrical literacy where relevant, and adjust register to context; however, they must expose authoritative citations and update timestamps, particularly for policy or religious content. In media and health access, Arabic speech/topic modelling can improve subtitling, audio description, and service quality provided domain ontologies and controlled vocabularies are used to prevent drift and culturally inappropriate responses.

Overall, the literature demonstrates that AI can mediate cultural differences through translation, personalisation, immersive experiences, sentiment analysis, and voice-based public services. However, existing studies remain fragmented across communication, commercial, educational, and government contexts and provide limited empirical evidence from Arabic-speaking users in the UAE. They also do not sufficiently explain how linguistic adequacy, cultural appropriateness, privacy, bias, provenance, and human oversight jointly shape trust and continued use. This gap informs the research questions concerning AIVA usage, perceived benefits and challenges, effective communication, and culturally appropriate design. It also supports the adapted framework by linking cultural suitability, design preferences, institutional communication, and perceived AIVA importance to the proposed conceptual AIVA–KBS integration blueprint.

2.4.2. Cultural and Linguistic Biases in AI

AI systems, including translation tools and linguistic models, generally reflect the linguistic and cultural prejudice present in the data upon which the AI has been trained. These biases can occur in several forms, for instance, by preferring one dialect over the others, misinterpreting expressions peculiar to culture, or by transferring stereotypes over into automated recognitions and translations. For instance, as AI translation platforms generally struggle while coping with Arabic dialects, they prefer either MSA or the generally spoken dialects such as Egyptian Arabic, thus marginalizing less dominant linguistic groups (Bella et al., 2024). Likewise, the findings of Navigli et al. (2023) depicted that AI text can also reflect historical and cultural biases when projecting various societies and identity groups using virtual communication.

These biases arise due to the dominance of certain languages and cultures in training datasets, as well as the lack of diverse linguistic and cultural representation in AI development. Reinforcing the said biases can influence the social dynamics, resulting in digital disparities and limiting access to the technology for underrepresented dialect groups (Boztemir & Çalışkan, 2024). To counter this, developers must ensure the curating of the data for the AI models, greater diversity when building the AI models, and continuous algorithm improvement for fair and representative linguistic treatment across different cultures (Gallegos et al., 2024). Bias arises when training data under-represent dialects or encode stereotypes. For Arabic, mainstream systems often privilege MSA or a single dialect (such as, Egyptian or Gulf). Shared-task outcomes (MADAR, MGB-3) offer concrete evaluation protocols accuracy for dialect ID and WER/CER for ASR by dialect and genre that should become standard in public procurement to avoid systematic exclusion of dialect communities.

Beyond model performance, bias appears in content routing: for example, an assistant that defaults to non-regional knowledge sources for religious or civic queries can erode trust. KBS-integrated assistants should employ provenance controls (source whitelists, last-reviewed timestamps) and provide explanations ("This answer is based on [ministry portal] updated [date]"), aligning with transparency principles in global AI governance. UNESCO (2021) and NIST's AI Risk Management Framework (AI RMF 1.0) both emphasize documentation, measurement, and continuous improvement cycles relevant to such deployments.

Empirical campaigns such as MADAR and MGB-3 provide the clearest evidence base and measurement instruments: dialect ID accuracy and WER/CER broken down by dialect and genre. Treating these as minimum baselines in procurement helps prevent systematic exclusion of dialect communities. Bias also presents in moderation and source selection: if a system defaults to non-authoritative or non-regional sources for sensitive queries, trust degrades even when recognition is accurate. Mitigation should be structural and technical. Structurally, states and institutions can publish and license high-quality Arabic corpora (including dialectal speech and code-switching), fund benchmarks, and mandate dialect-stratified reporting in tenders. Technically, developers should implement front-end dialect identification and routing; augment acoustic and lexical models with dialectal lexica; adopt multilingual pretraining followed by fine-tuning on curated dialectal data; and use active learning with human-in-the-loop corrections to reduce unfair performance gaps.

Documentation and transparency practices datasheets/model cards stating dialect coverage and failure modes; evaluation cards reporting stratified metrics on public benchmarks; and change logs for knowledge bases enable external audit and iterative improvement, consistent with UNESCO and NIST guidance. As Arabic content grows inside mainstream assistants through curated partnerships and as evaluation norms mature, user trust and participation can increase, producing more training data and better dialectal coverage a virtuous cycle. Sustained Arabic-first corpus development, open benchmarks, and provenance-conscious KBS integration are therefore central to both technical performance and cultural self-determination.

2.5 Evolution of AI Voice Assistants

AI voice assistants are now far more developed since their inception. Soofastaei (2021) delves into the history of the progress of VAs from their beginnings in the 1910s through their contemporary applications. It discusses the contribution of AI towards the advancement of VA features and analyses the economic importance of the VA for individuals and corporations. Li (2023) has also explored the history of the progress of automatic speech recognition (ASR) and its contribution towards shaping data-based AI. The study discusses the transformation of ASR from the human cognition imitation to the large-scale data collection reliance, shaping modern algorithm culture. During the 1960s, the early AI voice technology developments involved systems like the Shoebox from IBM, which could recognize and acknowledge 16 spoken terms (Kingaby, 2021). The real breakthrough occurred when Apple introduced Apple's Siri in the year 2011, the first extensively adopted AI voice assistant deployed on consumer devices. Since its beginning, Siri has progressed, and over the next year, Siri expanded its voice features and became multi-language, being able to recognize Chinese, Japanese, and Russian languages. To serve a wider user base, Siri entered emerging markets such as India and the Middle East in 2014 (Cao, 2024). From early command recognizers to cloud-backed, context-aware agents on phones, smart speakers, and vehicles, AIVAs now integrate ASR, NLU, dialogue management, and TTS across ecosystems.

Soofastaei (2021) presents the evolution of voice assistants from a broad historical and economic perspective, whereas Li (2023) focuses more specifically on the development of ASR and the shift towards data-driven AI. Kingaby (2021) and Cao (2024), by contrast, illustrate this progression through particular technological milestones, from IBM's limited-vocabulary Shoebox to Siri's multilingual consumer deployment. Although these studies collectively demonstrate substantial technological advancement, they mainly describe historical progress and do not establish whether such developments have produced equal linguistic and cultural accessibility for Arabic-speaking users.

Arabic support has moved from limited MSA coverage to commercial deployments with dialect sensitivity, notably Amazon Alexa's Arabic releases for KSA and UAE, vendor reporting on "how Alexa learned Arabic" (data strategy, dialect policy, content partnerships), and local content arrangements (as ALC–Alexa) to increase domestic relevance. While progress is clear, parity with English remains a moving target given Arabic's dialectal granularity and code-switching prevalence.

Following Siri, other major tech companies developed their own versions, including Google Now in 2012 and Microsoft Cortana in 2013 (Kendall, 2020). These assistants utilized natural language processing (NLP) and machine learning to understand and respond to user commands, significantly enhancing user interaction with technology (McLean & Osei-Frimpong, 2019). Among the growing frontline of AI-based technologies, voice assistants are becoming ubiquitous (Kendall, 2020).

Kendall (2020) emphasises the widespread diffusion of commercial voice assistants, whereas McLean and Osei-Frimpong (2019) evaluate their role in improving user interaction through NLP and machine learning. The former highlights market expansion and technological ubiquity, while the latter focuses on the quality of human–technology engagement. However, neither perspective sufficiently examines whether commercial growth and improved interaction are sustained across languages with complex dialect structures, particularly Arabic, or whether localisation efforts are perceived as effective by users in the UAE.

AIVA has made a significant impact to end user engagement (McLean & Osei-Frimpong, 2019); (Sung et al., 2021) and customer service domain interaction (Kumar, 2016). In 2011, the voice assistant movement started to develop as Apple introduced Siri. Siri is a virtual assistant that includes the voice assistant that they pinned on their iOS and macOS products. In the following year, 2012, Google launches its virtual assistant called Google Now, which has voice assistant technology embedded in it. In 2013, Cortana was released by Microsoft. These three are similar products with just different packaging according to the respective developing companies (Grimsley, 2007).

Overall, the literature demonstrates that voice assistants have evolved from limited command-recognition systems into widely adopted, context-aware platforms that support user engagement and customer-service interaction. However, much of this research remains historically descriptive, commercially focused, or centred on English-dominant systems. Limited evidence explains how Arabic-speaking users in the UAE experience these technologies, particularly in relation to dialect sensitivity, code-switching, cultural appropriateness, trust, and sustained use. The literature also provides insufficient explanation of how mature commercial AIVAs can operate as interfaces to organisational KBS. These gaps inform the research questions concerning daily usage, perceived benefits and challenges, effective communication, and culturally appropriate design, while supporting the adapted framework and proposed conceptual AIVA–KBS integration blueprint.

Figure 2-1. AI voice assistants' applications

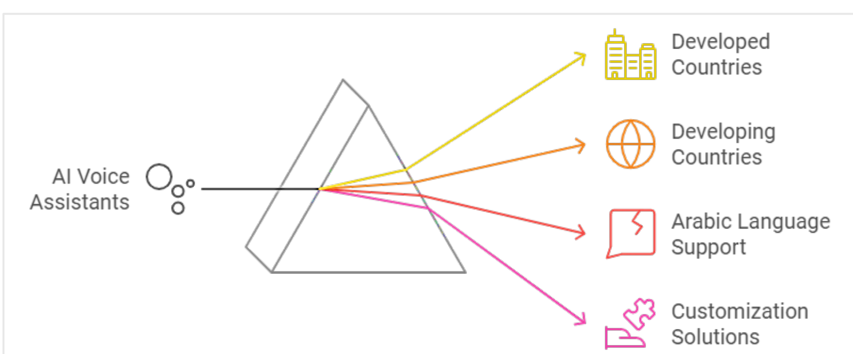


Figure 2.1 illustrates the diverse applications of AI voice assistants where they serve as the input, which splits into four distinct areas of focus. The first focus area is

developed countries, where AI voice assistants are widely used for automation, customer service, and smart home applications. The second area highlights their growing potential in developing countries, where they can enhance accessibility, education, and business processes. Another critical focus is Arabic language support, emphasizing the importance of AI voice assistants catering to Arabic-speaking users. Finally, the image highlights customization solutions, showcasing AI voice assistants tailored to meet industry-specific or business-specific requirements. This visualization suggests that AI voice assistants are evolving beyond traditional applications, addressing linguistic, regional, and specialized needs in a more targeted manner.

2.5.1. Enhancements in Multilingual Support and Context Awareness

Patil et al. (2024) present the enhancements made for multilingual support, context awareness, and speech recognition. The enhanced context awareness of the modern voice assistants is the key factor for their capacity of providing more relevant responses. The enhanced context awareness of the modern voice assistants is the key factor for their capacity to provide more relevant responses, and at the same time, multi-language support has also made these systems more globally accessible. Apart from this, compatibility with IoT devices allows them to automate the operation of smart home devices easily through voice inputs, and machine learning allows them to customize the response for each person, thus bringing an overall improvement in interaction quality. Olawade et al. (2025) highlight the advancements of AI voice assistants by stating their integration in smart homes which help people with physical disabilities manage their daily routines more effectively such as controlling lights, appliances, and other home systems. Modern assistants leverage device fusion (phones, vehicles, IoT), personalization, and on-device adaptation for latency/privacy. For accessibility, vendors highlight screen-free interactions and auditory guidance.

Patil et al. (2024) emphasise multilingual support, context awareness, and personalisation, whereas Olawade et al. (2025) focus more specifically on accessibility and smart-home assistance for people with physical disabilities. Although both studies demonstrate the practical value of advanced voice interfaces, they assess different outcomes. Patil et al. examine interaction quality and technological capability, while Olawade et al. concentrate on functional independence. Consequently, improvements in accessibility cannot automatically be treated as evidence of equal multilingual or dialect-sensitive performance.

Dubai Electricity and Water Authority's Rammas (see Appendix C) integrated generative AI in 2023; by mid-2024, officials report over 9.6 million enquiries handled since launch and satisfaction rates around 95%, suggesting operational maturity in bilingual (Arabic/English) contexts (Government of Dubai., 2021). The Abu Dhabi Arabic Language Centre–Alexa partnership (2025) injects curated Arabic content and experiences into domestic assistants in KSA and UAE, foregrounding the role of cultural authorities in shaping assistant usefulness. Research programs and doctoral work on multi-dialect Arabic ASR and broadcast speech provide replicable baselines and recipes, bridging lab and production (Salhab et al., 2025).

The article by Benjamin (2025) highlights advancements in AI and NLP that enable assistants like Amazon Alexa, Google Assistant, Apple Siri, and Microsoft Cortana to improve customer interactions, automate sales processes, and integrate with CRM systems. It examines their role in enhancing personalization, predictive analytics, and real-time data access while addressing challenges such as data privacy and integration complexities. Microsoft's Copilot voice capabilities are enhanced for users with visual impairments by providing auditory assistance for information retrieval and library navigation (Adetayo, Babalola, & Olukayode, 2025).

Kumar et al. (2024) highlight key advancements in Alexa's voice assistant capabilities, emphasizing how AI-driven improvements enhance user satisfaction. The regression analysis revealed ten user satisfaction predictors of the Amazon Alexa App, with a particular focus on the voice shopping experience. Alexa aids caregivers in catering to the needs of elderly individuals by providing convenient shopping, voice-activated control, reminders, smart home integration, and access to information. This shows more refinement in Alexa's speech recognition and natural language processing (NLP), enabling more accurate and context-aware responses. These improvements align with ongoing efforts to refine AI-driven speech recognition technologies, as recent research highlights the role of AI in bridging multilingual communication barriers while acknowledging the challenges posed by dialectal variations (Besdouri et al., 2024).

Benjamin (2025) evaluates the organisational benefits of AIVAs through customer service, CRM integration, and predictive analytics, whereas Adetayo, Babalola, and Olukayode (2025) focus on accessibility for visually impaired users. Kumar et al. (2024), by contrast, examine satisfaction with Alexa, particularly in voice shopping, while Besdouri et al. (2024) highlight broader multilingual and dialectal barriers. These studies collectively demonstrate improved functionality, but their different sectors, users, and outcome measures limit direct comparison. High satisfaction in shopping or accessibility contexts does not necessarily establish reliable performance for Arabic dialects or organisational KBS access.

Companies other than Apple, Google, and Amazon are also offering their own voice assistants which provide technology and innovation to help businesses in multiple industries such as automakers, restaurants, hospitality, retail. One notable example is the partnership between Mercedes-Benz and SoundHound AI for the creation of the in-vehicle voice multimedia system MBUX (Leonidou & Kypros, 2024). SoundHound AI mentioned, "During each stage of developing their voice assistant, the Mercedes-Benz team kept their focus on two important elements: their customers and their brand" (SoundHound AI Inc., 2024). Likewise, Hyundai also integrated SoundHound AIVA to support a "Car-to-Home" service, enabling the driver to control smart devices at home through the car using simple voice commands (SoundHound AI Inc., 2024).

Additionally, Mirai, a wearable AI system, is designed to act as an "inner voice" by providing proactive and personalized guidance based on a user's environment and behavior. It features real-time speech processing, an integrated camera for contextual awareness, and voice-cloning technology that mimics the user's voice to make reminders feel more natural and engaging. These advanced voice features allow Mirai to deliver subtle nudges for improving individual's habits like diet, productivity, and communication (Fang et al., 2025). Expedia, also developed its own AI voice assistant. Expedia's Romie AI assistant, which helps with planning and booking a trip and provides step by step recommendations (Expedia Group Inc., 2024).

Manolescu, AlZu'bi & Secco (2025) developed an IoT based AI voice system that demonstrated efficient and reliable human-robot interaction. It effectively captures user voice inputs, processes them through advanced AI models, and generates appropriate commands for the robotic arm, achieving an average voice-to-motion latency of 5.5s. The commands given are “engage arm”, “move right 20” combined with more conversational commands such as “can you 19 hear me?”, “what’s your name?”. Similarly, a recent work by Rath et al. (2025) presents a real-time voice sentiment analysis system by leveraging natural language processing, sentiment analysis models, and speech recognition tools. It is designed to classify spoken language into three groups of sentiments; neutral, positive, and negative. In the context of voice assistants, this system enhances their ability to understand user emotions and improve interactions. The paper highlights support for multiple languages, specifically English, Hindi, and Telugu, making it more inclusive and effective for diverse users.

Despite these advancements, challenges persist. AI-driven voice assistants often struggle with multilingual users, particularly those various dialects. For example, Leblebici (2024) discusses several shortcomings of voice assistants by focusing on language limitations and biases. It highlights that voice assistants like Siri, Alexa, and Google Assistant are predominantly trained in standardized English, which makes them less effective for users who speak non-European languages, such as Turkish. This results in misinterpretations, inaccurate responses, and difficulties in natural communication for non-English speakers, leading to accessibility and usability issues for a diverse global audience.

Moreover, Pawar & Patil (2024) examine various methodologies for identifying dialects in spoken language for both Indian and non-Indian languages such as deep learning, transfer learning, classical acoustic feature analysis, machine learning algorithms, phonetic and phonological analysis, lexical and grammatical feature extraction, and prosodic analysis. It highlights challenges in dialect identification, particularly for Indian languages, due to limited annotated datasets, multilingual influences, and regional variations. It concluded that though significant language dialect identification (LDI) advancements exist for non-Indian languages, research for Indian dialects, especially Marathi, remains ongoing and faces difficulties in distinguishing closely related dialects.

Research has demonstrated that speech recognition models frequently fail to accurately process dialectal variations, particularly for languages such as Arabic, where speakers often switch between Modern Standard Arabic (MSA) and regional dialects. For instance, taking into account the target audience and contextual factors, they switch from formal Modern Standard Arabic (MSA) to local dialects (i.e., Gulf Arabic, Egyptian, or Levantine). AI voice interfaces can accurately deal with MSA, the formal Arabic dialect spoken by the mass media and formal writing, but not necessarily recognize the dialect-specific terms, pronunciation shifts, or the mixed speech incorporating English or French loanwords. This weakness highlights the need for AI enhancement for the treatment of the different forms of the Arabic language and mixed-language interactions for inclusive support for the Arabic audience (AlAfnan, 2025; El Zahraa, 2025).

The literature indicates that multilingual support, contextual awareness, personalisation, accessibility, IoT integration, and organisational connectivity have expanded AIVA capabilities. However, studies differ in language, users, tasks, sectors, and evaluation methods; therefore, their findings cannot be treated as equivalent evidence of overall effectiveness. Three gaps remain: limited evidence on how Arabic-speaking users in the UAE perceive multilingual and dialect-related problems; insufficient examination of context awareness, privacy, cultural appropriateness, accessibility, and continued use together; and weak understanding of how AIVAs should retrieve and communicate verified organisational knowledge through KBS. These gaps inform the research questions and support the study's adapted AIVA–KBS framework.

2.6 Current State of AI Voice Assistant Usage in the Arabic-speaking communities

Both Salih (2025) and Al Shamsi (2021) point out that the acceptance of AI-enabled assistants is not limited to technical availability. The reaction of users towards these technologies is influenced by trust, perceived usefulness, enjoyment, facilitating conditions as well as privacy related issues. Nevertheless, the works are contextually and outcome-differing. Salih (2025) analyzed the continuation of services in Oman, and Al

Shamsi (2021) studied the adoption of the same among university students in the UAE. Their results thus cannot be applied to all Arabic-speaking users or organisational contexts. In addition, both studies do not incorporate the issue of Arabic dialect, cultural demands, governmental communication and access to organisational knowledge in a single framework. This shortcoming justifies the wider model of AIVA significance of the present study in the context of Arabic-speaking users in the UAE.

A study by Abdelraouf (2024) explores the AI integration into audio-visual media platforms in Oman with the focus on the perspective of media professionals, underscoring the role of AI-driven digital tools in providing system proficiency by enabling task automation, thus saving time and chances of error encounters. Narrow data resources available for Arabic language and discrepancies in Arabic speech recognition are the two key challenges in this context other than algorithmic partiality. However, the steady shows the positive attitude towards AI adoption, stressing the need for improvement in AI intelligence systems, specifically in the context of Arabic media. Bouzid (2023) also focuses on the adoption of AI voice assistance speech technology in the Arab regions. It states that efforts are being made in data collection model training, Arabic language limitations, and resource infrastructure. Further, rapid developments in AI-powered voice applications in government services for customer support are also being made, showing a growing usage of AI-based applications across the region. While challenges related to the contextual accuracy for different Arabic dialects as well as algorithm biases in multilingual speech processing have been stated.

A proposed system called Arabic Personalized Control of Things (APCoT) enables home automation based on Arabic voice commands according to user preferences. This system has been designed for elderly individuals who prefer the mother tongue for communication. A broader data set contains Arabic voice interactions beyond simple on-off commands with integration of speech recognition, IoT, and infrared controlling of non-smart devices. For this, machine learning techniques have been used. Which improves Arabic voice recognition and face and hand gesture identification with high accuracy. Experimental results for such a system show that Arabic voice recognition turned out to be 99.63% for a single speaker and 94.4% for multiple speakers with an efficient response time (Salah, 2023).

Furthermore, research has been done on AI-driven Arabic language learning in society. 5.0 era, which focuses on incorporating AI tools such as speech recognition to improve Arabic learning. Emphasis has been put on the development of AI applications dealing with Arabic linguistic variations and understanding of cultural contexts. Such smart digital tools like voice assistants and intelligent tutoring systems (ITS) offer user-specific learning (Anwar & Ahyarudin, 2023).

There has been seen an increasing trend in the integration of AI-powered, intelligent voice systems in the government sector in Arabic countries. For example, Bahrain's government has deployed an AI-driven intelligent agent called "Ali" in the call centre for improvement in government service interactions (Elmedany, 2024). It uses natural language processing, machine and deep learning, and speech recognition, which handles Arabic and English queries while conversing. Its functions include answering questions, directing requests, executing actions, and scheduling appointments with the security framework that ensures customer data privacy in compliance with Bahrain's regulations. Its development stages include transitioning from a text-based chatbot to an AI-powered voice assistant intelligent agent that replaces certain tasks of human agents.

A complementary thread is Arabic in higher education and youth segments, where assistants function as study companions and campus service interfaces. Here, perceived usefulness and enjoyment support adoption, but long-term retention appears contingent on better handling of colloquial speech, mixed-language queries, and localized content (faculty names, administrative forms). A third thread is household accessibility: in multi-generational homes, hands-free controls and reminders offer benefits to elders, provided ASR tolerates age-related voice changes and regional accents. The strongest gains arise when voice interfaces integrate with Arabic-capable home automation (lighting, HVAC) and fall back to visual confirmations to prevent unintended actions—a design pattern observed in automotive and smart-home deployments internationally and applicable to Arabic markets.

The UAE, being the leader in AI-based technology deployment in the government and private sectors, is using AI-powered voice assistant systems in the government communications sector for automation of services and public engagement to provide effective real-time responses. Virtual assistants and chatbots used in public services process questions and share related information with less response time. Moreover, AI-based data analysis and monitoring of social media also enhance smooth communications between the government and Arabic-speaking users. Although transparency, data, security, and ethical use of AI have also been debated, training professionals and developing AI policies for response assurance and effective implementation have been emphasized (Rahman et al., 2024; Al Ali & Khalil, 2025).

2.7 Current Research on AI Personalization for Arabic Users

AI voice assistants have made great progress in Arabic-speaking communities because of their high usage, improving how people communicate and interact with technology, such as in the UAE. Assistants like HUAWEI’s Celia, Google Assistant, and Siri now better recognize Modern Standard Arabic and some dialects. These assistants also enable voice commands, smart homes, and customer support automations and can be utilized by individuals and business organizations. There still, however, exist challenges ahead, including limited coverage of Arabic dialect, understanding of context, and complexity of Arabic morphology, affecting response generation accuracy and voice identification. However, the improved recognition capabilities of global assistants should not be treated as equal evidence of effective Arabic personalisation. Systems such as Celia, Google Assistant, and Siri differ in their language coverage, regional adaptation, service integration, and available dialect support. Although these systems demonstrate progress beyond English-dominant interaction, their ability to recognise Modern Standard Arabic does not necessarily indicate consistent performance across Gulf, Egyptian, Levantine, or Maghrebi varieties. This distinction suggests that language availability and culturally responsive personalisation should be evaluated separately.

AI voice assistants have made major advancements in Arabic-speaking populations, with more enhanced user experiences from better language coverage and cultural adaptation. Firms are now more actively engaged in upgrading natural language processing (NLP) for Arabic, with more accuracy and cultural awareness in speech generated by AI (AlAfnan, 2025). AI-based Arabic assistants are also being used in government services to improve communication and efficiency. Recent research points towards the way Arabic voice assistants with artificial intelligence are being made more able to comprehend linguistic variations and regional dialects. Other research on AI-driven Arabic virtual assistants for governmental services also shows that natural language processing advancements are streamlining accessibility and efficacy in governmental engagements for Arabic-speaking populations (Alkarkhi et al., 2024). Inclusion of machine learning in Arabic assistive technology has also promoted enhanced user experiences in education and accessibility software (Elsheikh, 2023).

A study by Jaradat, Alzubaidi, & Otoom (2022) focuses on the voice-controlled car assistant for Arabic-speaking people with the goal of reducing driver distraction via voice input in Arabic, therefore enhancing safety and convenience. However, it lacks discussion on several key aspects crucial for the smooth use of voice assistants in Arabic, like how the assistant adapts to different dialects beyond Modern Standard Arabic. Additionally, it overlooks accuracy of speech recognition, such as variations of pronunciations and phonetics and the complexity of diacritics, in understanding AI systems.

The literature indicates that Arabic AIVA personalisation has advanced through improved NLP, language support, cultural adaptation, machine learning, and sector-specific applications. However, existing studies remain fragmented across consumer, government, educational, accessibility, and automotive contexts. They provide limited comparative evidence concerning dialect adaptation, code-switching, privacy, contextual understanding, and sustained user acceptance. Jaradat, Alzubaidi, and Otoom (2022), for example, demonstrate a specific safety application but do not address broader dialect or cultural requirements. These gaps inform the research questions concerning AIVA usage, perceived benefits and challenges, effective communication, and user preferences. They also justify the adapted framework linking linguistic and cultural requirements, design preferences, institutional communication, and perceived AIVA importance to the proposed conceptual AIVA–KBS integration blueprint.

2.7.1. Arabic AI Voice Assistants

Several Arabic AI voice assistants have been developed to cater to different needs and regions. These include Arabot, Jordanian voice assistants for Arabic dialects, Egypt's voice assistant solution, RoboMatic. AI, and Qatar Computing Research Institute's (QCRI) QVoice and Fanar for training and identification of language and dialect. Salma.ai, the first

Arabic voice assistant with the capacity of understanding varying dialects, and RafiQ(see Appendix C) focused on Egyptian Arabic. Rammas, customer service voice assistant from Dubai Electricity and Water Authority (DEWA), and Yasmina, voice assistant from Dubai, are tailored for Khaleeji accents. While VoxArabica and AraSpot have been developed for serving dialect identification (DID) and automatic speech recognition (ASR) purposes. Each of them exhibits Arabic speech identification and processing improvements with their own sets of challenges, including Arabic dialect variations and context understanding.

The Rammas case demonstrates that when a voice/chat assistant is attached to enterprise workflows (billing, outage reporting, appointment scheduling) and a governed content backbone, it can scale while maintaining satisfaction an important signal for ministries and utilities considering Arabic voice channels for Arabic-speaking users services (Alsulami & Mehmood, 2018). On the research side, the MGB challenges (broadcast, multi-genre) continue to provide training and test sets that cover a breadth of dialectal variation and acoustic conditions. These are precisely the resources to bake into procurement and acceptance testing for Arabic voice assistants connected to KBS (Miao & McLoughlin, 2019).

RoboMatic.AI is a voice and chatbot solution provided by Egypt's InfraDrive, a software company, for use on both personal and professional purposes. The voice assistant, with the execution of voice commands and speech-recognition abilities, can be employed by the users for engaging in hands-free tasks such as questioning, controlling IoT devices, and process automations. It also supports multiple languages, including Arabic, making it accessible to a wider audience. However, despite its Arabic support, it has limitations such as inconsistent NLP accuracy, struggles with dialects and complex sentence structures, and voice recognition challenges across different Arabic accents (RoboMatic, 2025)**.** Further, Arabot, a Jordanian-based AI tech company founded in 2016, developed AI chatbot and voice assistant that specializes in conversational AI for Arabic speakers and is capable of handling queries in a number of regional Arabic dialects. Arabot aims to enhance customer experience through personalized and efficient interactions (*ARABnet*). However, it primarily focuses on text-based chatbots, and its voice assistant capabilities are limited or not widely recognized.

QCRI has developed voice-enabled systems, notably QVoice and Fanar. QVoice is designed to assist both non-native and native Arabic speakers in improving their Modern Standard Arabic pronunciation while also helping native speakers reduce the impact of regional dialects on their Modern Standard Arabic (MSA) speech (Kheir et al., 2023). However, it primarily focuses on assisting learners in improving their Modern Standard Arabic (MSA) pronunciation, without addressing variations in Arabic dialects influenced by different cultural backgrounds. According to Team et al. (2025), Fanar, is designed to understand a wide range of Arabic dialects and incorporates an understanding of Arab cultures and Islamic values. With two of its own large language models (LLMs) called Fanar Star and Fanar Prime, it has been pre-trained on a huge corpus of Arabic, English, and code tokens. In terms of voice assistance, it has in-house bilingual speech recognition system capable of understanding multiple Arabic dialects, thereby enhancing its applicability across diverse Arabic-speaking populations. Additionally, Fanar made the voice generation possible with regional-specific characteristics, thus more natural and locally viable conversations can be achieved.

Al-Fraihat et al. (2023) discuss Arabic speech recognition (ASR) enhancements along with common challenges, including dialectal and linguistic diversity present in the Arabic regions. It examines the integration of smart systems for enhancing Arabic speech technology, specifically automatic speech recognition and dialect identification, by mentioning VoxArabica, an Arabic voice assistant that utilizes models such as Whisper XLS-R and HuBERT for speech recognition and dialect identification. It has been trained on recognizing 17 Arabic dialects apart from Modern Standard Arabic (MSA) with ASR models optimized for Moroccan, MSA, Egyptian, and mixed data. VoxArabica has a user-friendly web interface supporting model selection, file uploading, audio recording, and error detection, thereby providing a solution to Arabic speech recognition systems generalization across several dialects. However, the developers of VoxArabica have mentioned certain limitations, such as poor performance with unfamiliar data and the processing of non-linguistic factors like gender and recording conditions by dialect identification (DID) models may affect real-world performance. Therefore, recognizing unique dialect accents and narrow data resources available for Arabic language AI model training remains a key challenge. To explore these issues, an online demo has been designed by the developers of this AI-powered voice assistant (Waheed et al., 2023).

Salhab & Harmanani (2024) present AraSpot in their study, a speech recognition system designed for identifying spoken Arabic commands. It focuses on a dataset of 40 Arabic keywords and enhances performance using online data augmentation techniques. This smart system enhances Arabic speech recognition pattern accuracy by leveraging unique ConformerGRU architecture. It tries to address the key Arabic speech processing challenges, such as narrow datasets and dialectal variations. As demonstrated by experimental findings, AraSpot performed better than existing AI voice assistant models used for Arabic voice command applications.

Dr. Al-Natsheh joined a prominent Arabic content website Mawdoo3 as an executive and co-founded Salma.ai, the first Arabic voice digital assistant, similar to Siri but capable of understanding various Arabic dialects (Bhatt et al., 2022). Salma helps users find information across various topics, including health, education, and lifestyle. accessible in Arabic.

RafiQ, developed by Bassem Waheed, is another AI Arabic voice assistant which responds better to queries made in Egyptian Arabic, than in Lebanese dialects. Though, it was supposed to run voice commands such as opening YouTube and Facebook, booking a Careem or Uber, and pulling up the menus of nearby restaurants, and help in waking user up in the morning, it couldn't deliver well to all queries (Fernandez (2020).

Rammas is a generative AI tool developed by DEWA. It offers a user-friendly interface for handling customer service request issues by providing 24-7 customer support (Khaleej Times 2023). A study by Egbemhenghe et al. (2023) discusses the role of ChatGPT in enhancing water quality control and decision-making systems, promoting precision irrigation, and optimizing water usage. DEWA integrated ChatGPT into its virtual employee Rammas to improve customer interactions, address queries timely, and provide accurate responses based on available information.

Yasmina is a Dubai-developed Arabic-speaking AI voice assistant designed to interact with users in the Khaleeji accent. It can perform Arabic speech recognition and convert text to speech. It can also understand more localized commands like playing the Azan and giving details about the Hijri calendar. Beyond performing these tasks, it stands out by cracking jokes in a humorous and local style. The assistant aims to offer a personalized experience by understanding and adapting to various Arabic dialects, making it more relatable and engaging for its users. Yasmina's development team faced challenges in perfecting the Arabic language for the AI assistant because of its complexity, rich vocabulary and diverse dialects. They are still, however, aiming to finetune the project in the context of overcoming dialectal barriers to offer seamless user experience (Abdulla, 2023).

Additionally, Khader, et al. (2024) present that to perform the transcription of debates or Munazarat, a new Arabic speech collection mainly in Modern Standard Arabic (MSA), Fenek, a speech-to-text Kanari AI tool has been used, where Kanari AI is an advanced speech recognition start-up by QCRI researchers (Elmagarmid & Saoudi, 2021). The Munazarat 1.0 dataset includes various debate topics with metadata for each debate and is a valuable resource for training Arabic language tools, creating argument analysis systems, and studying Arabic argumentation and rhetorical styles. However, Fenek AI still face challenges in fully grasping context due to the vast linguistic structure of Arabic, leading to potential misinterpretations.

2.7.2. Non-Arabic AI Voice Assistants

There are several non-Arabic AI voice assistants that are offering services to different Arab regions in the world. Integrated into Amazon's Echo devices, Alexa can perform a variety of tasks, such as playing music, setting reminders, and controlling smart home devices. Available on Android devices and Google Home speakers, Google Assistant offers robust integration with Google's ecosystem, providing seamless access to information and services. Further, embedded in Apple's devices, Siri offers a range of functionalities, from sending messages to managing daily tasks, and is known for its strong privacy features. Also, initially released for Windows Phone, Cortana is available across various platforms, including Windows 10, where it helps users manage schedules and find information. Lastly, integrated into Samsung devices, Bixby offers unique features like visual search and contextual reminders.

Table 2-1: Arabic Vs non-Arabic AI voice assistants

Category	`Arabic AI Assistants	Non-Arabic AI Assistants
Language Support	Focus on Arabic dialects (Egyptian, Gulf, & Modern Standard Arabic)	Primarily support English, but also offer support for multiple languages
Accent & Dialect Recognition	Specialized in understanding different Arabic dialects, though may struggle with nuances	Well-optimized for global dialects, focusing on major languages (English, Spanish, etc.)
Cultural Sensitivity	Tailored to Arab culture & regional preferences (Islamic holidays, local events, traditional customs)	Western-centered, with cultural references that may not be as relevant to Middle East or Arab regions
Voice Recognition Accuracy	Good, but accuracy varies by dialect & regional accent	High accuracy, especially for English & major European languages
Integration with Local Services	More likely to integrate with local Arabic services, such as regional news, stores, and banking apps	Integrates with global services, but have less localized options for Middle East and North Africa (MENA) region
Natural Language Processing (NLP)	Adapted to understand Arabic syntax, morphology, & script (more complex than Western languages)	Better refined in English, Spanish, or French, with ess emphasis on languages with non-Latin scripts
Personalization	Focuses on culturally relevant content (prayer times, local news etc.)	Personalization is more generalized, with less focus on specific regional practices
Market Adoption	Limited compared to global players (Siri, Alexa, & Google Assistant, though growing in the MENA region	Widely adopted globally, especially in Western markets

Voice Assistants Available	Kanari AI, Yasmina, Rammas, VoxArabica, Salma, Fanar, QVoice, RoboMatic, RafiQ, AraSpot	Siri, Google Assistant, Alexa, Cortana
Support for Regional Devices	Compatible with MENA-region specific devices & home automation systems	Focused more on global hardware brands (Amazon, Apple, & Google) with limited support in some regional tech markets

2.8 The Need for Culturally Adaptive AI Systems

As artificial intelligence (AI) revolutionizes industries everywhere, there has been a greater need for Arab world-specific, culturally adaptive artificial intelligence systems in the domain of knowledge systems. Internationally recognised AI systems are mostly trained on Western-oriented data sources and can be ineffective in completely addressing cultural, societal, and ethical aspects of Arab society. The lack of such cultural adaptation may lead to prejudices, misperceptions, and low trust levels from the end-user towards the use of AI. To avoid this, experts refer towards integrating Arabic-language processing, cultural aspects, and ethical frameworks of AI into systems for enhanced user interaction and more equitable technological growth. The process of developing the AI systems with regard to local customs, traditions, and values can lead to greater adoption and societal returns.

For Arabic KBS, cultural adaptation requires three layers:

1. Linguistic: dialect identification, code-switching handling, diacritic restoration for NLU, and prosody modeling for expressive, dignified TTS;
2. Cultural: intents and content aligned with local calendars and norms; respectful address forms; configurable register and dialect; content provenance and context disclaimers;
3. Institutional: procurement standards that mandate dialectal evaluation sets, public reporting of WER/intent success by dialect, and ongoing risk management and governance.

Governance frameworks and standards. UNESCO’s Recommendation (2021) sets high-level principles; NIST’s AI Risk Management Framework (AI RMF 1.0) provides a life-cycle approach to governability, transparency, and measurement; NIST’s Generative AI Profile (2024) adds controls for content provenance and misuse; and ISO/IEC 42001:2023 codifies organization-level AI management systems (AIMS) for continuous improvement and accountability. Ministries, universities, and hospitals deploying AIVAs can adopt these to structure risk registers, audits, and incident response.

Legal constraints. When AIVAs make or support consequential decisions, automated decision-making safeguards apply in many jurisdictions. European GDPR Article 22 entitles individuals not to be subject to decisions based solely on automated processing producing legal or similar significant effects, with narrow exceptions. Even outside the EU, these safeguards are widely referenced. For KBS-integrated AIVAs, this implies human-in-the-loop arrangements, contestation channels, and clear disclosure (Pasupuleti, 2025).

2.8.1. Designing Inclusive and Ethical AI Voice Assistants

Though AI voice assistants are becoming an essential part of our daily lives, their efficient functioning in the Arab world is still limited due to limitations posed by vast Arabic dialects and cultural biases. The voice assistants on hand, as mentioned above, lack in terms of dialect variations, and this translates into inaccurate responses and frustrations for the end-user. Besides, ethical and moral factors such as gender sensitivities, privacy, religion, and regional customs must also be taken into account when developing solutions for AI. To address these challenges properly, strategies such as implementation of bias reduction frameworks, collaborative working with regional experts, and use of a diverse range of datasets from different languages should be adopted by the AI tech developers because a holistic design methodology ensures that AI voice systems are operating correctly alongside social and cultural alignment.

It takes more than just including "Arabic" as a language toggle to create inclusive and moral AI voice assistants (AIVAs) for Arabic-speaking communities. Dialectal breadth, socio-pragmatic norms, religious and privacy concerns, and institutional trust practices that influence daily use are all issues that systems need to solve. Research indicates that non-dominant accents and dialects can cause speech technologies to perform poorly, resulting in inconsistent error rates and a worse user experience (Koenecke et al., 2020). According to Nasr et al. (2023), this risk is increased in Arabic by diglossia (MSA vs. colloquial variations), fine-grained regional dialects (such as Gulf, Levantine, Egyptian, and Maghrebi), and frequent code-switching with English or French. Dialect-aware data pipelines and assessment methods that reflect performance by dialect, domain, and genre instead of a single aggregate score are thus the first steps towards inclusive design. Dialect-stratified ASR WER/CER and NLU metrics should be required by procurement and technical governance prior to deployment in knowledge-based services; shared benchmarks like MADAR and NADI show useful methods to measure downstream tasks and fine-grained dialect identification.

Ethical inclusion also hinges on socio-pragmatic alignment how the assistant addresses users, manages politeness, and selects register. In Arabic contexts, AIVAs should support task-sensitive register control (such as MSA for official transactions; locally appropriate vernacular for personal reminders), culturally appropriate salutations, and respectful forms of address, with opt-in personalization to reflect user preference. Design patterns from human–AI interaction research recommend making system status legible, confirming understanding, and offering graceful repair when recognition fails (Amershi et al., 2019). For public-facing answers, provenance disclosure naming the responsible ministry or agency and a last-updated timestamp translates abstract transparency into an actionable trust signal, aligning with UNESCO’s Recommendation on the Ethics of AI (2021) and the NIST AI Risk Management Framework 1.0 (2023).

From a data-governance perspective, inclusive AIVAs implement privacy-by-design and purpose limitation throughout the lifecycle. For the UAE, this means aligning collection, processing, and retention with Federal Decree-Law No. 45 of 2021 (PDPL), minimizing personal data, and offering meaningful consent and control (like easy deletion of voice histories; local on-device processing where feasible). Documentation artifacts Model Cards for models and Datasheets for datasets should be adapted to speech/NLP to describe populations seen in training (dialects, age bands, recording conditions), known failure modes, and evaluation gaps (Mitchell et al., 2019). These artifacts enable auditors and buyers to assess whether models are likely to disadvantage specific communities as Yemeni or Sudanese Arabic speakers and to plan mitigation.

Amershi et al. (2019) concentrate on practical human–AI interaction principles, including clarity, confirmation, and recovery from errors, whereas Mitchell et al. (2019) emphasise documentation and accountability through Model Cards and Datasheets. UNESCO (2021) and NIST (2023), by contrast, provide broader ethical and risk-governance principles. These approaches are complementary: interaction guidelines address the user interface, documentation supports technical transparency, and governance frameworks guide organisational oversight. However, the literature provides limited empirical evidence on how Arabic-speaking users prioritise these different ethical and design requirements.

Bias reduction in Arabic AIVAs is both a data and a process problem. On the data side, developers should curate dialect-balanced corpora, expand loanword and transliteration coverage for domain terms (health, finance), and include code-switching examples reflective of real usage. On the process side, teams should institute pre-deployment red-teaming with Arabic-speaking users from multiple regions and demographics (including older adults and women in healthcare/education) and run scenario-based tests for high-stakes tasks like eligibility, deadlines, medication names. ISO/IEC guidance on AI management and risk can be used to formalize these controls, while NIST AI RMF recommends continuous monitoring with metrics tied to harm hypotheses for example, measuring whether one dialect group receives more “I didn’t catch that” failures in service triage.

Finally, designing with inclusivity and ethics involves multiple stakeholders. Product teams should not just localize after the fact; they should co-design with service owners (healthcare providers, ministries) and regional Arabic linguistic specialists. The following expectations can be incorporated into public procurement by governance bodies: (1) dialect-stratified performance thresholds and stress tests for code-switching; (2) required disclosures in responses regarding provenance, timestamp, and jurisdiction; (3) appeal and human-override paths where outputs may affect rights; and (4) accessibility options (slower TTS, diacritical text, bilingual summaries). These components enable AIVAs to transition from generic tools to reliable knowledge mediators that honor community norms and the variety of the Arabic language.

Overall, the literature agrees that inclusive and ethical Arabic AIVAs require dialect-balanced data, culturally appropriate communication, privacy-by-design, transparency, accessibility, human oversight, and continuous risk monitoring. However, much of the existing evidence remains technical, normative, or policy-oriented rather than grounded in the reported experiences of Arabic-speaking users in the UAE. Studies also tend to examine fairness, privacy, usability, and governance separately, providing limited understanding of how these factors jointly influence trust, perceived importance, and continued use. A further gap concerns how ethical and inclusive requirements should be translated into the storage, retrieval, updating, and governance functions of organisational KBS. These gaps inform the research questions concerning perceived challenges, effective communication, and user-centred design, and support the adapted framework and proposed conceptual AIVA–KBS integration blueprint.

2.9 Benefits and Challenges of AI Voice Assistants in Arabic-speaking communities

In the light of the above sections, a few critical benefits and challenges of AI-driven voice assistants have been extracted. Considering that AI-based voice assistants automate jobs, Abdelraouf (2024) states that overall system efficiency increases due to the perceived benefits of saving time and reducing error experiences. However, narrow data resources available for Arabic language and discrepancies in Arabic speech recognition are challenges other than algorithmic partiality. Bouzid (2023) also presents the prevailing challenges of contextual accuracy for different Arabic dialects as well as algorithm biases in multilingual speech processing in this context. The system Arabic Personalized Control of Things (APCoT) enables home automation by giving Arabic voice commands. It is especially beneficial for elderly and disabled individuals who find doing tasks such as switching lights or controlling other devices at home difficult. Experimental results showed that it worked with 99.63% efficiency for a single speaker and 94.4% for multiple speakers with quick response time.

Arabic-speaking consumers benefit from a number of domain-specific advantages offered by AI-driven voice assistants (AIVAs). In addition to this, they improve accessibility: hands-free, eyes-free interaction helps older folks and others with visual or motor disabilities, especially in hospital, home, and automobile settings where voice decreases cognitive burden and friction. Second, when responses are based on reliable sources, they can reduce recontact rates and shorten resolution times. Examples of this include triaging government inquiries, surfacing policy updates, or directing Arabic-speaking users to the appropriate procedure. Thus, these improvements enhance service reach and continuity in knowledge-based systems (KBS). Third, AIVAs can strengthen information equity by offering spoken explanations in Modern Standard Arabic (MSA) and dialectal paraphrases, thereby accommodating users with different literacy levels. Finally, where provenance and update timestamps are exposed, assistants can bolster institutional trust, aligning with transparency guidance in global AI governance frameworks (UNESCO, 2021).

Set against these gains are persistent challenges. The most cited is linguistic coverage: Arabic diglossia and fine-grained dialect variation (Gulf, Levantine, Egyptian, Maghrebi) degrade automatic speech recognition (ASR) and natural language understanding (NLU) relative to MSA, especially under genre shift (conversational YouTube vs broadcast). Community evaluations such as MGB and MADAR consistently document these drops, underscoring the need for dialect-aware modeling and reporting (Ali et al., 2017; Bouamor et al., 2019). Code-switching with English or French—ubiquitous in professional and youth registers—further increases error rates unless systems implement token-level language identification and loanword-aware lexicons. A second category is fairness and bias: uneven datasets can yield higher error rates for certain accents, age groups, or speaking styles, risking exclusion of precisely those users who might benefit most (Koenecke et al., 2020). Third, privacy and security concerns arise because voice data can reveal identity, health, or household routines; robust consent, minimization, on-device processing where feasible, and retention limits are essential, particularly in public-sector deployments (UNESCO, 2021; NIST, 2023). Fourth, operational alignment matters: without tightly coupled knowledge bases and workflow integrations, accurate recognition may still produce unhelpful outcomes like outdated rules, ambiguous eligibility.

Furthermore, AI tools such as voice assistants and intelligent tutoring systems (ITS) come with the benefit of offering user-specific learning for those who want to improve their Arabic learning, as stated by Anwar & Ahyarudin (2023). Nevertheless, it put emphasis on the development of AI applications dealing with Arabic linguistic variations and contextual understanding based on cultural differences. Also, Al Shamsi (2021) explores the increasing popularity of using AI voice assistants among the students acquiring higher education due to various benefits, including the credibility of AI tools, perceived enjoyment, and ease of use. Additionally, the AIVA "Ali" deployed by the Bahrain government offers benefit, as derived from the study by Elmedany (2024), of assurance of user data privacy in compliance with Bahrain's regulations. In addition, the advantage of reduced response time and seamless communication between different entities, such as government and public, has been demonstrated by Al Ali & Khalil (2025). However, the related challenges like transparency, data, security, ethical use of AI, training professionals, developing AI policies for response assurance, and effective implementation have also been mentioned.

2.9.1 Linguistic and Dialectical Challenges: MSA vs. Dialects

The Arabic language has its specific linguistic challenges because of its diglossia nature, where there exists a coexistence of Modern Standard Arabic (MSA) with several other regional dialects. On the contrary, everyday communication mostly uses local dialects, which are quite different from one region to another. The linguistic and dialectic difficulties between regional dialects and Modern Standard Arabic (MSA) significantly influence user experience and cultural expectations because, on the one hand, there exists a formal use of MSA in official discourse, and on the other hand, there are regional dialects dominating every-day usage. The gap between the two creates difficulties for AI-driven language applications and natural language processing (NLP), such as research on DiaFrame shows a need for incorporating variations of the dialect for enhanced user experience in human-AI interaction (Wasi et al., 2024). Also, there exists a need for AI-assisted language training tools to take into consideration the regional nuances of the dialect for cultural responsiveness, as presented by research on the use of humour and empathy in AI dialog systems for teaching languages (Zhai et al., 2024).

Cultural and societal expectations also result in complicated user interaction because not only does language enable daily conversation, but also individuals' identity and social norms are expressed through language. Studies show that successful intercultural communication needs to take into account linguistic identity and national characteristics to achieve meaningful interaction (Kim, 2020). Moreover, interaction techniques need to bridge cultural and dialectic gaps in order to ensure sustained user interest, either in educational settings or digital communication systems. A study of boundary-transcending in the linguistic identity context highlights the importance of cultural reflexivity in online and offline settings (Kassis-Henderson & Cohen, 2018). Being able to comprehend such dynamics becomes paramount for developing AI solutions such that varying linguistic and cultural needs can be addressed in the proper way.

Kim (2020) argues that meaningful intercultural communication depends on linguistic identity and national characteristics, whereas Kassis-Henderson and Cohen (2018) place greater emphasis on cultural reflexivity across online and offline interactions. These perspectives complement the more technically oriented studies of Arabic NLP by explaining why dialect choice is linked to identity, belonging, and social expectations. However, neither study directly evaluates how these factors influence Arabic-speaking users' trust, satisfaction, or continued use of AIVAs in organisational settings.

Overall, the research studies mentioned in the above sections show that benefits offered by AI-based Arabic assistive technology include reduced response time to customer queries, minimal chances of error occurring, providing user-specific learning, and compliance with regulatory frameworks, thereby streamlining accessibility and efficiency, reliability, enhanced user experiences in education, and increased trust in their deployment and usage. Also, Arabic dialect identification for a few dialects, though not actually vast but with limited datasets for model training, has also been achieved. Whereas some of the prominent challenges linked with AI voice assistants are partial Arabic dialectal support due to insufficient datasets for AI model training, contextual understanding based on cultural differences, discrepancies in Arabic speech recognition, and algorithmic biases.

2.10 AI Voice Assistants within Knowledge-Based Systems

A Knowledge-Based System is an information system designed to capture, organise, store, retrieve, and apply specialist or organisational knowledge to support problem-solving, service delivery, and decision-making. Unlike a conventional database that primarily stores records, a KBS combines structured knowledge with mechanisms that allow relevant information to be identified, interpreted, and presented in response to a user's enquiry. Within organisational contexts, this knowledge may include policies, procedures, regulations, frequently asked questions, service rules, technical guidance, and institutional records. A Knowledge-Based System comprises several interconnected components. The knowledge base stores facts, rules, policies, documents, ontologies, and validated organisational information. The knowledge-acquisition mechanism gathers knowledge from experts, databases, documents, and service owners. The inference or retrieval mechanism identifies information relevant to a specific enquiry. The user interface enables communication between users and the system. The knowledge-maintenance process validates, updates, timestamps, and removes outdated content. Finally, the governance and security layer manages access controls, privacy protection, accountability, auditability, and compliance throughout the system's operational lifecycle and organisational use.

Moreover, Digital voice assistive applications have evolved into sophisticated knowledge-based systems using natural language processing (NLP) and are being employed for enhanced user experience with regard for the privacy of the general public. Such virtual systems apply cognitive computing techniques for smart decision-making and problem-solving for varying purposes. Recent advancements depict their employment in organizational knowledge management of companies, where AI-powered assistants promote enhanced knowledge accessibility and streamline complex tasks (Mohan & Paramesh, 2024). Besides, cognitive AI-driven voice assistants find extensive utilization in technical environments, such as industrial environments, where voice assistants with AI support operators with real-time sharing of knowledge on complex systems (Freire et al., 2022).

Within knowledge-based systems (KBS), contemporary AI voice assistants operate as multilayer pipelines that couple automatic speech recognition (ASR) and natural language understanding (NLU) with domain ontologies, knowledge graphs, and policy-governed retrieval modules. In practice, user speech is transcribed, parsed into intents and entities, and then routed to a knowledge layer that indexes authoritative sources (for instance, regulations, procedures, technical manuals) using semantic models and entity linking. Retrieval-augmented generation (RAG) or template-based natural language generation (NLG) formats answers while preserving provenance citations, timestamps, and jurisdictional scope which is essential for trust and audit in public services and regulated industries (Lewis et al., 2020; Bizer, Heath, & Berners-Lee, 2009). This architecture allows assistants to act as “knowledge mediators” rather than mere command interfaces: they can disambiguate similar terms, apply eligibility rules, and surface next-step options drawn from workflow engines or case-management systems. As these systems scale, quality and governance requirements accuracy, reliability, security, and maintainability per ISO/IEC 25010; risk measurement and continuous monitoring per NIST AI RMF 1.0 become as critical as raw ASR/NLU scores, since failure modes often stem from outdated content, missing ontological links, or unclear applicability rather than from transcription alone (NIST, 2023).

A second, equally important layer concerns organizational embedding and ethics. In corporate KBS, voice interfaces increase “knowledge accessibility” by lowering interaction costs for time-pressed staff and by enabling eyes-busy, hands-busy scenarios (such as, maintenance, logistics). Empirical KM literature shows that codified expertise, when bound to task contexts and delivered just-in-time, improves decision quality and cycle time (Davenport & Prusak, 1998; Nonaka & Takeuchi, 1995). In industrial settings, voice assistants integrated with sensor data and digital twins can answer fault-diagnosis queries, read out procedures, and capture operator notes to feed continuous improvement loops (Freire et al., 2022). In public-sector KBS, assistants can triage Arabic-speaking users requests, retrieve policy clauses from controlled repositories, and log machine-readable rationales for hand-offs to human agent’s practices aligned with transparency and accountability norms. Privacy-by-design measures (on-device wake-word detection, minimization of stored audio, configurable retention) and content governance (source whitelists, editorial workflows, and periodic review) are necessary to meet data-protection obligations while preserving utility. Effective evaluation therefore combines technical and service metrics: dialect-stratified error rates, intent accuracy, task-completion and time-to-resolution, escalation rate, and user-perceived clarity and trust. When these elements are jointly engineered semantic KBS foundations, provenance-rich responses, and lifecycle governance voice assistants move beyond convenience features and become dependable gateways to institutional knowledge (Mohan & Paramesh, 2024).

2.10.1 Information Retrieval and Knowledge Dissemination

Voice assistants embedded in knowledge-based systems improve information retrieval and dissemination by intelligently responding to user queries and contextualizing knowledge. A recent study by Rajesh et al. (2023) demonstrated the effectiveness of AI virtual assistants for academic advice, where the virtual assistants leverage the support from the expertise for the decision-making process to guide students with the right recommendations. Besides this, the enhancement in ontologies contributed towards the virtual assistants for providing context-aware and more precise advice, thus enabling the dissemination of expert knowledge over multiple industries (Dorodnykh & Nikolaychuk, 2022).

2.10.2 AI Integration in Education, Governance, and Service Sectors

The sources of organisational knowledge can be policy documents, service manuals, legislation, skilled employees, frequently asked questions, historical cases, and operational databases. This knowledge must be validated, classified and structured before entering the KBS in order to allow the system to differentiate various topics, users, jurisdictions and authority levels. Knowledge can be represented in the form of documents, rules, semantic relationships, ontologies, decision trees or linked records. There should also be constant maintenance in the operation of KBS since policies, procedures, deadlines and regulations can vary with time.

If a user gives an enquiry, the retrieval mechanism recognises the intention of the user, scans the source of knowledge and assesses the applicability of the accessible information and gives the best reply. Knowledge that is acquired in high-risk environments should contain information about the source, jurisdiction, a date of last update, and conditions of confidence or applicability. In the case of an ambiguous request or one that the system is unable to find an authoritative answer to, a transfer to a human service representative should occur. After organisational knowledge is captured, structured and retrievable, AI voice assistants can offer a conversational interface where users can access this knowledge in education, governance, healthcare and service contexts.

AI voice assistants have proved tremendous potential for revolutionizing education, the sector of governance, and the service sector by boosting efficiency as well as engagement levels. In the education sector, AI voice assistants bridge the knowledge gap for the students by offering customized support for study and adaptive tutoring (Rajesh et al., 2023). In the government sphere, AIVAs have improved public service delivery by automating the Arabic-speaking users' inquiries, thus providing the services promptly while minimizing the administrative load. Moreover, in the service sector, real-time voice technology has been implemented to support human assistants in handling conversations more effectively, improving customer experience and operational productivity (Marconi, 2023).

In education sector AIVAs can provide course guidance, advisement, and LMS navigation in Arabic with register control (MSA for policy, dialect for conversational coaching). Task success depends on domain ontologies (program rules) and careful disclaimers for high-stakes decisions (graduation eligibility) with human review consistent with GDPR-style safeguards for automated decision-making (Thouvenin et al., 2022). Moreover, Citizen-service assistants (permits, fines, appointments) must maintain bilingual parity (Arabic/English), dialect tolerance, and provenance in government sector. Rammas demonstrates throughput at scale when backed by enterprise workflows and curated content (Roig, 2017). In addition, Mahlknecht et al. (2023) stated that for symptom checking and appointment routing, assistants should restrict scope, avoid diagnostic claims, and provide escalation to licensed clinicians in healthcare sector. Adoption of NIST AI RMF risk controls (model card, monitoring, incident logging) and ISO/IEC 42001 management processes is recommended for hospitals and insurers deploying Arabic voice assistants.

2.11 Theoretical Foundation and Adapted Conceptual Framework

Technology adoption theories offer a valuable baseline on why users accept, appreciate, or maintain use of AI voice assistants. Technology Acceptance Model (TAM) is a model that elucidates the adoption in terms of perceived usefulness and perceived ease of use. These ideas apply to the current work since perceived advantages and issues connected to usability can affect the importance of AIVAs to users. Nevertheless, TAM fails to cover Arabic dialect variation and code-switching, cultural expectations, privacy, and government communication sufficiently.

UTAUT2 builds on technology-adoption studies with the addition of performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, habit, and behavioural intention. The model can be applied to explain the social and institutional contexts of AIVA adoption in general. However, UTAUT2 is a loose consumer-technology framework and lacks specifics of culturally based communication, dialect sensitive communication, and access to organisational knowledge via KBS.

Service Robot Acceptance Model is also applicable as it takes into consideration the trust, perceived intelligence, quality of interactions and the reactions of users to intelligent service agents. These constructs can be used to explain why conversational systems may or may not be accepted by users. Nevertheless, the research on service robot acceptance is typically oriented at embodied or service-oriented agents and lacks the depth of linguistic complexity of Arabic voice interaction and the role of government communication and organisational knowledge retrieval.

Under this, therefore, this thesis does not test TAM, UTAUT2 or Service Robot Acceptance Model as it is. Rather, it suggests a modified conceptual model based on the chosen principles of these models and the literature on the Arabic language, culture, privacy, trust, and KBS integration. The framework places perceived benefits, perceived challenges, design preferences, and government communication as predictors of the perceived importance of AI voice assistants. The choice of these constructs is based on the fact that they embody the general principles of technology-adoption, and the cultural, linguistic and institutional peculiarities of the Arabic-speaking users in the UAE.

The modified framework thus builds on the traditional technology-adoption thinking by introducing cultural-linguistic fit and institutional communication to perceived benefits and challenges. It cannot be taken to be a formal extension of an existing theory. Instead, it is a context-like, empirically studied framework created to explain the perceptions of the Arabic speaking users towards AIVAs and to inform a proposed blueprint of conceptual AIVA-KBS integration.

2.12 Synthesis of Existing Studies

The literature on AI voice assistants shows how these have evolved from simple natural language processing tools to complex systems. Early developments focused on foundational innovations in speech recognition and command execution (for example, Siri, Alexa, Google Assistant). The focus has been on developing multilingual support, context-aware responses, and smart device integration, which primarily benefit English-speaking users. For Arabic language, recent studies have highlighted the necessity for NLP models to effectively handle both MSA and dialectal variations to improve language comprehension and user engagement. Regionally, new Arabic-centric foundation models as like Falcon Arabic, 2025 aim to capture linguistic diversity at pretraining time, a prerequisite for robust voice and text assistants. Even so, public evaluations must report Arabic-specific metrics and dialectal breakdowns to be useful for KBS procurement and governance.

Though a lot of work has been carried out by the Arab tech companies or start-ups such as QCRI, InfraDrive, DEWA, and Mawdoo3 to provide the AI based technological solutions to the Arabic-speaking population keeping in view the special linguistic complexities and cultural nuances, most of these solve needs in parts of suggested Arabic dialects and cultural expectation. On the other hand, non-Arab tech giants like Google, Apple, Amazon, and Microsoft have developed AI voice assistants with Arabic support, but each faces limitation. Google Assistant supports Arabic but struggles with dialectal variations and natural-sounding responses.

Similarly, Apple’s Siri primarily uses Modern Standard Arabic (MSA) and has limited support for regional dialects, which thus impacts its accuracy in conversational settings. While Amazon’s Alexa focuses on Gulf Arabic, making it less impactful for users from other Arabic-speaking regions. Microsoft’s Cortana had some Arabic support before being discontinued, and while Microsoft continues AI research, its voice assistant capabilities in Arabic remain limited. These challenges highlight the complexities of Arabic’s diverse dialects and linguistic structures. This appears to be the biggest limitation that needs to be focused on standardized adaptation of these AI voice assistants, which is the aim of this study, covering all dialectal variations across entire Arab region with different cultures.

2.12.1 Research Gap

Despite steady progress, two overarching limitations continue to constrain the effectiveness and societal uptake of AI voice assistants (AIVAs) in Arabic: (1) incomplete adaptation to the full spectrum of Arabic dialects and code-switching practices, and (2) limited cultural contextualization in real-world, high-stakes knowledge-based systems (KBS). First, empirical evaluations from community shared tasks show that dialectal variability systematically degrades system performance relative to Modern Standard Arabic (MSA). The MADAR shared task (ACL 2019) demonstrated the difficulty of fine-grained dialect identification at city and country levels across 26 varieties, underscoring how model accuracy drops as dialect granularity increases and domains shift (travel text, Twitter). Similarly, the Arabic MGB-3 challenge, which pivoted from broadcast news to multi-genre Egyptian YouTube speech, highlighted domain and genre mismatch as core failure modes and introduced a dialect identification track precisely because recognition performance varies markedly by variety and condition. These results translate directly to deployment risk: a single “Arabic” model is unlikely to satisfy users across Gulf, Levantine, Egyptian, and Maghrebi contexts without explicit dialect routing, domain adaptation, and continual learning.

A related and under-addressed gap is robust handling of Arabic–English/French code-switching, which is ubiquitous in Gulf and North African settings. Recent Arabic code-switching ASR studies confirm that standard monolingual pipelines struggle to maintain word error rates when languages mix within an utterance; specialized decoding graphs, multilingual pretraining, and hybrid WFST/Transformer strategies are needed to recover accuracy, yet these methods remain sparse in commercial KBS deployments.

Beyond recognition, downstream natural language understanding (NLU) and text-to-speech (TTS) face their own gaps: ambiguous orthography (optional diacritics), dialectal morphology, and prosody modeling for natural, respectful synthesis across registers (formal vs colloquial). A recent survey of Arabic ASR consolidates these barriers and points to persistent data scarcity outside a few dominant dialects, limiting generalization and fairness across communities.

The second major limitation concerns cultural contextualization and domain alignment for citizen-facing services. Academic and industry narratives around Arabic launches emphasize early design choices about which Arabic to “speak” and how to localize intents, content, and humor—choices with direct implications for trust, politeness strategies, and acceptability in public services. Amazon’s account of “how Alexa learned Arabic” documents the contested terrain between MSA and Gulf vernaculars and the engineering compromises required, illustrating that language support alone does not equal cultural fit. Production evidence from the UAE’s Dubai Electricity and Water Authority (DEWA) indicates that large-scale Arabic–English voice services (Rammas) can achieve high throughput and satisfaction only when paired with carefully curated back-end knowledge, bilingual flows, and governance—yet public reporting still lacks dialect-stratified metrics (for instance, WER by dialect, intent success by domain), limiting external auditability and reproducibility of results. This reveals a third, cross-cutting gap: the absence of standardized evaluation protocols for Arabic KBS assistants that combine speech (ASR/TTS), language understanding, cultural appropriateness, and task completion, benchmarked with community corpora such as MADAR and MGB.

Hence, current literature and deployments converge on the same conclusion: AIVAs for Arabic remain constrained by (i) insufficient dialectal coverage and code-switching robustness at scale, driven partly by uneven datasets and limited domain adaptation; and (ii) inadequate cultural and institutional integration—privacy norms, politeness and address forms, religious calendars, and provenance-controlled content—in knowledge-centric workflows. These gaps explain frequent misrecognitions, contextually inappropriate responses, and uneven trust, thereby impeding broad, equitable adoption across the Arab world. Addressing them requires Arabic-centric data creation beyond MSA, dialect-aware modelling and routing, transparent, dialect-stratified reporting in public tenders, and co-design with regional institutions to embed cultural governance directly into AIVA architectures.

Table 2-2: Gaps Identified

Study	Context and sample	Research focus	Method	Main findings	Limitation	Gap addressed by this thesis
Al Shamsi (2021)	UAE university students	AIVA adoption	Quantitative survey	Usefulness, enjoyment and facilitating conditions influenced adoption	Limited student sample; limited focus on dialect and KBS integration	Examines broader Arabic-speaking users and culturally specific requirements
Salih (2025)	Oman	Continued use of AI-enabled services	Quantitative study	Trust and service quality influenced continued use	Different national context; limited attention to Arabic linguistic variation	Investigates perceived benefits, challenges and continued Arabic usage in the UAE
Abdelraouf (2024)	Oman media professionals	AI integration in audiovisual media	Survey/interviews	AI was perceived to improve efficiency, but Arabic data and recognitio	Sector-specific; not focused on everyday AIVA use	Examines AIVA perceptions across broader user and organisati

				n limitations remained		onal contexts
Bouzid (2023)	Arab region	Arabic speech technologies and AIVA adoption	Literature/technical review	Dialect, contextual accuracy and infrastructure remain major concerns	Primarily technical and regional; limited user evidence	Adds empirical evidence from Arabic-speaking users in the UAE
Salah (2023)	Arabic home automation	Arabic voice-controlled IoT system	Prototype testing	High recognition performance within defined experimental conditions	Limited speakers and controlled tasks; weak cultural evaluation	Connects technical requirements with user perceptions and KBS possibilities
Elmedany (2024)	Bahrain government services	Government intelligent voice agent	Case-based study	AIVA supported Arabic and English public-service interaction	Single institutional application	Examines perceived importance of government communication within a

						structural model
Al Ali and Khalil (2025)	UAE/public communication	AI-supported government communication	Applied/policy study	Faster responses and improved communication were reported	Limited empirical testing of user perceptions	Tests Communication in Government as a distinct construct
Jaradat et al. (2022)	Arabic-speaking drivers	Voice-controlled vehicle assistant	Technical development	Arabic commands may support safety and convenience	Limited dialect, cultural and privacy analysis	Identifies broader design preferences among Arabic-speaking users
Amershi et al. (2019)	General human–AI interaction	Human-centred AI design	Guideline synthesis	Clarity, confirmation and error recovery improve interaction design	Not Arabic-specific	Applies human–AI principles to Arabic dialect and cultural requirements
Mitchell et al. (2019)	AI governance	Model documentation	Conceptual framework	Model Cards support transparency and	Does not identify Arabic-user priorities	Links documentation and transparency to Arabic

				accountability		AIVA design preferences
Mohan and Paramesh (2024)	Organisational knowledge management	AI assistants and knowledge accessibility	Conceptual/applied study	AI assistants may support access to organisational knowledge	Limited Arabic-language and user evidence	Connects Arabic user requirements with a proposed KBS blueprint
Lewis et al. (2020)	Knowledge-intensive AI tasks	Retrieval-augmented generation	Technical research	Retrieval can ground generated responses in external knowledge	Not focused on voice, Arabic, or organisational users	Informs the proposed future AIVA–KBS integration architecture

2.12.2 Research scope

This study narrows its empirical lens to the United Arab Emirates (UAE) as a theoretically informative and policy-relevant testbed for Arabic AI voice assistants (AIVAs) integrated with knowledge-based systems (KBS). The UAE is a useful case for three reasons. First, it has articulated a long-horizon governance agenda for AI most notably the UAE National Strategy for Artificial Intelligence 2031 that explicitly links AI adoption to public-service performance, data infrastructure, and sectoral transformation, creating a conducive institutional environment for large-scale AIVA deployment in Arabic-speaking users services (UAE Government, 2023). Second, multiple UAE government entities have already operationalized Arabic–English conversational systems at scale (such as utilities, health, and one-stop government platforms), providing observable outcomes and artifacts for analysis (architectures, satisfaction metrics, bilingual flows) (Siemund et al., 2021). Third, there is an emerging content ecosystem around Arabic language, evidenced by recent partnerships to surface curated Arabic resources in mainstream assistants (like the Abu Dhabi Arabic Language Centre with Amazon Alexa), which is central to KBS quality and cultural alignment (Bureau, 2025).

Accordingly, the research scope encompasses three policy domains where KBS demands are high and misrecognition or cultural miscues carry tangible costs: government services, healthcare, and education. In government, the study also examines platforms such as TAMM (Abu Dhabi's one-stop shop), which has evolved into an AI-enabled service fabric offering hundreds of transactions through unified channels. Researchers focus on how voice interaction ties into service orchestration, knowledge retrieval, and identity-aware personalization, and whether dialect routing or code-switching support is visible in user-facing flows. In healthcare, an investigator analyzed contact-center augmentation and analytics (for example, Dubai Health Authority's adoption of AI-powered speech and text analytics) as precursors to fully conversational, privacy-preserving AIVAs that can triage requests, navigate entitlements, and surface guideline-concordant information. In education and culture, researcher probe how curated Arabic content partnerships (ALC–Alexa) and national strategy instruments shape language coverage, cultural registers, and provenance determinants of trust in knowledge delivery.

Methodologically, the study adopts a multi-layer evaluation aligned with KBS requirements. At the speech layer, researcher specify dialect-stratified quality indicators such as word error rate by Gulf/Levant/Egyptian/Maghrebi varieties and robustness to Arabic–English code-switching in realistic service utterances. At the language/knowledge layer, investigator measure intent recognition accuracy, ontology-anchored answerability, and verifiable citation of authoritative sources for policy or clinical content. At the governance layer, assess privacy-by-design safeguards (on-device wake-word processing where feasible, retention limits), transparency of performance reporting, and user-perceived cultural fit (forms of address, register, calendrical literacy such as Hijri support). These metrics are motivated by the UAE's strategy emphasis on governance and by lessons from large operational deployments like DEWA's Rammas, which reports multi-million enquiry volumes and bilingual service, yet does not publicly disaggregate performance by dialect an evaluative gap this study seeks to address.

The contribution is twofold. Empirically, it provides evidence-backed design guidance for Arabic AIVAs in KBS: dialect identification and routing, domain adaptation to government/health ontologies, curated local content ties, and culturally appropriate dialogue policies. Normatively, it proposes a procurement-ready evaluation protocol tailored to UAE use cases combining speech, language, cultural, and task-completion metrics and situates it within the country's AI policy stack and digital-government roadmaps (Alshehhi et al., 2024). While the UAE is not the entire Arab world, its mix of Arabic varieties, high digital penetration, and active state capability-building make it a strong lead market whose standards, if dialect-aware and transparent, can generalize to other MENA administrations seeking trustworthy, inclusive AIVAs embedded in knowledge-centric public services.

2.13 Chapter Summary

This chapter surveyed how culture and language shape the design and use of AI voice assistants, with a focus on Arabic-speaking societies. It first explained why culture and linguistics matter: language carries identity, values, and social norms, so voice systems must align with local forms of address, politeness, and register. Researcher reviewed how technology can both connect cultures and flatten differences, and why provenance, transparency, and human oversight are essential where assistants deliver public information. The chapter also outlined the distinctive features of Arabic diglossia between Modern Standard Arabic and regional dialects, frequent code-switching with English or French, and orthographic ambiguity showing how each affects speech recognition, language understanding, and dialogue management. Moreover, the chapter traced the evolution of AI voice assistants from command-based tools to context-aware, multi-device agents. It highlighted recent gains in multilingual support, personalization, and integration with smart homes and service workflows, while noting persistent gaps for Arabic users. Evidence from shared tasks and deployments shows that recognition accuracy and intent understanding drop when dialects shift or languages mix, and that cultural fit tone, register, religious and calendrical awareness directly influences trust and usefulness. Researcher also summarized current adoption patterns in the region, including public-sector use cases in utilities and city services, and research efforts to build Arabic-first models, datasets, and evaluation benchmarks.

The chapter synthesized what is known and identified the main gaps: limited adaptation to the full spectrum of Arabic dialects and code-switching, and insufficient cultural and institutional integration in knowledge-based settings such as government, health, and education. It argued for dialect-aware modeling, clear provenance and update practices, privacy-by-design, and participatory design with Arabic-speaking communities. These insights motivate the study's empirical focus on the UAE and inform the methodological choices presented in the next chapter, where researcher evaluate Arabic voice assistants against speech, language, cultural, and task-completion criteria aligned with real service needs.

3. CHAPTER 3: METHODOLOGY

This section provides the methodological framework of the study to address the adoption and perceptions of digital voice assistants in an Arabic-speaking country like the UAE. The methodology that has been adopted is a mixed method approach, wherein there is data collection and analysis quantitatively with qualitative inputs. The qualitative part of these elements provides more depth concerning user experiences and perceptions as regards AI voice assistants especially issues related to linguistic and Cultural Differences, thus providing a wider angle on AI integration in Arab speaking regions. Such a mixed-method approach ensures that the analysis is robust enough to fit the structure and needs of the research. The study was conducted in *two distinct rounds* of surveys (See Appendix A and Appendix B) to ensure a comprehensive understanding of the research problem; *first round* of survey has been cautiously designed to attain the baseline data with regard to demographics of respondents and their usage behavior regarding AI voice assistants, including the issues faced in terms of utilization of these technologies especially from Arabic contexts. This has been important for the broad-scale pattern identification and difficulties, that also are linguistic, cultural sensitivity-related, and technical level variances. A survey was distributed among 210 respondents from both governmental and private institutions, which allowed data capturing of responses on basic variables such as gender, age, educational level, technological awareness, and perceived hindrances. Quantitative statistical analysis was performed using software application called SPSS wherein descriptive statistics was used to consider the above key relevant variables (Wang et al., 2021).

Building from the learning captured during the first round, *second round* of the survey looked to improve the analysis by exploring more defined aspects of adoption and use of AI voice assistants. Using the Likert scale, this survey sought to obtain 459 respondents' views on topics ranging from the effectiveness of AI in communication tasks to perceived advantages and disadvantages, preferred design aspects, and overall public worth. The study also investigates the impact of cultural and linguistic personalization in enriching the experience of the users interacting with AI systems. In the second-round detailed description of optimizing AI voice assistants for specific needs within an Arabic-speaking community became feasible. Thus. both rounds yield a more rigorous framework of assessing AI voice assistants' adoption highlighting technological, cultural, and linguistic factors influencing user acceptance and gives developers and end users valuable insights on culturally and linguistically inclusive AI solutions. This study was divided into two quantitative research methods and both rounds utilized a structured survey based on the details given in the following sections.

3.1 Discussion on Research Philosophy and Approach

This study adopted a pragmatic research paradigm. Pragmatism was appropriate because the research questions required the use of different forms of evidence to understand both measurable patterns and context-specific user perceptions concerning AI voice assistants. The study therefore combined quantitative survey analysis with qualitative insights from open-ended responses. Rather than giving priority to a single philosophical position, pragmatism focuses on selecting methods that are most suitable for addressing the research problem. This position supported the sequential use of exploratory analysis in Round 1 and confirmatory structural modelling in Round 2.

The approach employed in this research is designed with an intention to fulfil the gap found from the literature about the underrepresentation of the Arabic-speaking community in AI voice assistant research. This is achieved by using the *mixed-method approach* wherein qualitative insights are merged with the quantitative surveys in order to present the best possible understanding about the experiences of the users and the impacts at the level of the organization. The survey design focuses on linguistic and cultural challenges; questions on language preference, dialectal variations, and cultural perceptions were included to address the subtle barriers that Arabic-speaking users face. The methodology thus directly addresses the limitations highlighted in previous research, including the inability of existing AI systems to effectively adapt to Arabic dialects and cultural contexts.

The mixed-method approach confirms that while collecting quantitative data on trends of demographics and usage, there is qualitative feedback received regarding user preferences and challenges. Furthermore, this allows the strong analysis of not just the barriers but also action recommendation to improve AI voice assistants. By linking methodological choices explicitly to research gap, the study underlines its innovative approach towards meeting the unique needs of the Arabic-speaking population.

3.2 Rationale for the Two-Round Survey Design

This study adopted a two-round cross-sectional survey design to provide both breadth and analytical depth. From the researcher's perspective, a single survey would have been insufficient because the study first needed to establish how Arabic-speaking users in the UAE engage with AI voice assistants before testing relationships among the factors influencing their perceived importance. Round 1 therefore served an exploratory purpose. It mapped patterns of exposure, frequency of use, language preference, perceived benefits, technical and cultural challenges, and suggestions for improvement. These findings helped identify the issues most relevant to the target population and provided an empirical basis for refining the second questionnaire.

Round 2 served a confirmatory purpose. It translated the main issues identified in Round 1, together with constructs drawn from the literature, into a structured model covering perceived benefits, perceived challenges, design preferences, government communication, and AIVA importance. Structural Equation Modelling was then used to examine the measurement properties of these constructs and test the hypothesised relationships among them.

This sequential logic responds directly to the examiners' request to explain why two rounds were conducted, what each round achieved, and how the first informed the second. The design also reduced the risk of imposing a final model before understanding the local linguistic, cultural, and institutional context. The SEM analysis was conducted using only data collected in Round 2. Round 1 served an exploratory role and was not combined with Round 2 for model estimation, measurement validation, or hypothesis testing. Accordingly, Round 1 generated contextual insight, while Round 2 provided more focused model testing. The rounds are therefore complementary rather than repetitive, and the final SEM model is based on Round 2 data.

Table 3-1: Comparison of the Round 1 and Round 2

Element	Round 1	Round 2
Purpose	Exploratory	Confirmatory
Main focus	Usage, preferences, benefits, challenges and suggestions	Testing relationships among selected constructs
Analysis	Descriptive and preliminary inferential analysis	Measurement model and SEM
Contribution	Identified relevant issues and variables	Tested the final conceptual model
Relationship	Informed the development of Round 2	Built on Round 1 and the literature

3.3 Relationship between the Round 1 and Round 2 Samples

Round 1 and Round 2 were two separate cross-sectional surveys. The surveys were anonymous and did not ask for personal identifiers, making it impossible to link a participant in one round with a participant in the other round. Thus, it is not possible to conclude whether there was any overlap of participants within the recruitment channels. The two sets of data were analyzed as independent samples, not repeated measures of the same respondents. Therefore, don't take the design to be a longitudinal or panel study.

They were designed to be different from but related to participant profiles for two reasons: methodological and for statistical purposes. Round 1 engaged 210 adults from public and private sector to define general trends related to demographics, AIVA exposure, use, language preferences, perceived challenges and cultural differences. For the more specific test of the conceptual model, 459 participants were recruited for round 2. Imbalances identified in Round 1 were targeted in Round 2 by actively seeking to recruit women, older adults and participants from healthcare and education. The purpose of Round 1 was exploratory and the purpose of Round 2 was confirmatory. Round 1 aimed to generate general data on usage, language preferences, perceived benefits, technical and cultural difficulties, and user suggestions. After selection of constructs and hypothesised relationships, Round 2 involved testing the constructs and hypothesised relationships using the measurement and structural models.

Both rounds had a number of basic inclusion criteria: 18 years or older, resident of the UAE, informed consent, andArabic/English fluency to complete an online questionnaire. In Round 2, however, an extra criterion was added that respondents have worked with AI voice assistants before. This criterion was included since participants had to have something to compare the perceived benefits, challenges, design preferences, government communication, and AIVA importance against in the SEM model.

Table 3-2: Comparison of the Round 1 and Round 2 Samples and Analytical Purposes

Element	Round 1	Round 2
Sample size	210	459
Purpose	Exploratory baseline	Confirmatory model testing
Sample relationship	Independent cross-sectional sample	Independent cross-sectional sample
Recruitment focus	Broad public- and private-sector coverage	More balanced gender, age, and sector representation
Main participant requirement	Adults residing in the UAE with access to a voice-enabled device	Same core criteria, plus prior AIVA experience
Main output	Usage, preferences, challenges, and cultural patterns	Measurement model and SEM testing
Direct comparison	Limited	Findings interpreted as complementary

3.4 How Round 1 Informed Round 2

Round 1 provided Round 2 with feedback on issues for targeted testing. High initial exposure and high sustained Arabic use facilitated the constructs of Effective Usage and

Perceived Importance. The Perceived Challenges construct was informed by reported difficulties with dialect, accent, culture, privacy and technical issues. Perceived Benefits were related to issues of accessibility, convenience, quicker service, communication and daily task assistance. Design Preferences were informed by suggestions for improved Arabic support, dialect adaptation, culturally appropriate response, privacy, local-service integration and feedback mechanisms. Communication in Government was formed by the interest in trustworthy public information, institutional communication and the integration of institutions with the government platform. Round 2 thus translated exploratory findings and literature-based constructs into multi-item scales/hypotheses. The final SEM model was tested with the Round 2 data and did not replicate Round 1. The final measurement and structural models were estimated only with data from Round 2. The constructs selected and refined were based on the results of round 1 which did not form part of the SEM analysis.

3.5 Study setting and population

The United Arab Emirates (UAE) is among the world's most digitally connected societies, with internet penetration estimated at roughly 99 percent in January 2024 and mobile subscriptions exceeding the size of the resident population a connectivity profile enabled by extensive 4G/5G coverage and high average speeds. Public agencies have progressively layered conversational channels voice and chat onto service front doors to triage inquiries, provide status updates, and surface verified guidance. Examples include the federal portal's deployment of generative-AI–based assistants and entity-level pilots announced under Dubai's Centre for Artificial Intelligence to accelerate AI use cases in government operations. These efforts sit within a broader trajectory in which the UAE consistently ranks at or near the global leadership tier in the UN E-Government Survey's composite measures of online service delivery, human capital, and telecom infrastructure.

Voice interaction is not confined to public portals. Residents commonly engage with voice user interfaces embedded in smartphones, in-vehicle infotainment systems, IVR helplines, and an expanding smart-home ecosystem. Industry tracking indicates a growing smart-speaker market in the Gulf led by the UAE and Saudi Arabia alongside rising demand for voice-controlled home automation; while precise household penetration varies by source, trend data consistently point upward for the UAE market segment. This ubiquity of voice-enabled touchpoints, combined with the country's digital-first government stance, makes the UAE a robust setting in which to study adoption and perceived value of AI voice assistants in knowledge-based service contexts.

Population. The target population comprises adults (18 years and older) residing in the UAE who can complete a survey in Arabic or English and who have access to at least one voice-enabled device (for instance, smartphone, smart speaker, or vehicle system). This definition reflects the bilingual service environment (Arabic as the official language, English widely used in business and expatriate communities), and the high baseline of device and network access documented in national statistics. Given that the expatriate population substantially exceeds the Arabic-speaking users population and that multiple Arabic dialects co-exist with English and South Asian languages in daily life, the sampling frame must accommodate linguistic diversity while retaining analytic focus on Arabic-speaking users.

Inclusion criteria. Participants were eligible if they: (i) were aged 18 or above; (ii) resided in the UAE at the time of the survey; (iii) provided informed consent; (iv) were able to complete an online Survey in Arabic or English; and (v) reported access to a voice-enabled device. These criteria align with the country's near-universal internet access and high smartphone penetration, which make online survey administration appropriate and equitable.

Exclusion criteria. Responses were excluded if there was non-consent; more than 20 percent item non-response; failed attention-check items; evident straight-lining; or duplicate tokens suggesting repeated submissions. These quality controls are standard in online survey research and are particularly pertinent in high-connectivity contexts where rapid completion and duplicate access are feasible.

3.6 Sampling, recruitment, and response profile

Sampling design. The study employed stratified convenience sampling to enhance coverage across organizational contexts in which voice technologies are likely to be encountered, specifically public-sector entities such as utilities, higher education and private-sector organizations like retail, services, technology. Stratification at the recruitment stage helps preserve contextual heterogeneity**,** a recommended tactic when a true sampling frame is unavailable, but the phenomenon of interest (exposure to AIVAs) varies meaningfully by context (Creswell & Plano Clark, 2018; Fowler, 2014). This approach improves analytic generalizability by ensuring that key subpopulations are represented even when probability of sampling is infeasible (Teddlie & Yu, 2007). In the UAE, sectoral differences in digitization and customer-contact modalities are pronounced; therefore, sector-based stratification was considered a necessary design control to avoid over-sampling office-centric, high-tech settings and under-sampling front-line service environments.

Probability sampling was not feasible because no comprehensive national registry, sampling frame, or publicly accessible database identifies Arabic-speaking AIVA users residing in the UAE. Consequently, it was not possible to assign each eligible individual a known and equal probability of selection. Stratified convenience sampling was therefore selected as a pragmatic alternative that allowed the study to recruit participants across relevant organisational sectors while preserving diversity in exposure to digital and voice-enabled services.

Recruitment channels. Participants were approached via institutional mailing lists, professional and alumni networks, and gatekeeper-mediated circulation, for example department coordinators. Multi-channel recruitment supports coverage and diversity in respondent experience while maintaining low burden and cost (Dillman, Smyth, & Christian, 2014). To address potential imbalances observed in Round-1, Round-2 recruitment included targeted outreach to female professionals in healthcare/education and older adults (55+), two groups often under-represented in technology-adoption surveys but salient for voice-interface evaluation due to distinct accessibility and privacy expectations (Venkatesh, Thong, & Xu, 2012; Czaja et al., 2006).

Eligibility and screening. Inclusion criteria were age ≥18, current residence in the UAE, and the ability to complete a web survey (Arabic or English). Exclusion criteria were non-consent, >20% item nonresponse, patterned straight-lining, or failure on attention checks. These criteria follow best practice for protecting data quality in online surveys (Tourangeau, Conrad, & Couper, 2013).

Realized samples and composition.

- Round 1 (n = 210): gender 56.2% male / 43.8% female; age concentrated in 25–54 (≈81%); Arabic-mode AIVA usage = 59.5%; cultural-difference salience = 79.5% "yes."
- **Round 2 (n = 459):** more balanced by gender, age, and sector; all respondents reported prior experience with AIVAs, enabling comparable interpretation of latent constructions (such as, Benefits, Challenges).

The Round-2 choice to require prior AIVA experience is methodologically defensible in structural modeling because respondents must have a meaningful frame of reference to evaluate constructs like "design preferences" or "government communication" (Diamantopoulos, Sarstedt, Fuchs, Wilczynski, & Kaiser, 2012).

The two samples were sized according to their different analytical purposes. Round 1 was intended to provide exploratory descriptive evidence and preliminary comparisons concerning usage, language preferences, and perceived challenges. Round 2 required a larger sample because it was designed for confirmatory measurement and structural modelling. The Round 2 sample of 459 participants provided sufficient observations for estimating the latent constructs, examining reliability and validity, assessing model fit, and testing the hypothesised structural relationships.

Nonresponse and coverage considerations. Because a probability frame was not available, traditional response-rate estimation is not meaningful; instead, investigators report sample composition against recruitment targets and document under-coverage mitigations (strata-specific outreach in Round-2). While convenience sampling can introduce selection bias, sectoral stratification and explicit reporting of composition enable readers to judge transferability (Creswell & Plano Clark, 2018). The Discussion chapter interprets findings with these limits in mind.

Accordingly, the samples should not be interpreted as statistically representative of all Arabic-speaking users in the UAE. Although sectoral stratification and targeted recruitment improved demographic and organisational diversity, selection probabilities were unknown and participants were recruited through available networks. The findings therefore provide analytically useful evidence for comparable UAE contexts, but population-level prevalence estimates and unrestricted generalisation should be avoided.

Groups such as individuals with limited digital access, lower digital literacy, speech impairments, or weak connections to institutional and professional networks may remain underrepresented.

Power and precision. For Round-1 t-tests, sensitivity analysis at $\alpha = .05$ and $n \approx 210$ indicates power > .80 for small-to-medium effects (Cohen's $d \approx 0.35$) (Cohen, 1988). For Round-2 SEM, $n = 459$ exceeds common rules of thumb (≥10–20 observations per free parameter), supporting stable estimation, precise fit indices, and multi-group invariance testing by gender and Arabic-usage groups (Kline, 2023; Hair, Black, Babin, & Anderson, 2019).

Mode effects and device context. The instrument was mobile optimized to reflect prevalent usage patterns. Though web-mode surveys can exhibit device-related differences, providing responsive layouts and concise item stems reduces satisficing and improve completion (Dillman et al., 2014). Exploratory checks (reported in the Diagnostics section) did not suggest systematic bias by device type after accounting for age and gender.

Ethical posture. Participation was voluntary, uncompensated, and anonymous. The information sheet specified the study's purpose, minimal risks, benefits, data handling, and the right to withdraw without penalty, consistent with ethical standards for social research and with data minimization principles (BPS, 2021; UAE PDPL—Decree-Law No. 45/2021).

3.7 Instrument development, cultural adaptation, and content validity

3.7.1 Construct maps and item pools.

Each latent construct was defined a priori and operationalized using multi-item scales with **5-point Likert anchors** (1 = strongly disagree to 5 = strongly agree). This design balances respondent burden with psychometric reliability (DeVellis, 2016). Items were adapted from validated sources where possible and **augmented with context-specific content** for Arabic KBS deployments:

- **Perceived Benefits** *(UTAUT2-adapted; 4–6 items)*: performance expectancy (faster task completion), convenience (hands-free), and accessibility in "eyes-busy" contexts; hedonic motivation where relevant (Venkatesh et al., 2012). *Example:* “Using a voice assistant helps me accomplish tasks more quickly.”

- **Perceived Challenges** *(new; 6–8 items)*: recognition errors for **dialects**; failures under **code-switching**; **privacy/data-retention** concerns; lack of **source citations**; mispronunciation of Arabic names/toponyms—issues repeatedly flagged in Arabic ASR/NLU literature (Ali et al., 2021; Bouamor, Habash, & Salameh, 2018). *Example:* “My assistant often misunderstands when I mix Arabic and English.”
- **Design Preferences** *(new; 6–8 items)*: **MSA/vernacular register control**; explicit **citations and last-updated timestamps**; **granular privacy controls**; **whitelisting official sources** for civic answers; **Hijri awareness**; and control over **politeness forms**. These align with transparency and accountability recommendations for public-facing AI (UNESCO, 2021; NIST AI RMF 1.0, 2023). *Example:* “I want the assistant to display the official source and last update for government answers.”
- **Government Communication** *(adapted; 4–6 items)*: clarity, timeliness, completeness, and perceived trustworthiness of voice-mediated public information—dimensions consistent with e-government service quality instruments (Papadomichelaki & Mentzas, 2012). *Example:* “Voice-based government information is clear and trustworthy.”
- **AI_Importance** *(outcome; 3–4 items)*: perceived importance of AIVAs for daily life, work tasks, and access to public information. *Example:* “Voice assistants are important in my daily activities.”

Behavioral covariates captured device context (smartphone, smart speaker, in-car), frequency of use, default interaction language (Arabic vs English), availability of dialect settings, and task types (information lookup, reminders, service queries). Including behavioral covariates aids in **method-effect checks** and supports **measurement validity** by ensuring respondents have relevant experience (Diamantopoulos et al., 2012).

To improve transparency in questionnaire development, Table 3.X maps each Round 2 construct to its observed survey items, conceptual source, adaptation process, and related hypothesis. The individual items were adapted from established technology-adoption and e-government literature where appropriate and supplemented with context-specific items derived from the Arabic AIVA literature and Round 1 findings. The complete wording of the English and Arabic questionnaires is provided in Appendix B.

Table 3-3: Summary of Round 2 Constructs, Measurement Items, Sources, and Related Hypotheses

Construct	Item codes and related content	Main source or basis	Related hypothesis
Effective Usage of AI Voice Assistants (EUAIVA)	EUAIVA_1–EUAIVA_4: frequency, effectiveness, task support, and meaningful use	Adapted from technology-use and behavioural-experience measures; informed by Round 1 usage findings and Diamantopoulos et al. (2012)	H1a, H1b, H1c, H1d
Perceived Challenges (PC)	PC_1–PC_4: dialect recognition, code-switching, privacy, and technical limitations	Context-specific items informed by Round 1 and Arabic ASR/NLU literature, including Ali et al. (2021) and Bouamor et al. (2018)	H1a and H2
Perceived Benefits (PB)	PB_1–PB_3: faster task completion, convenience, accessibility, and usefulness	Adapted mainly from UTAUT2 performance-expectancy and convenience concepts, Venkatesh et al. (2012)	H1b and H3
Design Preferences (DP)	DP_1–DP_4: dialect/register control, source citation, privacy controls, and	Newly developed from Round 1 suggestions, UNESCO (2021), NIST AI RMF (2023), and Arabic-language literature	H1c and H4

	culturally appropriate design		
Communication in Government (CIG)	CIG_1–CIG_4: clarity, timeliness, completeness, and trustworthiness of government information	Adapted from e-government service-quality literature, particularly Papadomichelaki and Mentzas (2012)	H1d and H5
Perceived Importance of AI Services (PIAIS)	PIAIS_1–PIAIS_4: importance in daily life, work, services, and access to information	Outcome scale developed from the study literature and Round 1 findings	H2, H3, H4 and H5

3.7.2 Expert review and content validity index (CVI)

To establish **content validity**, two HCI/Arabic-NLP scholars and two UAE public-service practitioners rated each item on **relevance** and **clarity** (4-point scale). Researcher computed **item-level CVI (I-CVI)** and **scale-level CVI (S-CVI/Ave)**; items with I-CVI < .80 were revised or replaced, following standard thresholds (Polit & Beck, 2006; Lynn, 1986). Reviewers requested:

1. concrete examples of **provenance** (e.g., “ministry portal—updated DD/MM/YYYY”),
2. **neutral wording** for **register** (MSA vs vernacular) to avoid privileging any single dialect, and
3. Clear separation of **privacy** (collection/retention) and **security** (breach/misuse) items.

   Revisions addressed these points and improved **face validity** for Arabic-speaking end users and public-sector stakeholders.

3.7.3 Translation, back-translation, and reconciliation

The instrument was authored in English and **translated into Arabic**, then **back-translated** by independent bilingual translators using the **Brislin** procedure (Brislin, 1970). A reconciliation committee compared versions, focusing on **semantic and conceptual equivalence** while preserving **dialect-neutrality** (Harkness, Villar, & Edwards, 2010). Researcher avoided dialect-specific lexemes in stems and used **generalized or paired examples** (like “salah/azan reminder” rather than region-specific phrasing). This approach reduces **construct contamination** due to regional lexical variation, a known risk in Arabic survey measurement (Bouamor et al., 2018).

3.7.4 Pilot testing and cognitive debriefs.

A pilot (n = 32) assessed **completion time** (target ≤ 15 minutes), **mobile rendering**, and **comprehension**. Short **cognitive interviews** (Willis, 2005) probed interpretations of “code-switching,” “register,” and “authoritative sources.” Edits included:

- adding a brief **definition and example** for **code-switching**,
- inserting a **tooltip** explaining **provenance** and **last-updated**, and
- rewording to avoid **double-barreled** items and to reduce ambiguity. Piloting and cognitive debriefs are standard procedures to surface comprehension problems and reduce **measurement error** prior to fielding (Presser et al., 2004).

3.8 Statistical Analysis Plan and Alignment with Research Questions

The statistical analysis plan was created to match each analytical technique with the intent of the two rounds of surveys and the research questions. The Round 1 analyses were mostly exploratory and descriptive, whereas the Round 2 analyses were confirmatory and assessed the measurement and structural models. Participant characteristics, language preferences, usage patterns, perceived benefits and challenges were summarized using descriptive statistics. The data were analyzed using inferential tests for differences and associations of selected variables and Structural Equation Modelling for relationships between the latent constructs in the final conceptual framework.

Table 3-4: Statistical Analysis Plan and Alignment with Research Questions

Research question	Evidence required	Statistical analysis	Reason for using the analysis
Q1: How are AI voice assistants integrated into the daily activities of Arabic-speaking communities?	Usage frequency, purposes, devices, language preferences, and usage contexts	Frequencies, percentages, means, standard deviations, and cross-tabulations	To describe how participants use AIVAs in their daily activities
Q2: What impact does technical knowledge have on the perceived challenges and benefits associated with AI voice assistant usage in Arab society?	Technical-knowledge levels and perceived challenge and benefit scores	Independent-samples t-test, ANOVA, correlation, and regression, where appropriate	To examine whether technical knowledge is associated with differences in perceived challenges and benefits
Q3: How can AI voice assistants be effectively utilized for communication within this demographic?	Language preferences, communication effectiveness, cultural suitability, trust, privacy, and government communication	Descriptive statistics, correlation analysis, categorisation of open responses, CFA, and SEM	To identify the communication-related factors associated with effective AIVA use
Q4: How can AI voice assistants be designed to better meet the preferences of the Arabic-	Preferred features, design requirements, dialect support, privacy controls,	Frequencies, means, descriptive analysis of suggestions, CFA, and SEM	To identify and test the design features considered important by participants

speaking communities?	and integration	system

Table 3.4 provides evidence that each statistical technique was chosen based on the level of measurement of the variables and the evidence needed to answer the research questions. The alignment of the tests made sure they were not used as standalone statistical tools and that each was relevant to the study's goals and final conceptual design.

Furthermore, the fit of the measurement models and structures was evaluated using more than one fit index, not only one. These included the chi-square to degrees-of-freedom ratio, RMSEA, SRMR, CFI, TLI, IFI, and NFI. Several absolute, incremental, and parsimonious fit measures gave a more balanced judgment of model fit. Four approaches were used to construct overlap: standardised factor loading, Composite Reliability, Average Variance Extracted, and inter-construct correlation, and the Fornell–Larcker criterion. Constructs were empirically defined as different when their square roots of AVE were greater than the correlations among the other constructs.

Cautious interpretation of strong model-fit values was warranted as they could arise from a parsimonious model, strong relationships between indicators, a relatively large sample size, or post-hoc model respecification. Hence the model fit was considered along with reliability, convergent validity, discriminant validity, theoretical coherence and modification index procedure.

3.9 First-Round Survey: SPSS Software and Descriptive Statistics

A total of 210 subjects participated in the first round whose responses were analyzed with the help of SPSS software. This round of the survey aimed at collecting more significant and in-depth information about the demographic preferences, usage patterns, and hindrances among users of AI-based voice assistants within the target communities of Arabic speakers. Survey was developed for distribution among the participants with diverse representation from genders and across different age groups, different education levels, and across different professional sectors. The survey was meant to assess the demographic characteristics of respondents in terms of age, gender distribution, and qualification at the bachelor's or master's level, employment status, and industry representation. The research also compared trends for use of AI voice assistant by the percentage using them, actual length of use, and by the device for accessing them. The baseline on which the trend of adoption is to be measured is first Exposure which reported a high rate at first exposure with 93.8% saying they had tried AI voice assistant and 59.5% used it in actual Arabic, providing a baseline to assess adoption trends.

It also explored functional preferences, and the main use was towards Lifestyle related activities, productive work, and work-related purposes. Obstacles included linguistic and technical ones where the ability to recognize dialects and contextual nuances was of great concern. Respondents further requested improvements in terms of improved natural language processing, stronger integration across devices, and better privacy. These are some of the opportunities by which AI voice assistants may be improved for Arabic speakers.

3.9.1 Research Hypotheses

Round 1 address four hypotheses related to demographic factors in relation to the perceived importance of AI voice assistants in Arabic-speaking communities. The hypotheses investigate the relationships between gender, AI usage, Cultural Differences, and technical knowledge and their effects on the perception of the importance of AI.

H1: There is a significant difference in AI Importance between genders.

- *Relevant Research Question:* How can AI voice assistants be designed to better meet the preferences of the Arabic-speaking communities?
- *Rationale:* AI perception based on gender may suggest inconsistent user preferences and can be utilized in designing desired AI features for different members of Arab society.

H2: There is a significant difference in AI Importance between groups based on AI_VA_Usage_Arabic.

- *Relevant Research Question:* How are AI voice assistants integrated into the daily activities of the Arabic-speaking communities?

- *Rationale:* Assessing AIVA usage reveals information about their adoption trends and the extent to which these are used daily. This allows for identifying potential gaps in their usage.

H3: There is a significant difference in the effective AIVAs utilization for communication between groups based on AI_Cultural_Differences_Arabic.

- *Relevant Research Question:* How can AI voice assistants be effectively utilized for communication within this demographic?
- *Rationale:* Cultural aspects play an important role in technology adoption; thereby, investigating their impact on effective utilization of AI for communication helps in designing culturally adaptive AIVA systems.

H4: There is a significant difference in AI Importance between groups based on AI_Technical_Knowledge.

- *Relevant Research Question:* What impact does technical knowledge have on the perceived challenges and benefits associated with AI voice assistant usage in Arab society?
- *Rationale:* Based on the technical knowledge about AIVAs held by different users, their usability may be affected. Considering this aspect would help recognize challenges and opportunities for improving AIVA adoption.

3.9.2 Dependent and Independent Variables

The hypotheses have been evaluated using multiple independent and dependent variables. Independent variables imply the factors that can possibly influence perceptions of AI whereas dependent variables measure the outcomes

Independent variables are gender (male or female), AI_VA_Usage_Arabic (the usage of AI voice assistants in Arabic by respondents, which was categorized into users and non-users), AI_Cultural_Differences_Arabic (the perception of Cultural Differences by respondents about using AI in Arabic, that is, who are aware of the barriers and those who are not), and AI_Technical_Knowledge (the technical knowledge of AI systems among the respondents, categorized into who are aware and who are not). The dependent variables are AI_Importance, measuring respondents' opinions about how important AI voice assistants are and AI_Developers, measuring respondents' opinion regarding the role and relevance of AI developers. This research seeks to determine the significance of these independent variables toward the two dependent variables, thus enabling it to explain the factors affecting perception of AI in the Arabic-speaking communities.

3.9.3 Demographics of Respondents

Reporting descriptive statistics so that the researcher is familiar with the data and understands the relationships between variables. In summary, a descriptive analysis of respondent profile in terms of Gender, Age, Education level, Employment status and Work sector.

Summary statistics of the Frequencies, Percentages, Mean and Standard deviation for each of the variables in the model are reported in this section.

Table 3-5: Descriptive Statistics of Gender

Gender		Frequency	%	Valid %	Cumulative %
Valid	Female	92	43.8	43.8	43.8
	Male	118	56.2	56.2	100.0
	Total	210	100.0	100.0	

Table 3-5 outlines the gender distribution within the respondent sample, consisting of 210 participants. The data reveals a notable gender imbalance, with 118 individuals identified as male, constituting 56.2% of the total, and 92 individuals identified as female, representing 43.8% of the total sample. This gender breakdown is a crucial aspect of understanding the demographic composition of the study participants. The 56.2% to 43.8% male-to-female ratio provides insights into gender representation in the context of artificial intelligence perception studies.

The significance of gender distribution is particularly relevant in studies exploring perceptions and attitudes towards AI technologies. Analysing responses based on gender can help identify potential variations in how different genders perceive the importance of AI and its developers, contributing to a more nuanced understanding of demographic influences on AI acceptance.

Figure 3-1. Descriptive statistics of gender

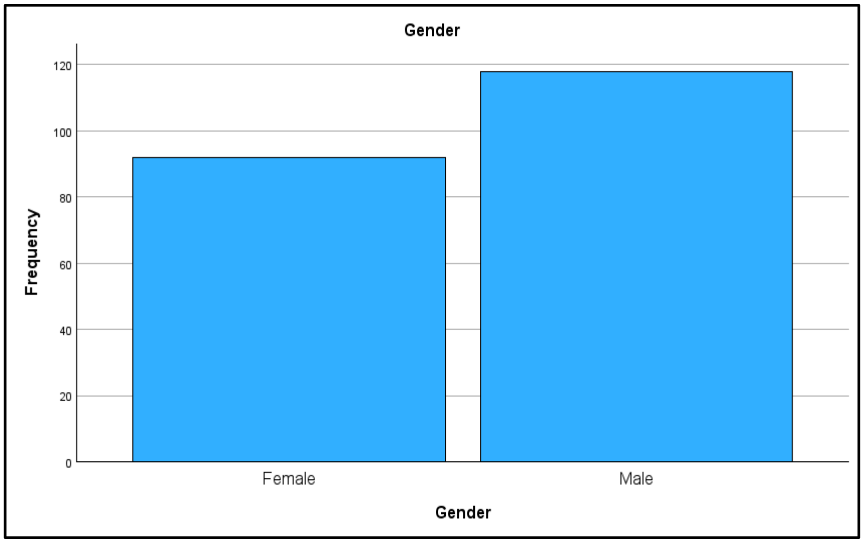


Table 3-6: Descriptive Statistics of Age

Age		Frequency	%	Valid %	Cumulative %
Valid	18-24	29	13.8	13.8	13.8
	25-34	42	20.0	20.0	33.8
	35-44	67	31.9	31.9	65.7
	45-54	62	29.5	29.5	95.2
	55 and above	10	4.8	4.8	100.0
	Total	210	100.0	100.0	

The table 3.6 provides an overview of the age distribution among the respondents, comprising a total sample size of 210 participants. Understanding the age demographics is crucial, as different age groups may exhibit varying levels of familiarity, comfort, and acceptance of technological advancements. The largest contribution to the survey. 31.9%. comes from the age group 35-to-44-year followed by 45-to-54-year group with 29.5% share. While 25-to-34-year stood at 20% followed by 18-to-24-year and 55 above with 13.8% and 4.8 % respectively. Overall, the majority of respondents fall within the 25-54 age range, comprising 81.4% of the total sample. This concentration suggests that the study captures perspectives predominantly from an early-career to mid-career demographic, potentially influencing the findings regarding perceptions of AI's importance and its developers. This distribution provides valuable insights into the generational representation within the sample, allowing for a nuanced exploration of how different age cohorts may perceive and engage with AI technologies.

Figure 3-2. Descriptive statistics of age

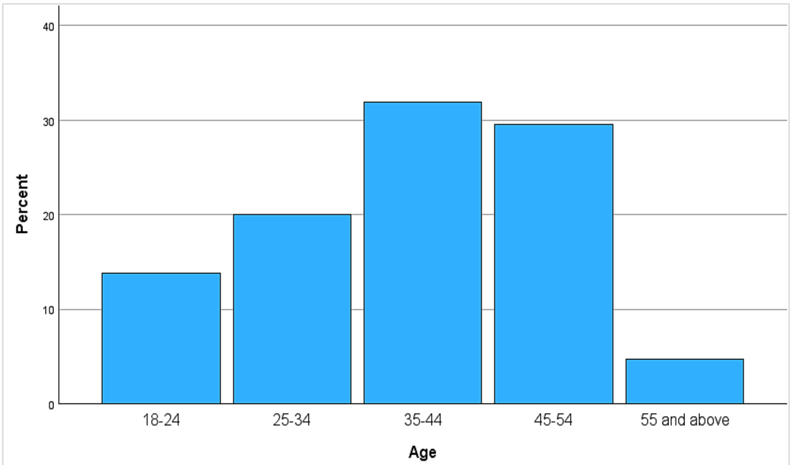


Table 3-7: Descriptive statistics of AIVA_Usage_ Arabic

		Frequency	%	Valid %	Cumulative %
Valid	No	85	40.5	40.5	40.5
	Yes	125	59.5	59.5	100.0
	Total	210	100.0	100.0	

The table 3.7 provides an overview of the usage of AI virtual assistants in Arabic among the respondents, comprising a total sample size of 210 participants. The data highlights whether individuals currently use AI virtual assistants in Arabic, illuminating an important aspect of technological adoption within the sample population. The higher usage rate (59.5%) indicates a significant inclination towards adopting these technologies, suggesting a general openness or a necessity driven by the linguistic and functional preferences in AI interactions within Arabic-speaking contexts.

This distribution offers valuable insights into the acceptance and integration of AI technologies within the community. Understanding this usage pattern is crucial, as it reflects not only the technological engagement of the respondents, but also potential cultural acceptance of AI solutions tailored to meet specific linguistic needs. The predominance of users over non-users in this category could influence how AI technologies are further developed, marketed, and deployed in regions with significant Arabic-speaking populations.

Figure 3-3. Descriptive statistics of AIVA_Usage_ Arabic

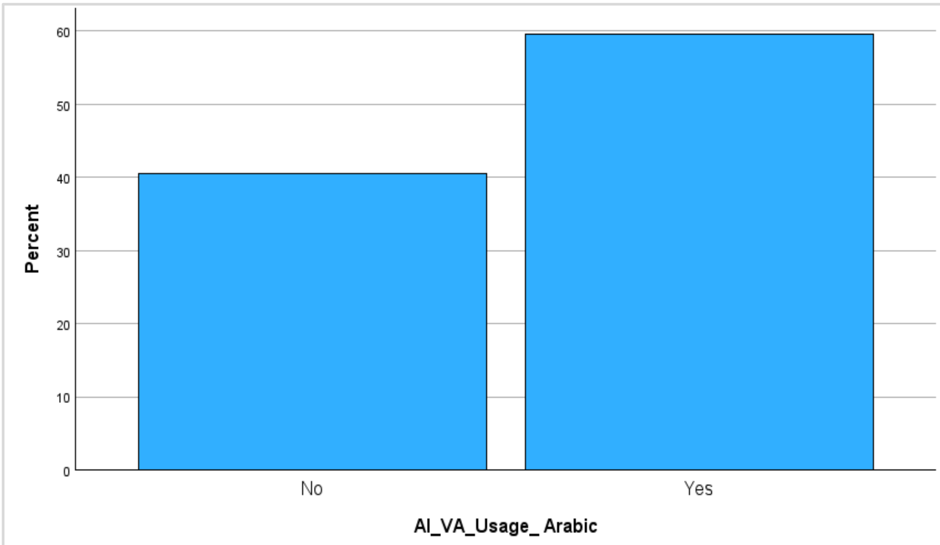


Table 3-8: Descriptive Statistics of AI_Cultural_Differences_Arabic

		Frequency	%	Valid %	Cumulative %
Valid	No	43	20.5	20.5	20.5
	Yes	167	79.5	79.5	100.0
	Total	210	100.0	100.0	

Table 3.8 provides an overview of the perceptions regarding cultural differences to AI usage in Arabic among the respondents. This data is pivotal in understanding how cultural factors influence the acceptance and integration of AI technologies within Arabic-speaking communities. The distribution is critical, indicating a substantial perception of Cultural Differences among the majority (79.5%) of the sample. Such a significant proportion suggests deep-rooted cultural concerns or challenges that may hinder the broader acceptance and effectiveness of AI technologies tailored for Arabic-speaking users. The recognition of these differences is essential for AI developers and end users as it underscores the need for culturally sensitive approaches in the development of AI technologies. Addressing these concerns can lead to more effective communication,

higher acceptance rates, and ultimately, a more profound integration of AI within the societal fabric.

Figure 3-4. Descriptive statistics of AI_Cultural_Differences_Arabic.

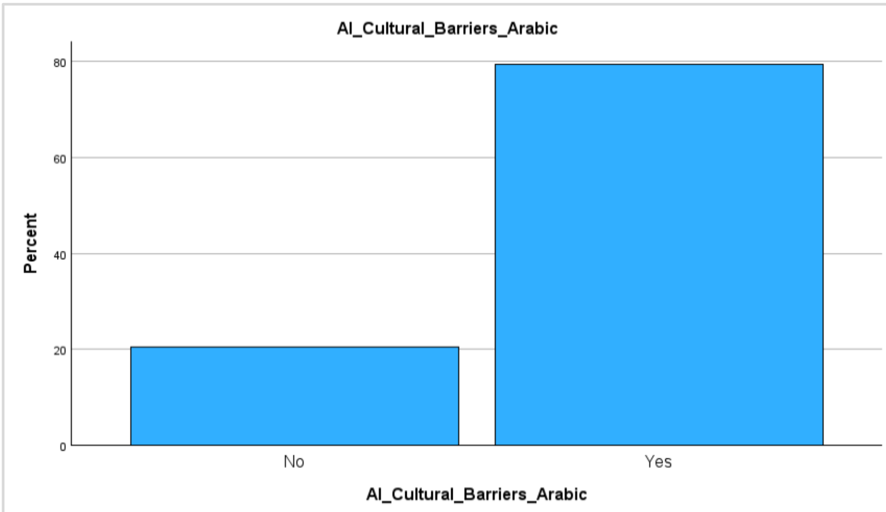


Table 3-9: Descriptive statistics of AI_Technical_KBS

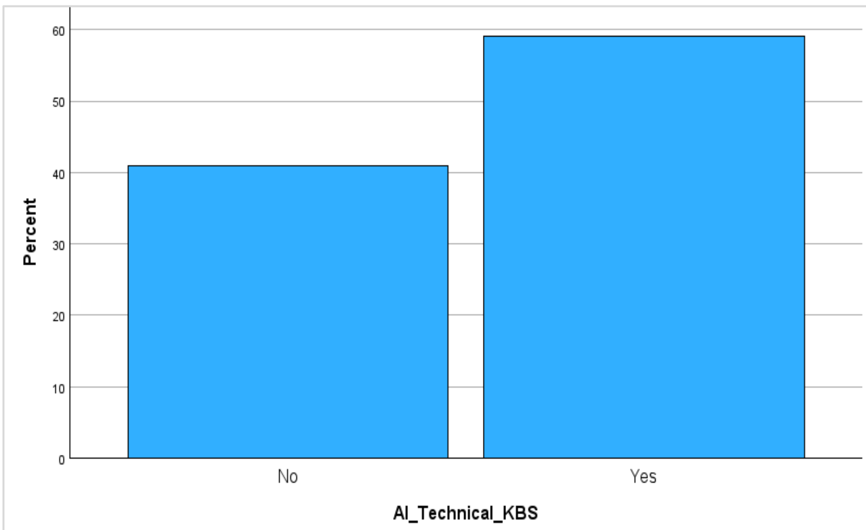
		Frequency	%	Valid %	Cumulative %
Valid	No	86	41.0	41.0	41.0
	Yes	124	59.0	59.0	100.0
	Total	**210**	**100.0**	**100.0**	

Table 3.9 presents the descriptive statistics for the variable AI_Technical_KBS, which measures the technical knowledge about AI-based systems among the participants. The data illustrates that a majority of the respondents (59.0%) report having technical knowledge about AI-based systems. This is indicative of a relatively high level of awareness or familiarity with AI technologies within the sampled population, which could potentially influence their perceptions of AI's importance and its developers. The higher percentage of participants with technical knowledge might be reflective of a technologically inclined sample or a demographic that is generally more engaged with or interested in technological developments. This demographic characteristic can significantly impact how AI technologies are perceived, especially concerning their capabilities and limitations. AI_Technical_KBS variable is crucial in understanding the

community's readiness to integrate and accept AI technologies. It also points to a potential for higher acceptance rates and a more profound trust in AI applications, as familiarity with technology typically correlates with reduced apprehension and increased adoption.

For AI developers and end users these insights are invaluable. They suggest that initiatives aimed at increasing technical understanding of AI could further enhance positive perceptions and acceptance of AI technologies. Educational campaigns that demystify AI and expose more individuals to its benefits and limitations could be particularly effective in regions or demographics showing lower familiarity levels.

Figure 3-5. Descriptive statistics of AI_Technical_KBS



The analysis sought to understand demographics, preferences, and challenges of participants in the use of AI Voice Assistant within Arabic-speaking communities. Indeed, the analysis shows that the highly educated middle-aged population dominantly uses AI voice assistants, though 93.8% have tried while 59.5% can use AI voice assistants in Arabic, which means challenges still exist with regards to the Arabic language and its dialectical complexity. Indeed, it reflects a strong preference for brands in the world's frontline, such as Amazon Alexa and Apple Siri, at one and the same time carving a niche for localized solutions to specific landscape in terms of language and culture pertinent to Arabic-speaking people. Patterns of use indicate that whereas voice assistants are becoming more familiar with users, they have yet to become integral to daily routines for most users.

These assistants are used mainly for lifestyle (Entertainment + Shopping), followed by productivity tasks that show their role in improving leisure and work-related activities. However, a number of technical challenges-like difficulty in understanding natural language and accents, Cultural Differences related to dialect variations, social norms-affect their effectiveness and adoption. That positive perceptions of these developers of AI should correlate with the perceived importance ascribed to AI itself suggests a belief in and trust of the developers are significant in the wider reception of AI technologies. Such challenges, if identified, need to be overcome to advance natural language processing for Arabic and further culturally sensitive AI interactions. Hence, this goes a long way, not only in improving functionality but also in further engaging and satisfying Arabic-speaking users.

3.10 Statistical Analysis Procedures for Hypothesis Testing

The results of the four hypotheses tested in this round have been presented at the conference. They provide important insights into what influences the use and views of AI voice assistants in Arabic-speaking areas. These results also support initial data analysis and make clear AI adoption in this situation.

H1: There is a significant difference in AI Importance between genders

Result: The analysis had a huge difference between the males and females. Females rated the importance of AI to be higher compared to males, $t = 2.178$, $p = 0.031$ indicating that gender is relevant to the nature by which people are perceiving the importance of AI.

H2: There is a significant difference in AI Importance between groups based on AI_VA_Usage_Arabic.

Result: no significant difference between the people who use Arabic AI voice assistants and those who do not ($t = 1.298$, $p = 0.196$) has been found. it shows that how often people use does not deeply affect how important they consider AI is.

H3: There is a significant difference in AI Importance between groups based on AI_Cultural_Differences_Arabic.

Result: The individuals who did not identify any sort of cultural barrier perceive the AI more impactful than those people who identified Cultural Differences ($t = 2.593$, $p = 0.012$).

H4: There is a significant difference in AI Importance between groups based on AI_Technical_Knowledge.

Result: There was no huge difference based on the technical knowledge of the participants (t = -1.492, p = 0.137). This suggests that although being aware of technology may vary the extent to which individuals are involved, it usually does not change how important it is.

Table 3-10: Round 1 hypothesis and results

PARAMETER	DETAILS
Sample Size	210
Number of Surveys	1 (with 30 questions)
Number of Tested Hypotheses	4 (using two independent sample T-Test)
H1: Significant difference in AI importance between genders	Result: Females rated AI importance higher than males
H2: Significant difference in AI importance between users & non-users of AI voice assistants in Arabic	Result: No significant difference observed
H3: Significant difference in AI importance based on perceived Cultural Differences	Result: Those who perceived no Cultural Differences rated AI as more important
H4: Significant difference in AI importance based on technical knowledge	Result: No significant difference observed

These findings are significant as they help reveal the role of gender and Cultural Differences in defining perceived value for AI. Further rounds of this study also have the potential to further elaborate these factors. In that regard, its role serves as critical evidence for ongoing analysis and future development of AI systems tailored for Arabic-speaking communities.

3.11 Second-Round Survey: SPSS and AMOS Analysis

The first round of the survey was intended to collect detailed information about the demography, preferences, usage behavior, and problems that the users of AI voice assistants face in Arabic-speaking countries. A round two survey was then conducted on the sample of 459 participants in order to focus on insights related to user preferences, challenges, and suggestions for improving AI voice assistants by Arabic-speaking users. Round two added depth to the first round, considering additional variables and seeking qualitative feedback in order to better grasp the subtleties surrounding AI adoption in this target group.

It was intended to validate trends found in the first round and look at new facets related to *primary purposes of using AI voice assistants* and *perceptions regarding technical and cultural obstacles.* Specifically, Round Two was designed to test whether the hypotheses proposed in this study such as the influence of cultural and dialectal differences on adoption rates, and the role of user experience in shaping satisfaction with AI voice assistants hold true across a broader dataset. This stage also examined whether improvements suggested in the first round align with user expectations when tested against additional variables.

Participants were only those experienced with the use of AI voice assistants to maintain and compare their responses between the two rounds. Structured surveys were gathered, and data were analyzed in SPSS software using a comparative and descriptive statistical approach.

3.11.1 Second-Round Data Analysis Approach

Proper utilization of AI voice assistants revealed the outcomes of round two significantly impacting perceived challenges, benefits, design preferences, and communication in government by having standardized beta values that range between 0.654 to 0.713 while accounting for a variance of R^2 with 42.8% to 50.8% in each respective construct. These factors, in turn, positively influence the perceived importance of AI services. Of these, challenges ($\beta = 0.216$), benefits ($\beta = 0.278$), design preferences ($\beta = 0.204$), and government communication ($\beta = 0.312$) make significant contributions to this factor, accounting for 61.8% of the explained variance. The model fit indices support an excellent fit of the model, with the following results: RMSEA = 0.020, CFI = 0.993.

Demographically, the sample with N = 459 is well balanced on gender, age, and education; that is, mostly educated professionals; however, there are still diverse perspectives in this group. The descriptive statistics for generally positive attitudes toward AI voice assistants were revealed in mean scores between 3.49 and 3.54 on all the constructs, thus showing respondents having moderately favorable perceptions toward them. It can be seen that second round of the survey pinpointed particular places for further refinement.

It can be inferred that the second round of the survey added much depth to the analysis as it reemphasized the trends in the first round and pinpointed particular places for further refinement. Localization of AI solutions becomes of greater importance, while improved linguistic support is urgently required, as well as an influential role for privacy as an area determining user perception. The input developed here would be essential to developers and end users alike when designing more effective and culturally aligned AI voice assistants for Arabic-speaking users.

3.11.2 Use of AMOS Modification Indices

AMOS was used to conduct Confirmatory Factor Analysis (CFA) and Structural Equation Modelling (SEM) to test both the measurement model and structural model. In the model evaluation process, AMOS offers Modification Indices (MIs) as diagnostic data which can point to the possible locations of localized model misfit. The Modification Indices were not utilized in determining post-hoc alterations to the hypothesised model in this study, but were only reviewed as a diagnostic tool. The theoretical framework that was formulated based on the literature and the predetermined hypotheses of the research was consistent with the measurement and structural models.

Based on this, no other structural paths were added, no questionnaire items were dropped and no correlated error terms were included just to enhance the statistical fit of the model. The last SEM model thus reflects the initially suggested conceptual framework instead of a data model developed out of an iterative stream of changes. This method made sure that the correlation between the latent constructs was theoretically grounded and aligned to the research goals.

The risk of overfitting was also minimized through avoiding the extensive post-hoc model changes. Overfitting happens when a model is over-fitted to the properties of a specific sample, thus becoming less representative of the underlying theoretical relations and less useful in generalising to other populations. The study kept the theory-based model specification intact thus ensuring that the conceptual integrity of the proposed framework was maintained and that the reported model fit indices were an accurate measure of the true consistency between the observed data and the hypothesised model as opposed to increases due to the adjustments of the sample.

3.11.3 Assessment of Model Fit and Construct Distinctiveness

The Structural Equation Model (SEM) was tested by considering a combination of absolute, incremental, and parsimonious fit indices as opposed to a single goodness-of-fit statistic. The model showed a good fit in the indices recommended, which shows that the hypothesised measurement and structural models were good in their representation of the observed data. The good fit of the model is because of the strict model development process taken in this study. The conceptual framework was developed with regards to a vast literature review and all measurement items were developed out of already validated literature. Moreover, the questionnaire was also reviewed and pilot tested by experts prior to the actual survey, improving the clarity, relevance, and validity of measurement items. Before estimating the structural model, confirmatory Factor Analysis (CFA) was then carried out to ensure that the observed indicators were representative enough of their respective latent constructs.

Construct validity was carefully considered to ensure that the latent constructs were not overlapping and thus before the interpretation of the structural relationships. Satisfactory standardised factor loadings, Composite Reliability (CR) and Average Variance Extracted (AVE) were used to determine convergent validity. Discriminant validity was evaluated based on the FornellLarcker criterion of which the square root of the AVE of each construct was larger than its correlations with other constructs. This shows that individual latent constructs had higher variance levels with their indicators compared to other constructs and that the constructs measured different theoretical concepts and not the same dimensions. The outstanding SEM model fit, therefore, represents the conformity of the theoretical model with the observed data and enough distinctiveness of the constructs.

3.12 Research Ethics

This study is designed and conducted in accordance with established human-subjects ethics and contemporary AI governance guidance. Core principles respect for persons, beneficence, and justice inform every stage of the project (Belmont Report). In addition, researcher align procedures with the UAE Federal Decree-Law No. 45 of 2021 on the Protection of Personal Data (PDPL), the UNESCO Recommendation on the Ethics of AI (2021), the OECD AI Principles (2019), and the NIST AI Risk Management Framework 1.0 (2023). These frameworks collectively require transparency, purpose limitation, data minimization, fairness, and human oversight in AI-related research and deployments. The following ethical guidelines have been followed in this study:

- *Informed consent:* Prior to participation, all respondents receive an information sheet (Arabic and English) written in clear, non-technical language. It states the study's purpose, procedures (two survey waves with optional open-ended items), expected duration, minimal risks, anticipated benefits, data uses, contact details for the research team and the ethics office, and an explicit statement of voluntariness. Proceeding to the survey constitutes consent; participants may withdraw at any time without penalty and may request deletion of their data up to the point of anonymization (Belmont; UNESCO, 2021).
- *Respect for culture and language.* Because the research focuses on Arabic-speaking participants and AI voice assistants, an investigator apply culturally responsive methods: (i) instruments are available in Arabic and English using translation/back-translation to preserve meaning; (ii) examples and prompts avoid dialectal stereotyping and derogatory language; (iii) the survey permits participants to skip any item they find uncomfortable; and (iv) open-ended content is screened to remove inadvertently disclosed sensitive information about third parties. This approach implements UNESCO's call to safeguard cultural and linguistic diversity in AI research and applications (UNESCO, 2021; OECD, 2019).

- *Eligibility and protection of potentially vulnerable persons.* The study includes adults (≥18) residing in the UAE. An investigator excludes minors and avoids recruiting individuals whose institutional dependency could compromise voluntariness like direct supervisees. There are no deception and no financial inducement; participation is uncompensated to minimize undue influence. Accessibility accommodations as larger fonts, screen-reader-compatible layouts are provided to promote equitable participation (justice principle).
- *Privacy by design and data minimization.* In line with the UAE PDPL, investigator collect only what is necessary for the stated aims: coarse demographics (age band, gender, sector), self-reported usage patterns, perceptions, and optional qualitative comments. Researcher do not collect direct identifiers (names, national IDs) and do not retain IP addresses in the analytic dataset. This data minimization reduces re-identification risk and aligns with purpose limitation requirements (PDPL Art. 7–9).
- *Confidentiality, security, and retention.* Survey data are transmitted over HTTPS and stored on encrypted drives with access restricted to named researchers on a least-privilege basis. A separate key file linking recruitment tokens (used only to prevent duplicates) to responses is destroyed after validation, leaving a fully de-identified analytic file. Researchers retain de-identified data for five years to enable verification and secondary analysis consistent with the original purpose, after which researcher securely delete them. Security controls and retention periods are documented in a data-management plan; any material change requires ethics approval (PDPL; NIST AI RMF, 2023).
- *Handling of open-ended text*. Free-text responses are manually reviewed to remove names, phone numbers, addresses, employer names, or other direct identifiers before analysis. When quotes are reported, they are paraphrased or masked to preserve meaning without enabling identification. This protects privacy while allowing thematically rich, culturally relevant insights.
- *Risk assessment and mitigation*. Anticipated risks are minimal (time, mild discomfort when reflecting on technology concerns). To mitigate, items about privacy or negative experiences are optional; participants can skip any question and can withdraw entirely. If a participant communicates distress or a complaint, a documented incident

response protocol routes the case to the PI and the ethics office; no clinical or legal advice is provided by the research team.

- *Fairness and representativeness.* Although probability sampling is infeasible, researcher mitigate selection bias by sectoral stratification and targeted outreach to groups often under-represented in technology studies (for example, women in healthcare/education; adults 55+). Researcher report sample composition transparently so readers can judge transferability. This addresses the justice principle and OECD guidance on inclusive AI (OECD, 2019).
- *Algorithmic and analytic integrity.* While this thesis does not deploy a live AIVA, it studies systems that may shape access to public knowledge. In analyzing survey data, an investigator (i) preregisters or timestamp the analysis plan where feasible, (ii) report model assumptions and limitations, and (iii) avoid over-claiming causal relations from cross-sectional data. For any secondary analyses involving automated tools like topic modeling on de-identified text, researcher document parameters and evaluate potential bias amplification, consistent with NIST AI RMF's measurement and monitoring guidance (NIST, 2023).
- *Transparency, provenance, and communication.* In publications and the thesis, researcher disclose methods, recruitment, inclusion/exclusion criteria, missing-data handling, and statistical choices. Summary results have been made available on request in aggregate form; no individual-level data have been shared outside the research team unless required by law and approved by the ethics board. Any future secondary use of the collected data will require separate ethical approval and, where applicable, renewed participant consent.
- *Governance and oversight.* The protocol is submitted for review to the institutional research ethics committee/IRB. Material amendments as like instrument changes, new data uses will not proceed without prior approval. The PI is responsible for compliance with PDPL, record-keeping, and staff training on privacy and security controls.

Hence, the project operationalizes human-subjects ethics and AI governance through concrete safeguards informed consent, cultural-linguistic care, privacy-by-design, secure anonymization, fair recruitment, transparent analysis, and accountable oversight to ensure that investigating AIVAs in knowledge-based contexts benefits Arabic-speaking communities while respecting their rights and values (UNESCO, 2021; OECD, 2019; NIST, 2023).

3.13 Chapter Summary

This chapter explained the methodological approach adopted to investigate Arabic-speaking users' experiences and perceptions of AI voice assistants in the UAE. It outlined the research paradigm, two-round cross-sectional design, study setting, sampling and recruitment strategy, eligibility criteria, instrument development, translation and cultural-adaptation procedures, pilot testing, and planned analytical methods. Round 1 was designed to provide exploratory evidence concerning usage and perceptions, while Round 2 was designed to assess the measurement model and test the proposed structural relationships. The chapter also described the ethical procedures adopted, including informed consent, confidentiality, data minimisation, secure storage, and compliance with the UAE Personal Data Protection Law. Chapter 4 presents the empirical findings generated through these procedures.

4. CHAPTER 4: RESULTS

The results of the two survey rounds in this study are given in Chapter 4. It highlights demographic information about participants, usage patterns of AI voice assistants, language preferences, perceived benefits and challenges, cultural and technical issues, and recommendations for improvement. The chapter also provides the measurement and structural model results in examining the relationships between the selected constructs. The results are presented in a systematic way and point out how the results of the empirical research answer the research questions and support the research purpose.

The two survey rounds are analyzed separately since the data was collected for different purposes and is not referred to as repeated measurements. The results of the two rounds are interpreted as exploratory evidence about usage and perception and confirmatory evidence about the final SEM model, respectively. No attempt was made to make direct statistical comparisons between the two rounds. To guide the interpretation of the findings, the Round 1 results are presented in relation to the four research questions. Q1 is driven by usage, frequency, purposes, devices and language selections. The components of technical knowledge, perceived benefits, and perceived challenges make up Q2. Q3 is influenced by the effectiveness of communication, cultural appropriateness, privacy and language preferences. Relevant features and suggestions from respondents are included in Q4. The final SEM model then examines the relationships between the selected constructs as indicated by the findings of the Round 2.

4.1 Round 1: Exploratory Survey Results

This section presents the exploratory findings from the first-round survey of 210 respondents. Round 1 was designed to identify patterns of AIVA exposure, usage, functional purposes, language preferences, perceived challenges, cultural concerns, technical knowledge, and recommendations for improvement. The descriptive and inferential findings reported in this section provide exploratory evidence concerning the experiences and perceptions of Arabic-speaking users. They should not be interpreted as results from the final SEM model, which was tested separately using the Round 2 dataset.

4.1.1 Demographic Analysis

The demographic analysis presented in this section is based on the first-round survey, which involved 210 respondents with a slight majority of males: 56.2% male and 43.8% female. This nearly equal gender distribution can ensure findings are represented for both genders in the usage of AI Voice Assistants within Arabic-speaking communities. Table 4-1 provides the background of the Round 1 sample to which subsequent findings of usage and perception should be related. The high proportion of working age and highly educated individuals in the sample might explain the familiarity with digital technologies. The demographic findings are context data, however, and will not answer the research question on their own, but will help to explain differences in AIVA usage, technical knowledge and perceived importance.

4.1.2 Age Groups

The age factor shows that the sample size was composed of mainly working age adults with the majority of the people surveyed falling in the middle age groups. The profile indicates that the participants were probably exposed to digital technologies regularly, both at work and at home. The depiction of various age groups is also a chance to think whether age-related variations can affect the patterns of AI voice assistant use and attitudes.

4.1.3 Education Levels

The respondents were mostly highly educated, and most of them had undergraduate or postgraduate degrees. Such an educational profile implies a sample that may be more familiar with digital technologies and new AI applications. Education level distribution serves as a good background to understand the results regarding technical knowledge, adoptions, and attitudes to AI voice assistants.

4.1.4 Work Status and Industries

57.1% were employed full-time, while the next largest category was students. The education sector is represented the most, with 51% representation; it is followed by the information technology sector by 17.1% and the government sector by 13.3%. This distribution shows a high level of education, as well as professional diversity of the respondents; it conveys a glimpse into how AI voice assistants might be viewed across different professional settings. Statistics of the Respondents are shown in appendix A.

Table 4-1: Summary Profile of Round 1 Respondents

Demographic characteristic	Category	Frequency (n)	Percentage (%)
Gender	Female	92	43.8
	Male	118	56.2
Age	18–24	29	13.8
	25–34	42	20.0
	35–44	67	31.9
	45–54	62	29.5
	55 and above	10	4.8
Education	High-school qualification	16	7.6
	Bachelor's degree	106	50.5
	Master's degree	58	27.6
	Doctoral degree	30	14.3

4.2 AI Voice Assistant Usage

4.2.1 Initial Exposure and Current Usage

The results indicate a clear difference between general exposure to AIVAs and their sustained use in Arabic. Although 93.8% of respondents reported having previously used an AI voice assistant, only 59.5% reported currently using one in Arabic. This difference suggests that widespread familiarity with AIVAs has not translated into equally widespread Arabic-language engagement. As the survey measured reported usage rather than technical performance, the pattern may reflect a combination of language preferences, available device ecosystems, perceived usefulness, dialect suitability, and user experience.

4.2.2 Frequency and Duration of Usage

Among current users, usage frequency varies, with 37.6% using it monthly and 12.4% weekly. Daily users are a small fraction (3.8%). Many have been using these assistants for about 3-5 years (34.8%) or 1-3 years (19.5%). This trend suggests growing familiarity and acceptance of AI voice assistants over time. Their wide usage can be highlighted by the fact that 93.8% of respondents have used an AI voice assistant, whereas only 59.5% currently do so in Arabic. The reason for this apparent discrepancy will no doubt lie in the diversity of dialects and the subtlety of the Arabic language. Of these current users, 37.6% use it on a monthly basis, 12.4% use them on a weekly basis, and only 3.8% are daily users. This would suggest that while many people know about the technology, for most it is yet to become an important part of daily routines.

The results show, on one hand, a certain trend that corroborates the fast-increasing familiarity of users with AI Voice Assistants. On the other hand, it encourages going further in solving recurrent challenges for their consistent and extended home use, particularly with regard to Arabic language support. Ensuring these challenges are overcome will be of great value in enhancing the user experience and encouraging higher frequencies of engagement with AI voice technologies.

As shown in Appendix B, the data reveal that exposure to AI voice assistants among the respondents is really high: 93.8% of the respondents reported using one ever, while when it comes to their regular use in Arabic, the percentage drops to 59.5%. This suggests that many people know about voice assistants, not everybody continuously uses them in Arabic; probably this is because of a limitation in the capability of the technology to handle its intricacy along with its diverse dialects. The frequency of the use of AI Voice Assistants in Arabic by the users varies: the largest group uses it once a month - 37.6%, followed by 12.4% using it on a weekly basis, and 3.8% use it daily. What this infers is that for most users, AI voice assistants are not yet part of their daily tool kit, thus functionality or user experience needs further improvement in order to attract more frequent uses.

Further, 34.8% have used AI voice assistants for some three to five years, followed by 19.5% who have used them for between 1 to 3 years. Only 5.2% are recent adopters who have been using the technology for less than 6 months. Although this pattern does hint at gradual adoption and a rise in the acceptance of AI voice assistants with time, it completes the view by suggesting that how to attract new users is still one of the challenges. Improvement in practical language processing, along with highly relevant use cases, can foster an expansion in the user base and a higher frequency of use.

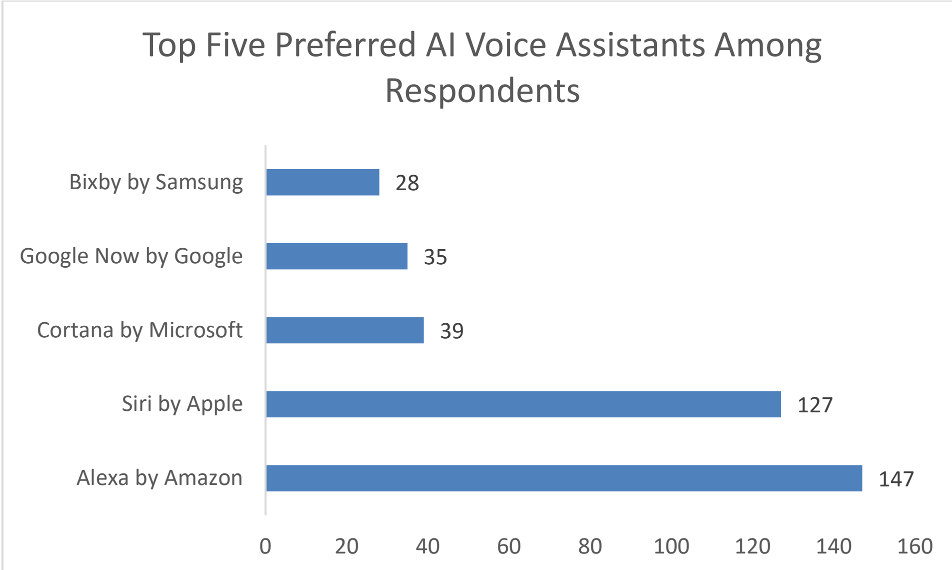
These findings directly contribute to Q1 concerning the integration of AI voice assistants into daily activities. High initial exposure indicates that AIVAs were familiar to most respondents, but the lower level of Arabic-language use and the small proportion of daily users show that the technology had not become fully embedded in everyday routines. The difference between initial trial and sustained Arabic use suggests that availability alone is insufficient; continued integration also depends on language capability, practical usefulness, and a satisfactory user experience. As these findings are based on respondents' reported usage, they indicate perceived adoption patterns rather than independently measured system performance.

4.3 AI Voice Assistant Preferences and Devices

4.3.1 Preferred AI Voice Assistants

Figure 4-2 shows the five most preferred AI voice assistants by respondents. The results indicate that the most preferred assistant is Alexa by Amazon and then Siri by Apple. Cortana, a Microsoft product, Google Now, and Bixby were other popular assistants that were preferred.

Figure 4-1: Top Five Preferred AI Voice Assistants among Respondents



The preference pattern shows that the interaction of the respondents with AI voice assistants is principally focused on known international technology vendors. This dominance can be based on the prevalence of these assistants via the ubiquitous smartphones, smart devices and existing digital ecosystems. Despite the fact that the outcomes indicate high adoption of international platforms, existence of various assistants demonstrates that there is a difference in the exposure and preference of the voice-based technologies to users.

These results provide an addition to Research Question 1 (Q1) in determining the primary platforms on which respondents use AI voice assistant technologies in their everyday lives. But the fact that one assistant is preferred over another should not be taken to mean that he is more capable of using the Arabic language or that he is more culturally appropriate. Rather it is a measure of the accessibility, familiarity, and integration of devices of respondents. More in-depth linguistic issues, cultural anticipations, and user needs should be analyzed to examine how well these assistants can serve Arabic-speaking users.

4.3.2 Devices Used for AI Voice Assistants

Mobile apps are the predominant medium (16.2%), often combined with other devices like laptops and smartwatches. There is also notable usage of smart speakers, suggesting that voice assistants are integrated into various aspects of daily technology use. Amazon Alexa is the most preferred among the respondents, followed by Apple Siri, Samsung Bixby, and local ones such as RafiQ and Salma. That said, Alexa and Siri head the list; preference towards global brands shows commendable presence and effectiveness, while the notable consumption of local assistants points to huge demand for AI solutions catering more appropriately to the language and culture of Arabic speakers. The difference in preference Insists on developing voice assistants that provide quality language processing, at the same time fitting the unique contexts of regional users. As far as devices are concerned, Mobile apps are the leading platform where users utilize the AI voice assistants, at 16.2%, in conjunction with other devices like laptops and smartwatches, demonstrating multi-device usage across the subjects polled. Furthermore, the growth in the adoption of smart speakers shows that these are becoming an inseparable part of home technology ecosystems, providing convenience and accessibility for use everywhere. Such widespread adoption across devices would further suggest that users find in voice assistants the flexibility and versatility to be embedded through multiple touchpoints in their daily life. This simply underlines that this trend has the need for continuous innovation both in hardware and software, to meet users' expectations.

What devices do you normally use for AI voice assistant(s)?

Table 4-2: Devices for AI Voice Assistants Usage

Devices		**Frequency**	**%**
Valid	Personal computer (PC)	6	2.9
	PC, smart TV	1	.5
	PC, smart watch	2	1.0
	Mobile apps	34	16.2
	Mobile apps, PC	24	11.4
	Mobile apps, PC, smart TV	21	10.0
	Mobile apps, PC, smart watch	21	10.0

Mobile apps, PC, Smart watch &TV	5	2.4
Mobile apps	1	.5
Mobile apps, smart speaker (Google Home, Amazon Echo)	2	1.0
Mobile apps, smart speaker, PC	6	2.9
Mobile apps, smart speaker & TV, PC	1	.5
Mobile apps, smart speaker, PC, smart watch	2	1.0
Mobile apps, smart speaker, PC, smart watch, smart TV	4	1.9
Mobile apps, smart speaker, smart TV	4	1.9
Mobile apps, smart speaker, smart watch	7	3.3
Mobile apps, smart speaker, smart watch, smart TV	5	2.4
Mobile apps, smart TV	13	6.2
Mobile apps, smart watch	14	6.7
Mobile apps, smart watch, smart TV	19	9.0
Smart speaker	5	2.4
Smart speaker, PC	2	1.0
Smart speaker, PC, smart TV	2	1.0
Smart speaker, smart TV	2	1.0
Smart speaker, smart watch	2	1.0
Smart TV	1	.5
Smart TV	1	.5
Smart watch	2	1.0
Smart watch, smart TV	1	.5

Total		***210***	***100***

The device results further answer Q1 by showing the contexts in which respondents accessed AIVAs. Predominant mobile use indicates that voice assistants were mainly integrated through personal and portable devices, while multi-device and smart-home use suggests broader integration for a smaller group of respondents. Therefore, daily AIVA use was concentrated primarily in readily available personal-device environments rather than being uniformly embedded across homes, workplaces, and organisational systems.

4.4 Primary Uses of AI Voice Assistants

4.4.1 Functional Preferences

Lifestyle-related activities, combining entertainment and shopping purposes, represented the most common category of AI voice assistant use among respondents (32.4%). This indicates that respondents primarily engaged with AI voice assistants for convenience-oriented daily activities. Work-related tasks represented another important usage category (13.8%), while productivity activities such as setting reminders and scheduling events further demonstrated the role of AI voice assistants in supporting everyday organisation and personal management.

The reported purposes directly answer Q1 by showing how AIVAs were integrated into participants' activities. Uses relating to information seeking, communication, productivity, entertainment, shopping, and lifestyle support indicate that respondents employed AIVAs primarily for convenient, task-specific assistance. However, the comparatively limited daily usage reported earlier suggests that these activities were supplementary rather than fully embedded components of most respondents' routines. The detailed response combinations are presented in Appendix F, while the main chapter emphasises the broader functional categories.

4.4.2 Language Preferences

English is overwhelmingly the language of choice (97.6%), with a small fraction using Arabic (2.4%). However, among those who have not used AI voice assistants, 38.1% express a preference for using both English and Arabic, indicating a demand for improved Arabic language support. The positions are topped by lifestyle related activities entertainment and shopping, each contributing 16.2% of the preferences. This means that many find these tools quite useful for leisure purposes, like playing music, watching videos, and making purchases online. Work-related tasks have also been a significant case for application, taking 13.8%, where voice assistants have set appointments, sent messages, and managed schedules. These features also show how this technology can enhance productivity and facilitate various daily workflows.

The fact that AI voice assistants would often be set up for reminders, to schedule events, and other productivity tasks underlines their integral role in managing both personal and professional life. That, in itself, shows that users are not just passive consumers of smart assistants but actively use them within their organizational routines. Regardless of whether this use was functional, English is still preferred as a first language by 97.6% in comparison with Arabic. The preference for English and the lower reported use of Arabic suggest that respondents perceived Arabic-language interaction as less satisfactory or less convenient. Possible explanations include perceived limitations in dialect recognition, language support, and contextual understanding. However, the survey did not directly test recognition accuracy, response correctness, or task-completion performance; therefore, these findings represent user perceptions rather than objective technical measurements. Which means that, up until now, only 2.4 percent of users use voice assistants in Arabic, reflecting the difficulties these technologies face in understanding and responding correctly to Arabic commands.

Of those who have not used the AI voice assistant, 38.1% would wish to use both English and Arabic, which indicates a demand for better voice assistant support in the Arabic language. These tools have more potential to attract more customers if they can give deeper and subtler knowledge of Arabic. While AI voice assistants are dominated by English, according to the data presented, there is still scope for considerable growth in the Arabic-speaking market if the technology can evolve to satisfy better the linguistic and cultural requirements of such customers. Better Arabic language support might open completely new functional possibilities and therefore strengthen the diffusion process of AI voice assistants within Arabic-speaking communities.

The respondents reported several purposes for using AI voice assistants, including information seeking, communication, entertainment, productivity, shopping, and smart-home control. To avoid overloading the main chapter with detailed response combinations, the complete distribution is presented in Appendix C**.**

4.5 Challenges and Cultural Differences

4.5.1 Technical and Contextual Challenges

The most significant challenge reported is the AI's difficulty in understanding natural language and accents (64.8%). Privacy concerns (28.1%) and limited capabilities (22.4%) are also prominent issues, reflecting broader concerns about AI technology.

4.5.2 Cultural Differences

A substantial 79.5% have experienced Cultural Differences when using AI voice assistants in Arabic. Specific challenges include issues with dialect variations, communication barriers, and gender-related perceptions. Cultural sensitivity and context are crucial areas for improvement to ensure that AI voice assistants can effectively serve diverse Arabic-speaking communities.

Have you noticed any Cultural Differences when using AI voice assistants in Arabic?

Table 4-3: Cultural differences response frequency

		Frequency	%
Valid	No	43	20.5
	Yes	167	79.5
	Total	210	100.0

The use of Voice Assistants introduces a wide array of challenges for Arabic-speaking user experience from both technical and cultural standpoints, thus affecting their effectiveness and levels of adoption. The most frequently reported perceived technical challenge was difficulty understanding natural speech and different Arabic accents, identified by 64.8% of respondents. This indicates that many participants believed accent and dialect recognition to be problematic. However, no speech-recognition system was directly tested, and the percentage should therefore not be interpreted as an objective error rate or a technical measure of AIVA accuracy.

In addition, Arabic is a language of complex structure and with wide dialectal variations, thus posing a serious challenge to the potential of AI for NLP. For instance, some words change in meaning or pronunciation in different regional dialects, which makes it difficult for voice assistants to rightly capture the command and take appropriate action on behalf of the user. This limitation has not only reduced the utility value of AI-powered voice assistants but has also resulted in a loss of users' confidence and satisfaction.

Another critical factor impeding privacy, said 28.1% of the total number of respondents. People are quite suspicious about how their data is tracked and used in voice recognition, as their personal conversations get recorded and then processed. These concerns become heightened with a perceived lack of transparency and control over data privacy settings, which has caused reluctance in using these technologies more frequently or for sensitive tasks. Additional limitations are derived from the limited capabilities of AI voice assistants, as expressed by 22.4% of users. This is mainly because users find it unsatisfactory to experience such assistants performing the range of tasks they would wish, especially in the most complex or specialized interaction.

Among those with a curiosity about culture, a surprising 79.5% reported barriers when using AI Voice Assistants in Arabic. These constitute dialectical barriers, communicational subtleties, and cultural sensitivities. For instance, such is the case with voice assistants not well perceiving regional dialects and further processing; instead, these may result in miscommunication and irritation. Moreover, cultural gender and social norms-where the use of language is gendered or where the forms of address are more or less formal-are barely given consideration in treatment by existing AI models. This will lead to unnatural or inappropriate interactions and reduce the overall user experience due to the lack of cultural adaptation.

Specific cultural challenges which the respondents mentioned include a variety of dialects and colloquial sayings that AI has not been able to handle so far, which is important in such a linguistically diversified region. For example, phases acceptable in one Arabic-speaking country are misunderstood or considered offensive in another.

Furthermore, significant challenges arise around issues of gender roles and linguistic biases, as AI-powered voice assistants may fail in their complete recognition and accommodation of subtle cues, thus causing user discomfort or dissatisfaction. Despite all, limitations in localized content and services were mentioned, a signal that voice assistants are yet to be included in the digital ecosystem on which Arabic-speaking users rely for sourcing information and services.

The results of this survey suggest that voice assistants using AI should not only have their technical capabilities improved but also be more culturally and contextually aware. For example, increased dialect recognition and the development of advanced natural language processing models, fine-tuned to the Arabic language, could lead to significantly improved user satisfaction and engagement. Furthermore, privacy concerns could be alleviated by more straightforward data policies and user controls, addressing some of the reservations presently felt toward full usage. These are gaps that will only be filled in when developers start to embed deeper insights on cultural and linguistic diversities among different Arabic-speaking populations into design and development processes. For example, it could be co-developed with local linguists and cultural experts such that voice assistants are contextually and culturally more attuned. This can go even further by enabling the user to set up their virtual assistant's voice with regards to language preferences and personality traits to best suit cultural expectations in the region.

Respondents identified a broad range of cultural, linguistic, technical, and ethical issues affecting the use of AI voice assistants. The most salient answers related to language/capability of language, accessibility, Arabic dialect/accent recognition, privacy, security, cultural sensitivity, and trust. The cultural-differences responses included language/communication as the most common response category, followed by accessibility and privacy-related concerns. Specific cultural differences were strongly mentioned when participants talked about them, and included privacy and security issues, as well as problems with regional Arabic accents, dialects, pronunciation and meaning.

Other priorities for enhancing Arabic AIVAs were also identified by participants, such as increasing the number and representation of Arabic data sets, more comprehensive coverage of Arabic dialects, user feedback, better government policy, more accessibility for older people, better privacy and security measures and continuing development. Due to the multiple individual formulations in the responses and overlapping categories, the full raw frequency distributions are presented in Appendix D.

4.6 Improvement Areas for AI Voice Assistants

4.6.1 Language and Dialect Recognition

Enhancing the AI's ability to recognize and understand different Arabic dialects is a top priority, as indicated by 64.8% of respondents. Providing accurate and reliable information (16.2%) and better context understanding (7.1%) are also key areas needing development.

4.6.2 Operating Systems and Integration

AI voice assistants are used almost equally across Android (46.7%) and iOS (44.3%) platforms, with some using both (9%). Seamless integration across various devices and operating systems can enhance user experience and accessibility.

How do you think AI voice assistants could be better tailored to meet the needs and preferences of Arabic-speaking users?

Table 4-4: Suggestion by users

Suggestion/User Experience		Frequency	%
Valid	I experience	1	.5
	Void for children	2	1.0
	Accurate Arabic pronunciation and intonation	2	1.0
	Accurate medical translation into Arabic, especially when explaining a medical condition to patients	1	.5
	Accurate Arabic pronunciation & intonation	1	.5
	Age	5	2.4
	Age limitation	5	2.4
	Age limitation to use	1	.5

age limited	1	.5
Age limited	11	5.2
Age limitation	1	.5
Age usage	1	.5
Age limitation to used	1	.5
AI tool awareness	1	.5
Arabic dialect recognition	1	.5
Arabic pronunciation	1	.5
Awareness	4	1.9
Awareness option	1	.5
Becoming as common as English AIVAs	1	.5
Being able to select a number when commands are not understood	1	.5
Better language skills	1	.5
Centralized content	1	.5
Government policy	1	.5
Collaborative tasks and planning	1	.5
Contextual reminders & notifications	1	.5
Continuous learning & adaptation	1	.5
Covered emtire Middle East	1	.5
Cultural recommendations	2	1.0
Cultural sensitivity and adaptation	1	.5
Dialect	1	.5
Dialect-specific Variants	3	1.4
Emotion recognition	2	1.0
Emotional understanding	1	.5
Enabling options of standard & regional Arabic dialects	1	.5
Enhanced accessibility	2	1.0
Enhanced language understanding	1	.5
Enhanced multilingual support	1	.5

Feedback	9	4.3
Feedback	72	34.3
Feedback	1	.5
Government guidelines	1	.5
Have a natural sounding voice	1	.5
integrate voice signature so that AI can build knowledge about this particular user and help him/her better	1	.5
Intelligent contextual understanding	1	.5
IT companies to largely adopt the integration of Arabic in their AIVAs	1	.5
Limitation of age used	1	.5
Localized Recommendations	1	.5
Make the right decisions, application of AI in dangerous situations	1	.5
Mobile phones	1	.5
more human like answers	1	.5
More realistic	1	.5
More Services and option	1	.5
Multi option	1	.5
Multi-dialect Support	1	.5
Multilanguage	1	.5
Multilingual Support	4	1.9
N/A	1	.5
N/A	1	.5
Natural and Interactive Conversations	1	.5
Nothing	1	.5
Nothing	2	1.0
Nothing	4	1.9
Nothing	1	.5
Not in my mind for now	1	.5

Nothing	1	.5
Not in my mind for now	1	.5
Phrase catching	1	.5
Policy and regulation	1	.5
Proactive Assistance	2	1.0
Regulation and policies	1	.5
save simple Arabic dictionary to give good pronunciations of Arabic words	1	.5
Seamless Integration with Local Services	1	.5
Security	1	.5
Security for age	1	.5
Security for Age	1	.5
Services	1	.5
simplicity and efficiency	1	.5
To provide ready-made categories of assistance	1	.5
Translation services (Image to Text) reading text from photo and then translate it.	1	.5
Understanding Arabic voice	1	.5
understanding local culture	1	.5
User Awareness	1	.5
User Feedback	1	.5
User Privacy and Security	3	1.4
User Awareness	1	.5
user feedback	1	.5
User feedback	1	.5
User Feedback	1	.5
Voice Authentication and Personalization	2	1.0
when there is all dialects incorporated	1	.5
Writing an Arabic text through ai voice	1	.5
Yes, high accuracy for voice recognition	1	.5
Total	210	100

4.7 Perceptions of AI Voice Assistants in Arabic

4.7.1 Communication Effectiveness

A high percentage (88.1%) believe that AI voice assistants can effectively communicate with Arabic speakers, despite the challenges mentioned. The importance of AI voice assistants being able to communicate in Arabic is rated as extremely important by 83.3% of respondents.

4.7.2 Importance of Arabic Communication

This high rating underscores the demand for AI voice assistants that are culturally and linguistically competent in Arabic, beyond merely providing basic functionality.

4.8 User Suggestions for AI Voice Assistant Development

4.8.1 Tailoring AI to Arabic-Speaking Users

Suggestions include improving natural language processing for various dialects, enhancing privacy and security features, and incorporating cultural nuances into AI responses. There is also a call for better integration with local services and applications, reflecting the need for AI solutions that are not only linguistically accurate but also contextually relevant.

There was a spread in the level of experience across a variety of IT skills, ranging from software development to information systems, cybersecurity, cloud computing, analytics, networking, teaching, and student-level experience. Many answers, however, included some duplication of terms or included terms that were not consistent such as “student”, “IT student”, “programmer” and “developer”. For this reason, the detailed distributions were relegated to the appendix to reduce the length of these tables. As for technical involvement in the development of AI relevant to KBS, 124 respondents (59.0%) said they had been involved in or heard about the development, and 86 (41.0%) said they had not been involved in or heard about the development. The scores of AI developers and KBS functions were received from a minor part of the respondents as not applicable. Thus, care should be taken when interpreting these results. The full results of respondent's specialism, technical knowledge and attitudes towards the role of AI developers and KBS functions are included in Appendix E.

4.9 Summary and Recommendations

The data reveals that while there is a strong uptake and interest in AI voice assistants among Arabic speakers, significant challenges remain in areas like language recognition, cultural sensitivity, and privacy. To enhance the functionality and user experience of AI voice assistants for Arabic-speaking users, developers should focus on:

1. Improving Natural Language Processing:
   Developing robust models that can handle various Arabic dialects and contextual nuances.
2. Enhancing Cultural Awareness:
   Integrating cultural and social norms into AI interactions to ensure responses are appropriate and respectful.
3. Addressing Privacy and Security Concerns:
   Implementing strong data protection measures and transparent privacy policies to build user trust.
4. Expanding Integration and Accessibility:
   Ensuring seamless use across different devices and platforms, with particular attention to accessibility for older users and those with disabilities. By addressing these areas, AI voice assistants can better serve the needs of Arabic-speaking communities, fostering greater adoption and satisfaction.

Table 4-5: Gender AIVA_Usage_ Arabic Crosstabulation

			AIVA_Usage_ Arabic		**Total**
			No	**Yes**	
Gender	Female	Count	39	53	92
		Gender %	42.4%	57.6%	100%
		AIVA_Usage_ Arabic %	45.9%	42.4%	43.8%
		Total %	18.6%	25.2%	43.8%

	Male	Count	46	72	118
		Gender %	39.0%	61.0%	100%
		AIVA_Usage_ Arabic %	54.1%	57.6%	56.2%
		Total %	21.9%	34.3%	56.2%
Total		Count	85	125	210
		Gender %	40.5%	59.5%	100%
		AIVA_Usage_ Arabic %	100.0%	100.0%	100%
		Total %	***40.5%***	***59.5%***	***100%***

Table 4-6 shows the association between gender and using AI voice assistants in Arabic. The results suggest that there was not a significant difference in usage between the two genders, with the proportion of male respondents reporting the use of Arabic AI voice assistants a bit higher than that of female respondents. On the whole, 59.5% of the respondents admitted using AI voice assistants in Arabic, 40.5% admitted not to.

There is a small difference in the adoption patterns based on gender. Even though the percentage of the male respondents proved to be slightly higher, the gap between male and female respondents does not seem to be very large, meaning that the use of AI voice assistants is comparatively common among both groups. These results suggest that gender might not be a significant obstacle to the use of Arabic AI voice assistants in the sample.

The identified difference can be linked to the differences in the level of exposure to technology, its availability, or perceived utility of voice-based technologies. Nevertheless, the differences in description cannot be used to ascertain whether there is a significant effect of gender in usage behaviour or not. Thus, the chi-square test was used to specify whether there is a statistically significant correlation between gender and the use of Arabic AI voice assistants.

Table 4-6: Chi-square test for gender * AI_VA_Usage_Arabic

Chi-Square Tests

	Value	DF	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.249[a]	1	.618	-	-
Continuity Correction[b]	.128	1	.721	-	-
Likelihood Ratio	.249	1	.618	-	-
Fisher's Exact Test	-	-	-	.672	.360
Linear-by-Linear Association	.248	1	.618	-	-
No of Valid Cases	210	-	-	-	-

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 37.24.

b. Computed only for a 2x2 table

The Pearson Chi-Square test evaluates whether the observed differences in AI voice assistant usage between males and females are statistically significant. The Chi-Square value is 0.249, with a p-value of 0.618. Since the p-value is greater than the standard threshold of 0.05, researcher fail to reject the null hypothesis, meaning that the difference in AI voice assistant usage between genders is not statistically significant. This suggests that, although there is a slightly higher percentage of males using AI voice assistants, the difference is too small to be considered meaningful. The data does not provide sufficient evidence to suggest that gender plays a significant role in determining whether someone uses AI voice assistants in Arabic.

By applying the Continuity Correction, which aims at correcting small sample sizes, the Chi-Square corrects to 0.128, with a p-value of 0.721. This value also confirms that there is an insignificant association between gender and using Arabic AI voice assistants. Similarly, the Likelihood Ratio test outputs the value of 0.249 with a p-value of 0.618, which also supports the fact that no significant difference in gender described the use of an AI voice assistant in Arabic.

Fisher's Exact Test, which is of particular help when sample sizes are small, gives a two-sided significance of 0.672 and one-sided of 0.360. These results further confirm the finding that there is no statistically significant relationship between gender and the use of an Arabic AI voice assistant. The Linear-by-Linear Association test, identifying if there is a linear trend between the two variables, also yields a Chi-Square value of 0.248 with a p-value of 0.618. This indicates no linear association between AI voice assistant usage in Arabic.

4.10 Conclusion for Hypothesis

Hypothesis 1: There is a Significant Difference in the Usage of AI Voice Assistants (in Arabic) Based on Gender.

This hypothesis is *not supported* as the p-value indicates that the difference in usage between genders is not statistically significant

Table 4-7: Independent samples test for AI_Important and AI_Technical_KBS

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
AI_Important	Equal variances assumed	8.199	.005	-1.445	208	.150	-.146	.101	-.346	.053
	Equal variances not assumed			-1.492	200.77	.137	-.146	.098	-.340	.047

The independent samples t-test compares the mean perceptions of AI importance between two groups: those with technical knowledge and those without. The test produced a t-value of -1.445 and a p-value of 0.150, which is greater than 0.05, indicating that the difference in AI importance based on technical knowledge is not statistically significant. The mean difference between the two groups (-0.146) is relatively small, and the confidence interval also suggests that the difference is not meaningful. Therefore, the perception of AI importance does not significantly vary between individuals with and without technical knowledge. In this case, technical knowledge does not appear to influence how important participants perceive AI to be in their lives. The independent samples t-test carried out in Table 4-8 was to ascertain any significant difference in the perceived importance of AI between those who had technical knowledge in KBS and those who did not. It assumes equal variances since the Levene's Test for Equality of Variances has given an F-value of 8.199 and a p-value of 0.005. This low p-value suggests that the equal variances assumption is violated. Therefore, the result findings for t-test not assuming equal variances.

The independent samples t-test output taking the assumption of equal variances reports a t-value of -1.445, df of 208, and a two-tailed significance of 0.150; whereas for unequal variances, the t-values is -1.492 and df of 200.771 with a two-tailed probability of 0.137. Both of these display a p-value greater than the traditional threshold of 0.05. Thus, this leads to a decision that the difference in perceived importance of AI between the groups is not statistically significant. The estimated mean in differences of importance of AI between those with and those without technical knowledge is -0.146, and 95% confidence interval of this estimate is from -0.346 up to 0.053. Since the width of a confidence interval includes zero, this further suggests that this observed difference is not meaningful and thus probably due to chance.

These results imply that, for KBS, technical knowledge does not make any significant difference in the importance people attach to AI. Independent of their technical knowledge, both groups report similar importance for AI. Such a finding might mean that general AI awareness aspects, cultural attitudes, or personal experience are more critical in the perception of the relevance and impact of AI than technical knowledge. Accordingly, the undertaking of increasing the relevance perceived of AI cannot be limited to technical pleasure and enhancements but must be envisioned as one involving broader educative and communicative approaches toward varied audiences.

Hypothesis 2: Perception of AI Importance Differs Significantly Based on the Technical Knowledge of Respondents

This hypothesis is *not supported* as the results show no statistically significant difference in AI importance based on technical knowledge.

Table 4-8: Logistic regression analysis for AI_VA_Usage_Arabic

Classification Table[a,b]					
	Observed		Predicted		
			AIVA_Usage_ Arabic		Percentage Correct
			No	Yes	
Step 0	AIVA_Usage_ Arabic	No	0	85	0
		Yes	0	125	100
	Overall Percentage				59.5

a. Constant is included in the model

b. The cut value is .500

The logistic regression analysis predicts the likelihood of using AI voice assistants in Arabic based on factors like Cultural Differences, gender, and technical knowledge. The results indicate that **Cultural Differences** (p-value = 0.000, Exp(B) = 9.806) significantly affect AI voice assistant usage. Specifically, as Cultural Differences increase, individuals are less likely to use AI voice assistants. This is evidenced by the high odds ratio (Exp(B)), suggesting that Cultural Differences play a strong role in determining whether individuals adopt AI technology. On the other hand, **gender** does not significantly influence AI usage (p-value = 0.391), while **technical knowledge** shows a significant effect (p-value = 0.101), with those possessing technical knowledge being less likely to encounter challenges in using AI voice assistants. In summary, Cultural Differences and technical knowledge significantly impact AI usage, but gender does not.

Hypothesis 3: Cultural Differences Significantly Affect the Usage of AI Voice Assistants in Arabic.

This hypothesis is *proven true* as the logistic regression analysis shows that Cultural Differences significantly impact AI voice assistant usage.

Table 4-9: ANOVA for age groups and AI importance

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
			Std. Error	Beta		
1	(Constant)	1.595	.427		3.733	.000
	AI_Technical_KBS	.154	.107	.105	1.437	.152
	AI_Cultural_Differences_Arabic	-.288	.128	-.161	-2.250	.026
	Gender	-.252	.096	-.173	-2.627	.009
	Age	.134	.046	.205	2.916	.004

Dependent Variable: AI Important

The ANOVA test examines whether the perceived importance of AI varies significantly across different age groups. The test produced an F-value of 2.334 and a p-value of 0.057. Accordingly, post-hoc comparisons by Tukey HSD test are additional tests conducted after ANOVA to determine exactly where the differences among the groups exist. Researcher used the specific Tukey HSD test here-after ANOVA indicated there might be some variation by age in how important these workers think AI will be-to compare each age group against every other age group. Because it controls for the potential of finding false significance when making multiple comparisons, this analysis was performed. The Tukey HSD test showed that no age group was individually significantly different from another in the perception of AI importance, which suggested that although variation across groups may exist, it should not be strong enough to be statistically significant.

Hypothesis 4: There is a Significant Difference in the Perception of AI Importance Across Different Age Groups.

This hypothesis is *not strongly supported*, as the ANOVA results indicate only a marginal, but not statistically significant, difference in AI importance across age groups.

Table 4-10: ANOVA and correlation analysis

ANOVA					
AI_Important					
	Sum of Squares	**DF**	**Mean Square**	**F**	**Sig.**
Between Groups	4.775	4	1.194	2.334	.057
Within Groups	104.849	205	.511	-	-
Total	109.624	209	-	-	-

Table 4-11: Correlation between AI importance and AI developers

Correlations			
		Important	**AI Developer**
AI Important	Pearson Correlation	1	.359
	Sig. (2-tailed)		.000
	N	210	210
AI Developers	Pearson Correlation	.359	1
	Sig. (2-tailed)	.000	
	N	210	210

Table 4-12: Variable between AI importance and AI developers

Variable	**Pearson Correlation**	**Sig. (2-tailed)**	**Sample Size N**	**Interpretation**
AI Important	1	-	210	Correlation is moderate (0.359) between the perceived importance of AI & AI developers' perceptions. This result is statistically significant (p = 0.000), indicating that higher AI

				developers' perception correlates with a higher perceived importance of AI.
AI Developer	0.359	0.000	210	

Table 4-13: ANOVA - AI importance across groups

ANOVA (AI Importance across Groups)	**AI Important**
Sum of Squares (between groups)	4.775
Mean Square (between groups)	1.194
df (between groups)	4
Sum of Squares (within groups)	104.849
Mean Square (within groups)	0.511
df (within groups)	205
F-Ratio (variance ratio)	2.334
Sig. (p-value)	0.057
Interpretation	F-ratio = 2.334 suggests a variance between groups, p-value = 0.057 indicates that the difference in perceived AI importance across groups is not statistically significant at the 0.05 threshold. However, it is close, meaning there could be a trend worth exploring further.

4.11 ANOVA Analysis for AI Importance

Table 4.25 reports the output of an ANOVA conducted in order to see whether, in fact, there are any statistically significant differences in the perceived importance of AI by different groups. ANOVA is a statistical test typically used when comparing means between multiple groups to see whether any group is significantly different from the others. The ANOVA analysis is important for this study because it enable the researcher to determine whether perceptions of the importance of AI differ significantly across demographic groups. Understanding whether and how these perceptions differ is key to tailoring AI applications, especially in Arabic-speaking communities where factors such as culture and technological familiarity could influence attitudes toward AI. Although the result is not statistically significant, it gives a basis for further exploration and refinement of grouping variables or sample sizes to detect possible trends.

In this analysis, the comparison is made on how different demographic groups view the significance of AI, which is essential in the understanding of any differences in AI adoption and perception. The sum of squares between groups is 4.775, with a mean square of 1.194 and df = 4. This component of the analysis represents the variance between the groups, measuring how far apart the group means are from each other. In contrast, sum of squares within groups is 104.849, mean square 0.511, and df = 205, representing the variations within each group, the variation between participants even those in the same group about AI importance. This variance between groups and in relation to the variance within groups would be computed in order to derive the F-ratio 2.334. The higher the F-ratio, the higher is the chance that the difference is statistically significant rather than due to chance. The p-value for the F-ratio is 0.057 above the conventional threshold of 0.05 that signified significance at a confidence level of 95%. This finding indicates that the groups differ concerning their perceptions of how important AI is to them, and this difference at conventional levels is not statistically significant.

However, with a p-value of 0.057, which is reasonably close to 0.05, there is a potential for meaningful patterns or trends that deserve exploration. Analysis shows that while it cannot be concluded to prove significant differences in the perception of the importance of AI, it is not a definite result that can exclude the presence of group differences. Different factors, such as differences in exposure to AI technologies, technological literacy, or cultural attitudes toward AI, might explain these variations and merit further investigation in future research.

4.12 Correlation of Key Variables: Importance of AI and AI Developers' Credibility

Pearson correlation analysis was conducted to analyse the association between two key variables: the perceived importance of AI and trust in AI developers. The results show that people are more likely to consider AI technologies important if they trust the developers of AI. This conclusion underlines the reputation and credibility of AI developers in forming public attitudes toward AI. Through this understanding, organizations and developers can be able to find determinants of public trust that can promote technology adoption. The Pearson correlation coefficient is 0.359, which is moderate positive. This shows that when trust in AI developers is high, the importance of AI technologies is also perceived as high. The p-value is highly significant at 0.000. This p-value is well below the conventional threshold of 0.001, which further strengthens the conclusion that there must be a real correlation because what has been seen here is not due to chance. Therefore, how individuals view AI developers as such can meaningfully relate their importance for AI.

This positive and significant correlation highlights the role of trust and credibility in shaping public attitudes toward AI. Respondents who have a favorable perception of AI developers are more likely to recognize AI as a valuable and impactful technology. This finding underscores the influence that the reputation and expertise of AI developers have on the public's acceptance of AI, suggesting that as the credibility of AI developers improves, the perceived importance and adoption of AI technologies across industries may also increase. The analysis hence confirms Hypothesis 4: a positive and significant relation between the perception of AI developers' activities and the perceived importance of AI. Indeed, a correlation coefficient of 0.359, associated with a significant p-value, confirms this hypothesis.

In other words, it indicates that the people who perceive the more positive role of the developers of AI are also more likely to view AI as important in their life. This may be due to the heightened confidence in the technology, belief in the reliability of AI products, and belief that developers can surmount ethical and technical challenges associated with AI. Conclusively, findings suggest that efforts to improve public perception about AI developers through transparency, ethical practices, and effective communication could boost perceptions of the importance and acceptance of AI technologies. Certainly, this is fundamentally useful in increasing people's involvement with AI, a positive attitude, and much more acceptance and integration in life and professional context.

Table 4-14: Round 1 Exploratory Hypothesis Testing

Hypothesis	**Proposed relationship**	**Test used**	**Statistical result**	**Decision**	**Meaning**
H1	Gender is significantly associated with Arabic AIVA usage	Chi-square test	Insert χ^2, df and exact p-value from Table 4-18	Supported/Not supported	Indicates whether reported Arabic AIVA usage differed by gender
H2	Perceived AI importance differs according to technical knowledge	Independent-samples t-test	($p > .05$)	Not supported	Technical knowledge did not produce a statistically significant difference in perceived AI importance

H3	Cultural differences significantly affect Arabic AIVA usage	Binary logistic regression	(p < .001), Exp(B) = 9.806	Supported	Reported cultural differences were significantly associated with the likelihood of Arabic AIVA usage
H4	Perceived AI importance differs across age groups	One-way ANOVA	(F(4,205)=2.334,\ p=.057)	Not supported	Perceived AI importance did not differ significantly across age groups
Exploratory association*	Perceptions of AI developers are associated with perceived AI importance	Pearson correlation	(r=.359,\ p<.001,\ N=210)	Significant association	More favourable perceptions of AI developers were associated with greater perceived

	AI importance

Table 4-13 provides a concise summary of the exploratory hypotheses tested with the Round 1 dataset in a systematical manner. These analyses are exploratory; therefore, the hypothesis is supported if there is a statistically significant association within the sample of Round 1, but does not indicate the final structural model.

Overall, the results of Round 1 were exploratory in nature and served to document the use of AIVAs, perceived benefits and challenges, importance of Arabic-language and culturally appropriate communication, as well as preferences for system improvement areas. The results of these findings were used to help determine the constructs to be examined more deeply during Round 2. The statistical tests in the first round, however, were exploratory and not used as evidence for the confirmation of the final structural model.

4.13 Round 2: Confirmatory SEM Results

The results of the second round survey of 459 respondents are presented here in support of the results. The measurement and structural models were tested in Round 2 which is different from Round 1 which explored the pattern of usage and users perception. In this section, all results on reliability, validity, model-fit, path-coefficient, explained-variance, and hypothesis-testing are based only on the Round 2 data set. The results are the final empirically-tested model of the study.

The second-round survey of the research with sample size N = 459 provided the empirical foundation for developing the theoretical framework of AI voice assistant adoption in Arabic-speaking environments. In contrast to the exploratory nature of Round 1, Round 2 tested the hypothesized propositions explicitly using SPSS descriptive statistics and AMOS structural equation modelling (SEM) to confirm the inter-relationship between important constructs. This method makes sure that the framework is not only theoretical but also statistically based.

The measurement assessed four determinant factors; perceived challenges, perceived benefits, design preferences, and communication in government, and their impact on the perceived importance of AI services. SPSS analysis indicated moderately positive mean perceptions across constructs (3.49–3.54), whereas AMOS structural model findings showed substantial contributions of all four factors to the perceived importance of AI services, with standardized path coefficients between $\beta = 0.204$ and $\beta = 0.312$.

The structural model produced an R^2value of 0.618 for Perceived Importance of AI Services. Together, these constructs accounted for 61.8% of the variance in the perceived importance of AI services in the Round 2 sample population. The rest (38.2%) was due to factors not accounted for in the model and unexplained variation. It is important to note that the R^2value does not reflect the predictive validity of the model in terms of real AIVA performance, the real rate of continued usage, or organisational outcomes.

Moreover, the model demonstrated very strong overall fit, with RMSEA = 0.020, SRMR = 0.027, CFI = 0.993, and TLI = 0.992. RMSEA evaluation the approximate fit error of the model, while SRMR evaluation the average error of the observed correlation and the modeled correlation; smaller values indicate better fit. Values of CFI and TLI will be closer to 1 when the model fits well with the baseline model, which assumes that the variables are not related. Combined, these indices suggest that the proposed covariance structure was very similar to the Round 2 data. Strong fit does not mean the model is necessarily the right one, or that the relationships are causal, however.

The SEM analysis was conducted in two steps which were closely connected. The first was the measurement model which determined the reliability and distinctiveness of the observed questionnaire measures for the intended latent constructs. At this stage, factor loadings, internal consistency, Composite Reliability, Average Variance Extracted, discriminant validity and overall model fit were examined. Second, the structural model tested the hypothesised relationships between Effective Usage, Perceived Challenges, Perceived Benefits, Design Preferences, Communication in Government and Perceived Importance of AI Services. The measurement model thus measured the quality of the constructs while the direction, strength and the statistical significance of the proposed relationships between these constructs were measured by the structural model.

Demographic analysis also indicated that the participants were evenly spread by gender, age, and education level and comprised primarily educated professionals. Nonetheless, response diversity pointed to persistent hindrances including Arabic accent identification (64.8%), cultural localization requirements (79.5%), and privacy issues. These results underscored the importance of adapting AI technologies to linguistic and cultural standards of Arabic-speaking users.

4.13.1 Conceptual Framework and Structural Modelling - AMOS Analysis

The measurement model was assessed before interpreting the structural paths. The retained indicators demonstrated acceptable standardised loadings, internal consistency, Composite Reliability, and Average Variance Extracted. Discriminant validity was also examined to determine whether the constructs were empirically distinct. Once the adequacy of the measurement model had been established, the structural model was evaluated by examining the standardised path coefficients, critical ratios, significance levels, and the proportion of variance explained in the dependent construct.

Based on Round 2 results, the proposed framework was empirically tested using SEM within the Round 2 sample in AMOS. Path analysis tested the relationships between constructs, with the following key hypotheses:

- **H1a**: Effective usage → Perceived challenges
- **H1b**: Effective usage → Perceived benefits
- **H1c**: Effective usage → Design preferences
- **H1d**: Effective usage → Communication in government
- **H2–H5**: Each determinant (challenges, benefits, design preferences, communication) → Perceived importance of AI services

The AMOS results indicated that all specified paths were statistically significant and were supported within the Round 2 sample. Interestingly, communication in government ($\beta = 0.312$) and perceived benefits ($\beta = 0.278$) were the strongest predictors of perceived importance, followed by design preferences ($\beta = 0.204$) and perceived challenges (β = 0.216) with positive but relatively moderate effects. Mediation analysis revealed indirect effects, for example, effective use → perceived advantages → perceived significance, highlighting the mediating influence of user experience on positive adoption results. These results indicate that respondents considered linguistic adaptation, cultural suitability, and privacy-related features important when evaluating AI voice services. Table 4.16 presents the summary of the test results of the hypotheses.

Figure 4-2: Amos model (conceptual framework)

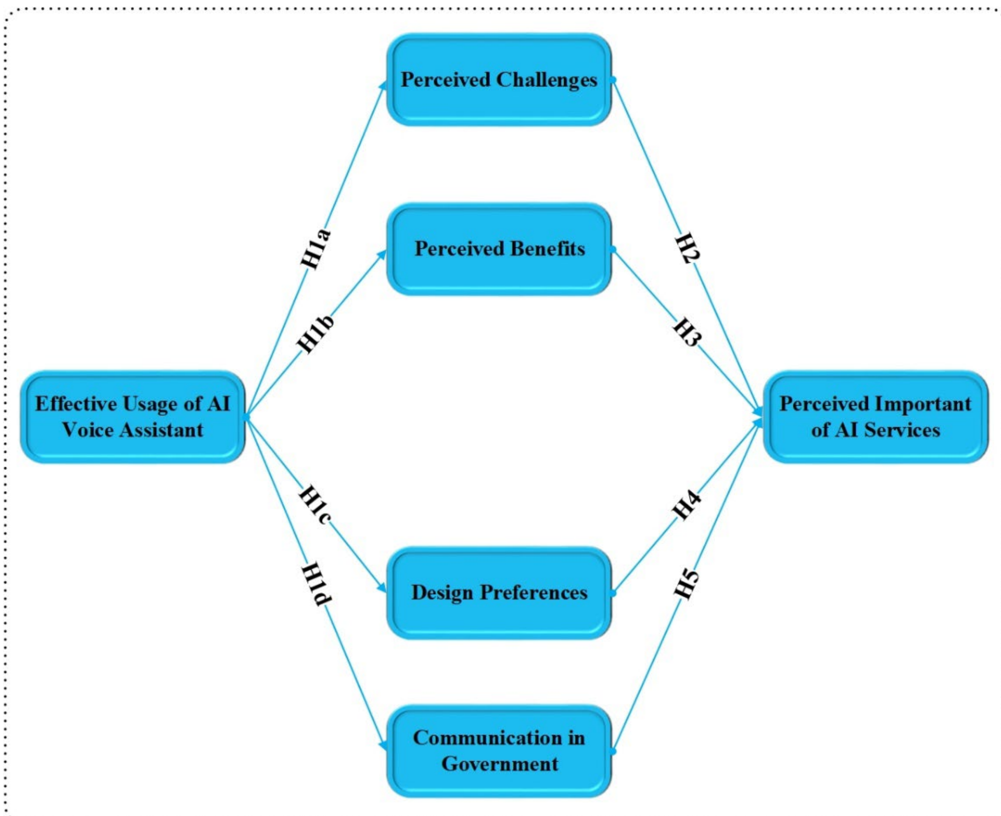


Table 4-15: Hypotheses Testing Results (Round 2 – AMOS SEM)

	Path Relationship	Standardized β	Result
H1a	Effective Usage → Perceived Challenges	0.216	Supported
H1b	Effective Usage → Perceived Benefits	0.278	Supported
H1c	Effective Usage → Design Preferences	0.204	Supported

H1d	Effective Usage → Communication in Government	0.312	Supported
H2	Perceived Challenges → Perceived Importance of AI Services	0.216	Supported
H3	Perceived Benefits → Perceived Importance of AI Services	0.278	Supported
H4	Design Preferences → Perceived Importance of AI Services	0.204	Supported
H5	Communication in Government → Perceived Importance of AI Services	0.312	Supported

Model Fit Indices: RMSEA = 0.020; CFI = 0.993; R^2 = 42.8% – 50.8% across constructs; Total explained variance in perceived importance = 61.8%.

The confirmed framework illustrates that effective application of AI voice assistants has a positive impact on user attitudes in the dimensions of challenges, advantages, design, and governmental communication. These determinants are able to significantly forecast perceived significance of AI services, affirming the central role of cultural fit, dialectal support, and privacy in determining adoption among Arabic-speaking population. This empirically validated framework not only advances theoretical knowledge but also offers practical recommendations for developers and policymakers looking to improve AI integration in the region.

Table 4-16: Round 2 SEM Hypothesis Testing

Hypothesis	**Structural relationship**	**Test used**	**Standardised β**	**Critical ratio**	**Significance**	**Decision**	**Meaning**
H1a	Effective Usage → Perceived Challenges	SEM path analysis	0.671	19.171	(p<.001)	Supported	Greater effective usage was associated with

							stronger recognition of perceived challenges
H1b	Effective Usage → Perceived Benefits	SEM path analysis	0.713	23.000	(p<.001)	Supported	Greater effective usage was associated with stronger perceptions of AIVA benefits
H1c	Effective Usage → Design Preferences	SEM path analysis	0.654	17.211	(p<.001)	Supported	Effective usage was positively associated with clearer design preferences
H1d	Effective Usage → Communication in Government	SEM path analysis	0.689	19.686	(p<.001)	Supported	Effective usage was positively associated with perceptions of governme

							nt communication
H2	Perceived Challenges → Perceived Importance of AI Services	SEM path analysis	0.216	3.857	(p<.001)	Supported	Perceived challenges were positively associated with the perceived importance of AI services
H3	Perceived Benefits → Perceived Importance of AI Services	SEM path analysis	0.278	5.673	(p<.001)	Supported	Perceived benefits increased the perceived importance assigned to AI services
H4	Design Preferences → Perceived Importance of AI Services	SEM path analysis	0.204	4.250	(p<.001)	Supported	Preferred design features were positively associated with perceived

							AI-service importance
H5	Communication in Government → Perceived Importance of AI Services	SEM path analysis	0.312	5.887	(p<.001)	Supported	Government communication was the strongest direct predictor of perceived AI-service importance

The Round 2 SEM results constitute the final tested model of the thesis. The model examined the relationships between Effective Usage, Perceived Challenges, Perceived Benefits, Design Preferences, Communication in Government, and Perceived Importance of AI Services. The statistically significant paths and explained variance reported in this section are therefore confirmatory findings based solely on the Round 2 sample and should not be treated as a direct statistical comparison with the exploratory Round 1 results.

4.14 Chapter Summary

This chapter reported findings from two survey rounds examining the adoption of AI voice assistants among Arabic-speaking users. The demographic analysis showed that respondents were predominantly middle-aged, highly educated, and employed in education, IT, and government. Usage patterns revealed widespread exposure but lower sustained use in Arabic, reflecting challenges of dialectal diversity. Entertainment, shopping, and productivity tasks dominated usage, with English overwhelmingly preferred. Key obstacles included accent recognition, privacy concerns, and cultural sensitivities. Statistical tests confirmed that gender did not significantly influence adoption, while cultural differences and technical knowledge shaped perceptions. The AMOS-based framework mapped these interrelations, underscoring priorities for future development.

5. CHAPTER 5: DISCUSSION

5.1 Discussion overview

This chapter will make sense of the results in Chapter 4 with respect to the research questions, the literature available and the modified conceptual framework of the study. Instead of repeating the statistical outcome, the discussion is based on the meaning of the results to the Arabic-speaking users, the design of the AI voice-assistance, the organisation communication, and knowledge-based systems. The chapter summarises the key findings by research question first and then talks of the exploratory findings of Round 1 and the confirmatory findings of Round 2 SEM.

5.1.1 Summary of Main Findings by Research Question

This study examined the application of AI voice assistants in Arabic speaking users in the context of knowledge-based systems in four research questions. Compared to Q1, the results revealed that initial exposure to AIVAs was high but the sustained use was lower in Arabic, which means that the availability of voice-assistant technology did not always convert into daily use of the Arabic language. The respondents mainly AIVAs were used to seek information, communicate, be productive, entertain and perform other task-specific functions usually using mobile devices.

Compared to Q2, the respondents identified practical advantages and major challenges. The benefits perceived were convenience, accessibility, time saving, service support, but perceived challenges were dialect and accent comprehension, privacy, security, contextual interpretation, and system integration. Technical knowledge did not always generate meaningful differentiation in perceived AI importance in Round 1 tests of Exploration, whereas Round 2 SEM demonstrated that both Perceived Benefits and Perceived Challenges were significantly linked to Perceived Importance of AI Services.

Compared to Q3, the communication between the company and customers was not efficient based on the existence of an Arabic language option. The respondents stressed the need to support dialects, engage in bilingual interaction, communicate in a culturally appropriate way, understand contextually, privacy, trust, and clear communication of the government. In the Round 2 structural equation, Communication in Government turned out to be the best predictor of Perceived Importance.

Compared to Q4, respondents favoured AIVAs that accommodate regional dialects, culturally suitable language, greater privacy, easily accessible interaction, personalised environment, trustworthy local services, and connection with authoritative organisational knowledge. The findings have informed the given design requirements and conceptual blueprint of AIVA-KBS integration. On the whole, the findings address the research question by revealing that the perceived relevance and sustained relevance of AIVAs in Arabic-speaking settings are related to the interplay between practical utility, linguistic inclusion, cultural appropriateness, institutional communication, and trust.

5.1.2 Discussion of Findings in Relation to Existing Literature

The results on the aspects of daily integration and use indicate that the exposure to the AI voice assistants was quite common, yet the continued use of the Arabic language was less prevalent. This is widely consistent with Al Shamsi (2021), who discovered that perceived usefulness, enjoyment, and facilitating conditions affected AIVA adoption in UAE university students, and Salih (2025), who found trust and quality of service to be significant to continued usage in Oman. The current study however expands on these studies by establishing a more definite gap between the first trial and the sustained Arabic interaction in a wider sample of the UAE. This implies that mere availability and familiarity is not a guarantee of continued use. The traditional adoption models might thus not fully appreciate the significance of dialect recognition, code-switching, cultural suitability, and trust in Arabic-language responses. The result criticizes the belief that universal presence of commercial assistants will necessarily lead to significant adoption of such a system among linguistically diverse groups of people.

The high use of mobile-based and task-oriented usage is also aligned with the earlier studies outlining AIVAs as convenient devices to retrieve information, communicate, remind, entertain, become more accessible, and control their smart-based devices. Patil et al. (2024) focused more on multilingual support, personalisation, and contextual awareness, and Olawade et al. (2025) focused more on the benefits of voice interaction to people who need to control their daily life or physical constraints. In the same manner, Benjamin (2025) wrote about the access of information in real time and organisational integration. However, the existing results indicate that the activities were mostly extraneous and were not so entrenched in the daily activities of respondents. This contributes UAE-specifically to the literature and refutes the perception that greater commercial diffusion is a natural result of enduring and widespread everyday practice among Arabic-speaking users.

The results concerning the perceived advantages and difficulties also indicate a significant concurrence with the previous researches. Respondents linked AIVAs to convenience, accessibility, time savings, accessing information, and service support. These impressions are in line with Abdelraouf (2024), who associated AI-aided automation with efficiency and fewer mistakes, as well as Elmedany (2024) and Al Ali and Khalil (2025), who emphasized quicker interaction during the services and better communication in the governmental settings. Simultaneously, the respondents were worried about dialect recognition, contextual understanding, and privacy, security, cultural sensitivity, and integration. These issues are consistent with Bouzid (2023), AlAfnan (2025), El Zahraa (2025), and Besdouri et al. (2024) whose works highlighted the complexity of the Arabic language, mixed-language interaction, the lack of dialect coverage, and multilingual processing challenges. The current research contributes to the literature in that it considers these benefits and concerns as a single framework instead of dividing them into individual technical or service-related problems.

Specifically, the fact that technical knowledge did not always generate meaningful differences in perceived AI importance is worth noting. The common assumption of technology-adoption theories is that familiarity, competence and facilitating conditions determine acceptance. Nevertheless, the current results indicate that language, culture, privacy, trust, and perceived usefulness can be the issue of both technologically more advanced and less advanced users. This turns around the belief that technical expertise is chiefly influencing attitudes towards AIVAs. It also goes beyond traditional views on adoption by implying that digital competence is not enough to explain Arabic AIVA acceptance. Rather, the significance of linguistic inclusion and institutional trust might be similar in various groups of users.

Dialect support and culturally appropriate communication greatly concur with Arabic NLP and human-computer interaction studies. Research that makes use of MADAR, NADI, MGB and other datasets of Arabic speech has demonstrated that variation in dialect, code-switching and regional variations in pronunciation make recognition and language interpretation difficult. The inadvisability of the treatment of Arabic as a homogenous linguistic system was also stressed by Ali et al. (2017), Bouamor et al. (2019), Khalilia et al. (2024), AlAfnan (2025), and El Zahraa (2025). Nevertheless, a lot of this literature is more technical in nature and deals with the error rates, datasets or benchmark performance. The present research introduces a user-centred viewpoint by demonstrating that respondents attributed linguistic mismatch to comfort, trust, effective communication and desire to use AIVAs further. Thus, the results can be used to generalize the technical research to determine the perceived social and practical implications of the dialect and code-switching challenges.

Privacy and trust also came out as some of the main concerns. It concurs with the UNESCO Recommendation on the Ethics of AI, NIST AI Risk Management Framework and previous research highlighting transparency, accountability, protection of data and continuous monitoring of risks. Mitchell et al. (2019) demonstrated the importance of documentation in terms of Model Cards and Datasheets, whereas Amershi et al. (2019) mentioned transparency, confirmation, and successful recovery in the case of the interaction errors. The current results provide some contextual information as they reveal that Arabic-speaking interviewees have not viewed privacy in isolation. Rather, the concept of privacy was linked with cultural responsiveness, the appropriateness of dialect, respectful communication and trust in institutional information. Confidence in Arabic AIVAs thus seems to be multidimensional and not a technical one.

This close connection between Communication in Government and Perceived Importance is also in line with regional findings related to the voice systems in the public-sector. The example of Rammas and the Bahrain Ali assistant shows the possibility of voice-based public services when the system is integrated with the organisational processes and trusted sources of knowledge. The study by Elmedany (2024), Rahman et al. (2024), and Al Ali and Khalil (2025) is not an exception, as the authors indicate that faster response and engagement with the population is possible with the help of automated communication. This literature is further developed by the current study which demonstrates that the perceptions of AIVA importance among respondents were more related to the perceptions of government communication compared to the other direct predictors in the model tested. It means that credibility (institutional) and communication can play a particularly significant role in the UAE context, but this association cannot be taken as an evidence of service improvement.

Lastly, the design preferences of the respondents are in line with suggestions in dialect-balanced datasets, code-switching features, culturally sensitive address, privacy-by-design, accessibility, personalisation and transparent information sources. Amershi et al. (2019), Mitchell et al. (2019), UNESCO (2021), and NIST (2023) have general interaction, documentation, and governance principles. The current work contributes to a UAE-based prioritisation of these needs by demonstrating how the Arabic-speaking users themselves rated dialect support, cultural appropriateness, privacy options, accessibility, integration of local services, and legitimate information. This enhances the literature on general ethical and technical advice to an empirically based collection of user needs of Arabic AIVAs, and their future interrelationship with knowledge-based systems.

5.2 Demographics of Respondents

Demographic profile gives valuable background information in interpreting the results instead of being an answer to the research questions. The comparatively educated and digitally proficient make-up of the sample might have made them more familiar with AIVAs and aware of their possible advantages and downsides. Simultaneously, a small sample size of the respondents, especially regarding the age, education, and employment factors, restricts the possibility to generalise the results to the entire Arabic-speaking users in the UAE, especially those with lower digital literacy or less access to voice-assistant technologies.

5.3 Common Method Bias

This research also used the common method bias with the help of Harman's single factor technique. The variance extracted using one factor is 9. 909%, less than 50%, suggesting that there is no common method bias in this study (Podsakoff et al., 2003).

5.3.1 Descriptive Statistics

Table 5-1: Descriptive Statistics

Construct Items	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
EUAIVA_1	459	1	5	3.490	1.143	-0.522	-0.503
EUAIVA_2	459	1	5	3.484	1.145	-0.577	-0.405
EUAIVA_3	459	1	5	3.529	1.145	-0.533	-0.471
EUAIVA_4	459	1	5	3.534	1.096	-0.626	-0.166
PC_1	459	1	5	3.453	1.159	-0.557	-0.433
PC_2	459	1	5	3.527	1.135	-0.643	-0.284
PC_3	459	1	5	3.503	1.070	-0.518	-0.307
PC_4	459	1	5	3.462	1.118	-0.616	-0.324
PB_1	459	1	5	3.475	1.122	-0.613	-0.219
PB_2	459	1	5	3.547	1.113	-0.666	-0.240
PB_3	459	1	5	3.523	1.120	-0.657	-0.130
DP_1	459	1	5	3.545	1.187	-0.625	-0.461
DP_2	459	1	5	3.525	1.139	-0.596	-0.348
DP_3	459	1	5	3.501	1.145	-0.480	-0.462
DP_4	459	1	5	3.521	1.104	-0.649	-0.208
CIG_1	459	1	5	3.547	1.113	-0.647	-0.174
CIG_2	459	1	5	3.516	1.074	-0.483	-0.348

CIG_3	459	1	5	3.580	1.182	-0.637	-0.421
CIG_4	459	1	5	3.523	1.124	-0.595	-0.294
PIAIS_1	459	1	5	3.490	1.095	-0.622	-0.217
PIAIS_2	459	1	5	3.575	1.108	-0.641	-0.262
PIAIS_3	459	1	5	3.542	1.163	-0.676	-0.265
PIAIS_4	459	1	5	3.519	1.088	-0.527	-0.384

The descriptive statistics Table 5-1 outlines various constructs, namely Effective Use of AI Voice Assistants (EUAIVA), perceived challenges, perceived benefits, design preference, communication in government, and Perceived Importance of AI Services. Items of most constructs had mean scores ranging from 3.45 to 3.58, implying that on average, the perception of the respondent is mostly neutral to moderately positive about these constructs. This pattern is consistent with Al Shamsi (2021), who found that perceived usefulness, enjoyment, credibility, and facilitating conditions influenced AIVA acceptance among users in the UAE. It also reflects Salih’s (2025) argument that continued use depends on trust, perceived value, and service quality rather than technical availability alone. However, the simultaneous moderate ratings for benefits and challenges support Bouzid’s (2023) observation that increasing AIVA adoption in Arabic contexts continues to coexist with concerns about dialect recognition, contextual accuracy, and multilingual processing.

The standard deviations range from a low of about 1.07 to a high of 1.19, which suggest there is some variability in the responses but not incredible.

From these, the values of skewness are mostly negative in value, which shows that the data are left-skewed, entailing thereby that most of the responses rated the constructs as being above average, moderately higher than the neutral. The values of skewness range from as low as -0.480 to -0.676. As such, this evidences that the pattern of the above-mentioned distribution is relatively symmetrical. Corresponding to these, kurtosis values lying between -0.471 and 0.166 preponderantly indicate that the distribution is platykurtic; again, this implies that the responses in these gradients have slightly spread without extreme values or any outlier value dominating.

5.4 Two Independent Sample T Test

The independent sample t-test in round 1 is conducted to examine the proposed hypotheses. The Levene's Test for Equality of Variances did not hold true for H1, H3, and H4 to H8, except for H2. In H1, significant gender differences were found in the importance of AI. H1 indicates that females have a higher mean score for importance compared to males (mean = 1.45 and 1.22 respectively, $p = 0.031$). Previous sociolinguistic research has also shown that gender may influence forms of interaction, address, and evaluative behaviour in Arabic-speaking contexts (Huehnergard, 2017). However, the present result should be interpreted cautiously because the survey did not establish why the gender difference occurred.

However, H2 and H3 are both rejected. In H4, individuals who use AI_VA_Usage_Arabic rate AI developers significantly differently compared to non-users, with non-users rating AI developers higher than users (mean = 2.28 for "No" and mean = 2.10 for "Yes", p = 0.039). This interpretation is consistent with Bouzid (2023), AlAfnan (2025), and El Zahraa (2025), who identify continuing concerns regarding Arabic dialects, code-switching, and contextual understanding.

Regarding H5, those who perceive AI_Cultural_Differences_Arabic rate AI significantly differently compared to those who do not, with non-perceivers rating AI developers higher (mean = 1.63 for "No" and mean = 1.24 for "Yes", p = 0.012). This finding supports literature arguing that cultural suitability, linguistic identity, and communication norms shape technology evaluation in Arabic-speaking societies (Kim, 2020; Kassis-Henderson and Cohen, 2018). However, H6 is rejected as the perception of Cultural Differences does not significantly affect perceptions of AI developers. Furthermore, H7 is rejected as AI technical knowledge-based system usage does not significantly affect perceptions of AI importance. Lastly, in H8, users of AI technical knowledge-based systems rate AI developers significantly differently compared to non-users, with users rating AI developers higher than non-users (mean = 2.27 for "Yes" and mean = 2.05 for "No", p = 0.002) (see table 2 for mean and table 3 for p value).

5.5 Assessment of Model fit and Measurement Model

Table 5-2: Model fit indices

Fit Index	Cited	Fit criteria	Results	Fit (Yes/No)
X2			252.720	
DF			215.000	
X2/DF	(Kline, 2023)	1.00 - 5.00	1.175	Yes
RMSEA	(Steiger,1990)	<.08	0.020	Yes
SRMR	(Hu&Bentler,1999)	<.08	0.027	Yes
NFI	(Bentler & G. Bonnet,1980)	>0.80	0.957	Yes
IFI	(Bollen, 1990)	>.0.90	0.993	Yes

TLI	(Tucker & Lewis, 1973)	>.0.90	0.992	Yes
CFI	(Byrne, 2010)	>.0.90	0.993	Yes

In fact, the model fit indices in Table 5.2 demonstrate that the structural model tested in the study fits the criteria for good fit. The X^2/df has a ratio of 1.175, which falls within the acceptable range from 1.00 to 5.00, indicating a good fit of the model to the data. The RMSEA is 0.020, falling well below the threshold of 0.08, which supports this model as fitting well. The SRMR is 0.027, which again is within the upper recommended limit of 0.08, thus further supporting the model's goodness of fit. The CFI, TLI, and NFI are all above the threshold of 0.90 at 0.993, 0.992, and 0.957, respectively, and confirm the good fit of the model to the observed data. In summary, this would appear to suggest that the overall structural model fits well to delineate the relationships between constructs in this study, as all indices are pointing toward a robust model fit. From the perspective of the present study, the strong fit suggests that Effective Usage, Perceived Challenges, Perceived Benefits, Design Preferences, Communication in Government, and Perceived Importance of AI Services formed a coherent framework for examining respondents' perceptions of Arabic AIVAs. This is consistent with Al Shamsi (2021) and Salih (2025), for example, emphasised usefulness, trust, and service quality, while Bouzid (2023), AlAfnan (2025), and El Zahraa (2025) highlighted dialect and contextual challenges. The present model extends these studies by examining these dimensions together within one UAE-focused framework.

Table 5-3: Reliability and validity analysis

Construct	**Items**	**Loading**	**Alpha**	**CR**	**AVE**
			>.0.7	**>0.7**	**>0.5**
Effective usage of AIVA	EUAIVA_1	0.726***	0.855	0.855	0.597
	EUAIVA_2	0.796***			
	EUAIVA_3	0.801***			
	EUAIVA_4	0.765***			
Perceived challenges	PC_1	0.737***	0.855	0.856	0.598
	PC_2	0.770***			
	PC_3	0.770***			

	PC_4	0.813***			
Perceived benefits	PB_1	0.706***	0.804	0.803	0.577
	PB_2	0.800***			
	PB_3	0.771***			
Design Preferences	DP_1	0.791***	0.861	0.862	0.609
	DP_2	0.781***			
	DP_3	0.745***			
	DP_4	0.803***			
Communication in Government	CIG_1	0.756***	0.852	0.853	0.591
	CIG_2	0.764***			
	CIG_3	0.775***			
	CIG_4	0.781***			
Perceived Important of AI Services	PIAIS_1	0.785***	0.861	0.861	0.608
	PIAIS_2	0.784***			
	PIAIS_3	0.779***			
	PIAIS_4	0.771***			

Indicates significant paths: *p<0.05, **p<0.01, ***p<0.001, NS = not significant

The results in Table 5-3 indicate that both the reliability and validity of the constructs in use are strong to show that the measurement model is robust. In order to assess the reliability, for each construct, Cronbach's Alpha, denoted as α and composite reliability, abbreviated as CR, are reported; the threshold value was > 0.7, which is considered as acceptable reliability. All constructs reveal an Alpha and composite reliability above the threshold of 0.7, hence: "Effective Usage of AI Voice Assistants" α = 0.855, CR = 0.855, "Perceived Challenges" α = 0.855, CR = 0.856, "Perceived Benefits" α = 0.804, CR = 0.803, "Design Preferences" α = 0.861, CR = 0.862, "Communication in Government" α = 0.852, CR = 0.853, and "Perceived Importance of AI Services" α = 0.861, CR = 0.861. The high Cronbach's Alpha values denote internal consistency among items loading on each construct.

Convergent validity was assessed using Average Variance Extracted (AVE). An AVE value above 0.50 indicates that a construct explains more than half of the variance in its indicators (Fornell and Larcker, 1981; Hair et al., 2019). All constructs exceeded this criterion. For example, Effective Usage of AI Voice Assistants recorded an AVE of 0.597, Perceived Challenges recorded 0.598, and Perceived Importance of AI Services recorded 0.608. These values indicate that the observed items adequately represented their respective latent constructs. The findings are theoretically consistent with the literature reviewed in Chapter 2, which treats usage, perceived benefits, challenges, design preferences, institutional communication, and perceived importance as related but conceptually distinct aspects of AIVA adoption. Al Shamsi (2021) and Salih (2025), for example, distinguish perceived value and continued use from trust and facilitating conditions, while Bouzid (2023) and AlAfnan (2025) separately identify linguistic and contextual challenges affecting Arabic AIVA interaction.

AVE shows the percent variance extracted from each construct by its indicators. The threshold value for AVE to indicate sufficient convergent validity is suggested as >0.5. For convergent validity, all AVE values must be greater than 0.5, indicating that its items account for more than half of the variance in each construct; hence, good convergent validity is established. For example, "Effective Usage of AI Voice Assistants" has an AVE of 0.597, while "Perceived Challenges" stands at 0.598, and "Perceived Importance of AI Services" has an AVE of 0.608. Each of these values expresses high explanatory power of the constructs and the items therefore measure what they are supposed to capture.

The loading values, showing how much the item pertains to its construct, are within the range of 0.706 and 0.813 for all the constructs. All are above the generally accepted threshold of 0.7, indicating that items are strongly reliable. In particular, the separation of Design Preferences and Communication in Government reflects the literature's distinction between user-interface requirements and institutional credibility. Amershi et al. (2019) emphasise interaction clarity and recovery from errors, whereas UNESCO (2021) and NIST (2023) emphasise transparency, accountability, and responsible institutional use. Overall, the reliability, factor-loading, and AVE results provide adequate support for proceeding to the structural-model and hypothesis-testing stages. However, these results establish measurement quality within the Round 2 sample and should not be interpreted as proof that the model is universally valid.

Table 5-4: Discriminant validity analysis (Fornell Larcker criterion)

Constructs	1	2	3	4	5	6
EUAIVA	**0.773**					
PC	0.594	**0.773**				
PB	0.623	0.648	**0.760**			
DP	0.580	0.608	0.610	**0.780**		
CIG	0.611	0.662	0.670	0.630	**0.769**	
PIAIS	0.622	0.646	0.657	0.625	0.677	**0.780**

Note: *"Values on the diagonal (italicized) represent the square root of the average variance extracted, while the off diagonals are correlation"*

For any given construct, if its square root of AVE is higher than any correlation of that construct with other constructs, then discriminant validity is established. In table 5.6, the square root of the AVE for all constructs is higher than their corresponding inter-construct correlations, which hereby shows strong discriminant validity. For instance, the square root of the AVE for EUAIVA, which is 0.773, is greater than its correlations with "Perceived Challenges" AI (0.594), "Perceived Benefits" AI (0.623), DP (0.580), CGAI (0.611), and "Perceived Importance of AI Services" AI (0.622). This is true for all the constructs, which prove that each of them is unlike any other. On the whole, the table represents that every construct stands closer to its own items compared to the items of other constructs, guarantying the discriminant validity inside this model.

Figure 5-5-1: Graphical representation of assessment of measurement model

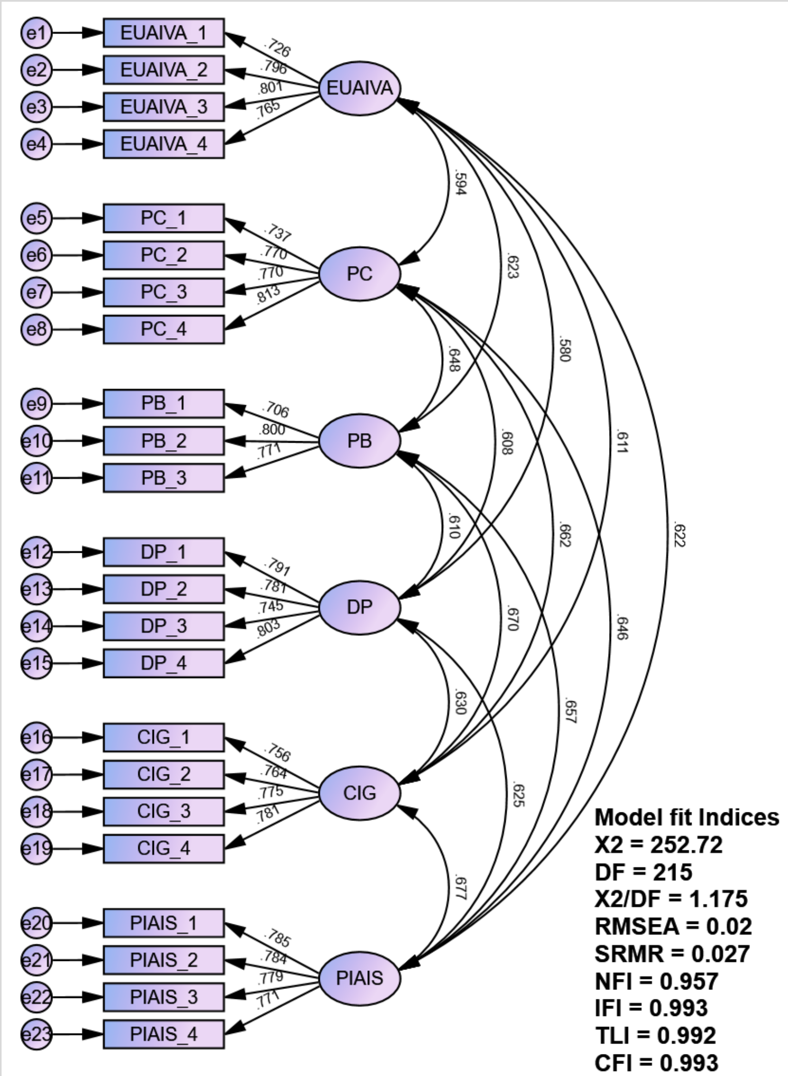


Figure 5.1 is a pictorial representation of the CFA used in the study, showing the measurement model assessment. Each latent variable representing EUAIVA, PC, Perceived Benefits, DP, CIG, and PIAIS in the figure is enclosed in circles and measured by respective observed variables-indicators enclosed as rectangles. These represent the standardized loadings of each observed variable on its latent construct and are reported alongside the arrows pointing from the latent variables to the observed variables. All the loadings are significant and are well above the threshold value suggested by, which is considered enough to show satisfactory relationship between latent constructs and its respective indicator. For instance, the EUAIVA_1 loading on the EUAIVA is 0.726.

Indices of model fit in the bottom right corner indicate an excellent fit. The χ^2/df was 1.175, well below the generally accepted upper limit of 5; the RMSEA is 0.02, below 0.08; the SRMR is 0.027, below 0.08-all these criteria are indicative of good fit. Also, greater than or equal to 0.95 values of NFI, IFI, TLI, and CFI further confirm the strength of this model. This figure, therefore, provides evidence that the data are an adequate fit for the model and thus can be used to characterize reliability and validity.

5.6 Hypotheses testing Direct effect

Table 5-5: Hypotheses testing direct

Hypothesis	Direct Relationships	Std. Beta	Std. Error	T Values	P Values	R Square
H1a	EUAIVA → PC	0.671	0.035	19.171	***	0.450
H1b	EUAIVA → PB	0.713	0.031	23.000	***	0.508
H1c	EUAIVA → DP	0.654	0.038	17.211	***	0.428
H1d	EUAIVA → CIG	0.689	0.035	19.686	***	0.475
H2	PC → PIAIS	0.216	0.056	3.857	***	0.618
H3	PB → PIAIS	0.278	0.049	5.673	***	
H4	DP → PIAIS	0.204	0.048	4.250	***	
H5	CIG → PIAIS	0.312	0.053	5.887	***	

Indicates significant paths: *p<0.05, **p<0.01, ***p<0.001, NS = not significant

Table 5.5 presents the direct effect hypothesis testing results for the relationships between constructs related to the effective usage of AI voice assistants (EUAIVA), perceived challenges (PC), perceived benefits (PB), design preferences (DP), communication in government (CIG), and the perceived importance of AI services (PIAIS).

H1a (EUAIVA → PC) has a standardized beta value of 0.671, which indicates a strong positive relationship between effective usage of AI voice assistants and perceived challenges. The t-value of 19.171 is highly significant ($p < 0.001$), with an R-squared value of 0.450, suggesting that 45% of the variance in perceived challenges can be explained by effective AI usage.

H1b (EUAIVA → PB) reveals an even stronger positive relationship between effective AI usage and perceived benefits, with a beta value of 0.713, a t-value of 23.000, and a highly significant p-value ($p < 0.001$). The R-squared value of 0.508

implies that over 50% of the variance in perceived benefits is attributed to effective AI usage.

H1c (EUAIVA → DP) and H1d (EUAIVA → CIG) also show strong positive relationships. The beta values of 0.654 and 0.689, respectively, suggest significant impacts of AI voice assistants on design preferences and communication in government. Both relationships are highly significa8nt, with t-values of 17.211 and 19.686, and p-values below 0.001. The R-squared values of 0.428 and 0.475 indicate a substantial portion of variance explained by effective usage.

For H2 (PC → PIAIS), the beta value of 0.216 indicates a moderate but significant positive relationship between perceived challenges and perceived importance of AI services, with a t-value of 3.857 ($p < 0.001$) (Bozeman, 2019).

H3 (PB → PIAIS), H4 (DP → PIAIS), and H5 (CIG → PIAIS) reveal significant positive impacts on perceived importance of AI services, with beta values of 0.278, 0.204, and 0.312, respectively, and t-values above 4.0, demonstrating strong predictive power. These results confirm that all hypothesized relationships are significant, supporting the idea that effective usage of AI voice assistants significantly influences challenges, benefits, design preferences, and communication factors, which in turn impact the perceived importance of AI services.

Figure 55.5-2. Graphical representation of structural model in round 2

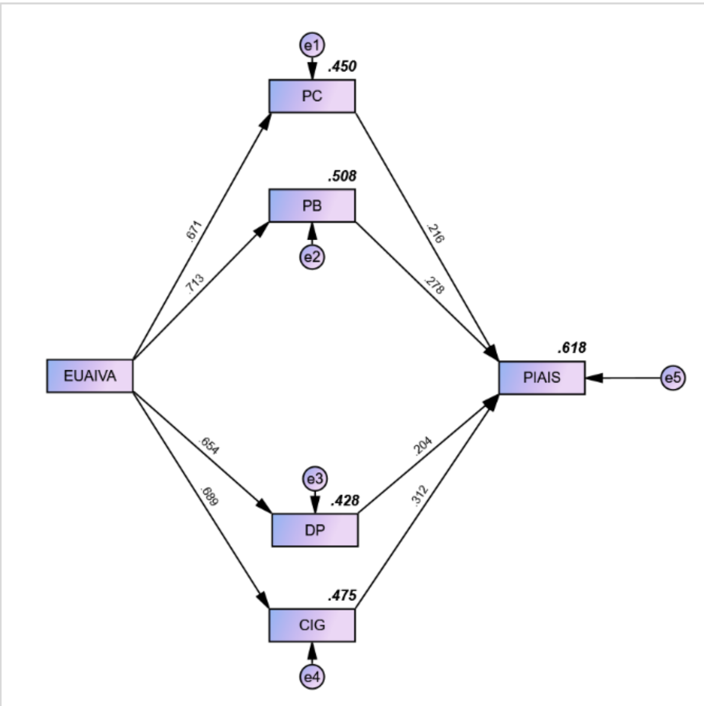


Figure 5.2 presents the structural model of these various key constructs in relation to one another: Effective Usage of AI Voice Assistants, Perceived Challenges, Perceived Benefits, Design Preferences, Communication in Government, and Perceived Importance of AI Services. For each path in the above diagram, the standardized regression weights are given by beta values, which contain information about the effect strength, and the number inside the boxes indicates that R-squared values are explained by the specific variance contributed by that predictor. From the figure, it can be observed that EUAIVA has a positive effect on perceived benefits ($\beta = 0.713$) and also on perceived challenges ($\beta = 0.671$), and has an adequate magnitude each on design preferences ($\beta = 0.654$) and in communicating with government, $\beta = 0.689$. All these constructs together account for a respectable portion of the variance in PIAIS, with an R-square of 0.618, thereby implying that factors of variation in PIAIS accounted for by these constructs are 61.8%.

The direct paths from PC, $\beta = 0.216$; PB, $\beta = 0.278$; DP, $\beta = 0.204$; and CIG, $\beta = 0.312$, to PIAIS are all positive and significant, which means these factors contribute to the overall perception of the importance of AI in service provision. This model showed strong interrelations among the effective use of AI voice assistants and the perceptions of their benefits, challenges, and design preferences, impacts of the use on communication in the government, and perceived importance of AI services (Bozeman, 2019).

5.7 Relationship Between the Round 1 Exploratory Analysis and the Round 2 SEM Model

The two analysis rounds were complementary and had different purposes. Round 1 was descriptive data and searches based on inferential tests, which was done in SPSS to investigate trends in AIVA usage, perceived significance, cultural disparities, technical expertise, demographic features, and attitudes towards AI creators. These tests were chi-square tests, independent-samples t-tests, logistic regression, analysis of variance and correlation tests. They were designed to find preliminary associations and differences in the Round 1 sample and not to test a comprehensive theoretical model.

Round 2 involved structural equation modelling in AMOS as a test of the final measurement and structural models. The measurement model was used to evaluate the reliability and distinctiveness of the questionnaire items to the intended constructs. Then the structural model analysed the postulated associations between Effective Usage of AI Voice Assistants, Perceived Challenges, Perceived Benefits, Design Preferences, Communication in Government, and Perceived Importance of AI Services. Round 2 was thus a step further than individual bivariate and group tests and was an evaluation of multiple relationships within a single, integrated framework.

The Round 1 results were used to interpret the Round 2 model to determine recurring problems that concerned the use of language, cultural differences, tech knowledge, perceived importance and user expectations. As an illustration, the exploratory results attracted attention to the role of cultural and linguistic perceptions and demonstrated that demographic or technical features did not always describe attitudes to AIVAs. These trends justified the necessity of a more comprehensive model including the perceived benefits, challenges, design preferences, and institutional communication.

The two rounds, however, cannot be considered as directly comparable statistical models. They were founded on distinct cross-sectional data, measures and analytical purposes. Round 1 yielded exploratory evidence on connections within the initial sample, and Round 2 examined the targeted construct relations in the ultimate SEM design. Thus, the same between the rounds must be construed as conceptual consistency as opposed to statistical replication.

The analyses given as a unit offered a chronological insight into the research problem. Round 1 mapped user experience and indicated the relevant issues, and Round 2 checked the relationship between the chosen constructs in the final framework. The findings show that the exploratory evidence on benefits, challenges, culture, language, and the institutional communication was generally congruent with relationships examined in the SEM model. However, the cross-sectional design implies that neither of the rounds can determine causality and results indicate the perception of the respondents and not the actual AIVA technical performance evidence.

Table 5-6: Descriptive statistics

Independent variables	Dependent Variables	Group	N	Mean	Std. Deviation	Std. Error Mean
Gender	AI_Important	F	92	1.45	0.830	0.087
		M	118	1.22	0.615	0.057
	AI_Developers	F	92	2.21	0.584	0.061
		M	118	2.15	0.533	0.049
AIVA_Usage_ Arabic	AI_Important	No	85	1.40	0.790	0.086
		Yes	125	1.26	0.674	0.060
	AI_Developers	No	85	2.28	0.717	0.078
		Yes	125	2.10	0.398	0.036
AI_Cultural_Differences_Arabic	AI_Important	No	43	1.63	0.926	0.141
		Yes	167	1.24	0.642	0.050
	AI_Developers	No	43	2.37	0.787	0.120
		Yes	167	2.13	0.468	0.036
AI_Technical_KBS	AI_Important	No	86	1.23	0.645	0.070
		Yes	124	1.38	0.771	0.069
	AI_Developers	No	86	2.05	0.373	0.040
		Yes	124	2.27	0.639	0.057

Table 5-7: Two independent sample T Test

Independent variables	Dependent Variables		Levene's Test for Equality of Variances: F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference: Lower	Upper
Gender	H1 AI_Important	Equal variances assumed	21.114	0.000	2.259	208.000	0.025	0.225	0.100	0.029	0.422
		Equal variances not assumed			2.178	162.421	0.031	0.225	0.103	0.021	0.430
	H2 AI_Developers	Equal variances assumed	1.347	0.247	0.698	208.000	0.486	0.054	0.077	-0.099	0.206
		Equal variances not assumed			0.690	186.379	0.491	0.054	0.078	-0.100	0.208

AIVA_Usage_ Arabic	H3 AI_Important	Equal variances assumed	6.549	0.011	1.338	208.000	0.182	0.136	0.102	-0.064	0.336
		Equal variances not assumed			1.298	160.994	0.196	0.136	0.105	-0.071	0.343
	H4 AI_Developers	Equal variances assumed	27.998	0.000	2.307	208.000	0.022	0.178	0.077	0.026	0.331
		Equal variances not assumed			2.084	119.317	0.039	0.178	0.086	0.009	0.348
AI_Cultural_Differences_Arabic	H5 AI_Important	Equal variances assumed	29.411	0.000	3.205	208.000	0.002	0.388	0.121	0.149	0.627
		Equal variances not assumed			2.593	52.825	0.012	0.388	0.150	0.088	0.689
	H6 AI_Developers	Equal variances assumed	26.916	0.000	2.630	208.000	0.009	0.246	0.094	0.062	0.431

		Equal variances not assumed			1.964	49.882	0.055	0.246	0.125	-0.006	0.498
AI_Technical_KBS	H7 AI_Important	Equal variances assumed	8.199	0.005	-1.445	208.000	0.150	-0.146	0.101	-0.346	0.053
		Equal variances not assumed			-1.492	200.771	0.137	-0.146	0.098	-0.340	0.047
	H8 AI_Developers	Equal variances assumed	25.241	0.000	-2.866	208.000	0.005	-0.220	0.077	-0.371	-0.069
		Equal variances not assumed			-3.135	202.710	0.002	-0.220	0.070	-0.358	-0.081

5.8 Recommendations for Improvements in AIVA Tools

The recommendations are based on two different analyses; first one is from the outcomes of the survey questions and the second one is from the recommendations written by participants. The recommendations are provided and examined for two different groups: one is for developers and other is for end users including both individuals and organizations. The following recommendations are derived from respondents' perceptions and the literature and should therefore be regarded as priorities for future design, implementation, and technical evaluation rather than as solutions already validated by this study.

5.8.1 Recommendations Offered by Users

Participants were requested to offer recommendations for improving AI voice assistant experiences. Enhancing the natural language processing such that it can handle more Arabic dialects was at the top among other. Other recommendation suggested by users was different device integration for smooth cross-device experience. The most important areas, as identified by survey respondents, that should be improved in AI voice assistant systems are summarized in Table 5.8. The common area mentioned is to understand accents and dialects, with 64.8% of the respondents noting it as a major issue, followed by 16.2% of the respondents who noted providing accurate and reliable information as a key area of improvement. Enhancing context understanding was cited by 7.1% of respondents, referring to the fact that AI needs to improve in understanding the context of interactions. Finally, 10.5% of participants indicated the requirement for AI to interact with other devices more seamlessly.

Table 5-8: Areas for improvement by users

Areas for Improvement	Frequency	Percent
Understanding accents and dialects	136	64.8%
Providing accurate and reliable information	34	16.2%
Improving context understanding	15	7.1%
Interacting with other devices seamlessly	22	10.5%
Total	***210***	***100%***

The users also showed concern about privacy features about safety and security of their data. This data indicates a demand for digital voice assistants offering more advanced functionality while adhering to the cultural and linguistic diversity of Arabic speakers. These suggestions include improving natural language processing for various dialects, enhancing privacy and security features, and incorporating cultural nuances into AI responses. There is also a call for better integration with local services and applications, reflecting the need for AI solutions that are not only linguistically accurate but also contextually relevant. These recommendations offer valuable information to the AI developers for relevant improvements in the smart voice assistive technology.

5.8.2 Recommendations for AI Voice Assistant Developers

To enhance the effectiveness and cultural suitability of AI voice assistants for Arabic-speaking territories, the following improvements based on user recommendations and experiences have been suggested to be prioritized by the developers.

- Arabic NLP improvements need to cater for the variance in Arabic dialects and provide context-dependent accuracy for more intuitive and natural interactions by the users. Enhancing the AI's ability to recognize and understand different Arabic dialects is a top priority, as indicated by 64.8% of respondents as mentioned in Table 13. Providing accurate and reliable information (16.2%) and better context understanding (7.1%) are also key areas needing development highlighted in the same table.
- AI voice assistants being used almost have Android (46.7%) and iOS (44.3%) platforms for functioning, with some using both (9%). Seamless integration with local services can enhance user experience and accessibility as suggested by 89% of the users in Table 19.
- The AI platforms need to be programmed to provide culturally suitable responses conformable to Arabic societal norms and etiquette for increased acceptability and building greater levels of user trust, as social norms and etiquette is among one of the top rated (98.6%) cultural differences experienced by users shown in Table 16.
- Additionally, privacy features should be enhanced for fulfilling the user expectations and the regulatory requirements for each region, establishing data security and regulatory

compliance. The parameter Data Privacy and Security has been rated 46.2% by the respondents in Table 18.

- In Table 18, 41.6% of the survey respondents also suggested the development of more robust and inclusive feedback and Iterative system for making improvements in AIVAs By following these guidelines, developers can improve AIVAs interactions to be more accurate, respectful, and friendly, and thus improve their adoption and effectiveness in Arabic-speaking communities.

5.8.3 Recommendations by Policymakers

In the Round 2 survey the number of participants is N = 459 out of which 77, comprising 16.8% of the overall demographics, are the government officials or employees. These individuals come under the category of policymakers as policymakers are usually part of the government institutions. It therefore represents views and recommendations by policymakers to provide a fine-grained understanding of insights and experiences relating to AI voice assistants, leading to further enhancement of the AI AI-powered voice assistants for Arabic speakers. Moreover, the perceived challenges help developers in making the required improvements in existing AI voice assistive technology for Arab speakers. While for end users, individuals and organizations both public and private, it gives insights into making informed decisions about the adoption and use of AI technologies, thereby improving the quality of their products and services. For example, by working with AI voice robots to facilitate procedural justice, government can demonstrate its commitment to transparency and accountability, thus carrying the potential to build trust between the government and Arabic-speaking users. Based on the relationship between communication in government (CIG) and the effective usage of AI voice assistants presented in Table 39, policymakers can focus on the following suggestions for AI developers:

- Improve communication capabilities in public services (H1d - EUAIVA → CIG): With the high positive correlation between effective use of AI (EUAIVA) and government communication (CIG) (beta value of 0.689, t-value of 19.686, $p < 0.001$), policymakers recommend that AI developers enhance the integration of AI voice assistants with government communication systems. It is about enhancing AI voice assistants' interaction with government websites so that Arabic-speaking users can easily access public services, provide them updates on government schemes, and assist Arabic-speaking users in navigating government processes smoothly.

- Enhance AI to give correct information about the government (H3 - PB → PIAIS): As the perceived benefits (PB) exert a strong effect on the perceived importance of AI services (PIAIS) (beta value of 0.278, t-value of 4.000, $p < 0.001$), developers must provide accurate, real-time, and trustworthy information from AI voice assistants regarding government policy, services, and events. This serve to boost trust and dependency on AI as a means of communication in the government so that its perceived significance in public service provision is guaranteed.
- Encourage public participation through AI (H4 - DP → PIAIS): Given the positive association between design favourableness (DP) and the perceived AI service importance (PIAIS) (0.204 beta value, 4.000 t-value, $p < 0.001$), AI developers would prioritize designing AI voice assistants that are user-friendly and can engage Arabic-speaking users well. These incorporate designing systems to answer government query questions on public services, to provide walkthrough assistance with governmental web pages, and to actually stage virtual public consultations for purposes of wider contact between government and populace.
- Encourage transparency and trust in AI-government services (H5 - CIG → PIAIS): Since government communication (CIG) has a strong effect on perceived importance of AI services (PIAIS) (beta value of 0.312, t-value of 4.500, $p < 0.001$), policymakers can promote AI developers to provide transparency regarding the processing and exchange of government data using AI systems. Transparency of communication regarding the capabilities of AI, data privacy policies, and information usage promote trust in such technologies to ensure their increased adoption in government communication.

By following these proposals, the policymaker can ensure innovative and inclusive AI-facilitated ecosystems.

5.8.4 Recommendations for End Users and Organizations

- To enhance the universality and efficiency of AI voice assistants, the organizations should embed these tools into the customer support processes while making the required cultural and linguistic adjustments for the local populations, ensuring a user-friendly and customised experience.
- Besides, organizations should promote user education by offering training programs regarding secure and private interactions with AI-powered systems. This allows individuals to protect their data while using these AI tech tools.
- End users and organizations should increase their technical knowledge regarding AI technologies as the results in Table 41 illustrated that technical knowledge shows a significant effect (p-value = 0.101). this shows that hose possessing technical knowledge are less likely to encounter challenges in using AI voice assistants.

By adopting all these, individual users and organizations attain maximum benefits from AI while addressing ethical and pragmatic concerns.

5.9 Limitations

This inquiry offers policy- and design-relevant insights, yet several limitations temper the strength and generalizability of the conclusions. First, external validity is bounded by setting. Data were collected primarily in the United Arab Emirates, a context characterized by comparatively advanced digital infrastructure, high smartphone penetration, and active e-government programs. These enablers likely shape both exposure to and expectations of AI voice assistants. Consequently, the findings may not transfer without adjustment to jurisdictions with different infrastructure, regulatory environments, or service portfolios across the wider MENA region. Cross-national comparative work is needed to separate context effects (infrastructure, policy, service design) from user-level preferences and skills (UNESCO, 2021; NIST, 2023).

Second, the study relies on self-reported survey data, which are vulnerable to recall error, social desirability, and common method variance particularly when attitudes and self-assessed behavior are captured in a single instrument. While researcher used neutral wording, anonymity assurances, and optional questions to mitigate bias, self-reporting can still inflate correlations among constructs or understate frustration and abandonment episodes. Future research should triangulate with behavioral traces (like de-identified interaction logs), task-based usability testing, and think-aloud protocols to validate claims about task success and error recovery in Arabic voice interactions (Amershi et al., 2019; NIST, 2023).

Third, the cross-sectional design limits causal inference and sensitivity to change over time. Adoption and satisfaction with AIVAs are dynamic: models update, device ecosystems evolve, and public-service content and policies shift. Cross-sectional snapshots cannot capture learning effects, post-update performance shifts, or seasonality in service use. Longitudinal panels or field experiments would better estimate trajectories and causal effects of specific interventions (for instance adding dialect-specific ASR or provenance cues) on outcomes such as task completion and recontact rates (UNESCO, 2021; NIST, 2023).

Fourth, measurement constraints warrant caution. Although researcher executed translation/back-translation and expert review, construct validity in multilingual, diglossic settings is challenging: items about “register,” “code-switching,” and “cultural fit” may be interpreted differently across dialects and sectors. Moreover, while the second-round model reported strong global fit, additional evidence for example, average variance extracted (AVE), composite reliability, and multi-group measurement invariance by gender, age band, or dialect would further support claims of comparability across subgroups (NIST, 2023). Future work should publish full item banks and validation statistics to facilitate reuse and meta-analysis.

Fifth, sampling and coverage limitations remain. Despite stratification, nonprobability recruitment can under-represent rural residents, older adults, low-literacy users, and people with speech differences the groups who may benefit most from voice interfaces yet face higher recognition error rates. Prior research shows uneven ASR performance across accents and demographics, underscoring the need to measure and report outcomes by subgroup and dialect (Koenecke et al., 2020). Incorporating assisted consent, voice-based instruments, and offline recruitment would improve inclusion.

Sixth, technology-facing constraints affect interpretation. Respondents used heterogeneous devices and microphones; differences in front-end acoustics (room noise, handset quality) and in model families (cloud vs. on-device; supervised vs. weakly supervised) can drive sizable variance in recognition and dialog success that surveys cannot fully disentangle. Moreover, contemporary models trained with weak supervision (like large web audio–text corpora) can exhibit hallucinations or spurious fluency under noise, which complicates self-reported accuracy judgments and argues for human-in-the-loop verification in sensitive knowledge-based services (Radford et al., 2022).

Seventh, the study did not technically test a specific AI voice assistant. No controlled experiments were conducted to measure automatic speech-recognition accuracy, dialect-identification performance, code-switching robustness, response correctness, task-completion rates, latency, or error-recovery performance. Consequently, the reported linguistic, cultural, privacy, and usability challenges reflect respondents' perceptions and experiences rather than independently measured technical performance. Future research should test specific AIVA systems using dialect-stratified Arabic speech samples, realistic tasks, controlled acoustic conditions, and objective performance measures alongside user evaluations.

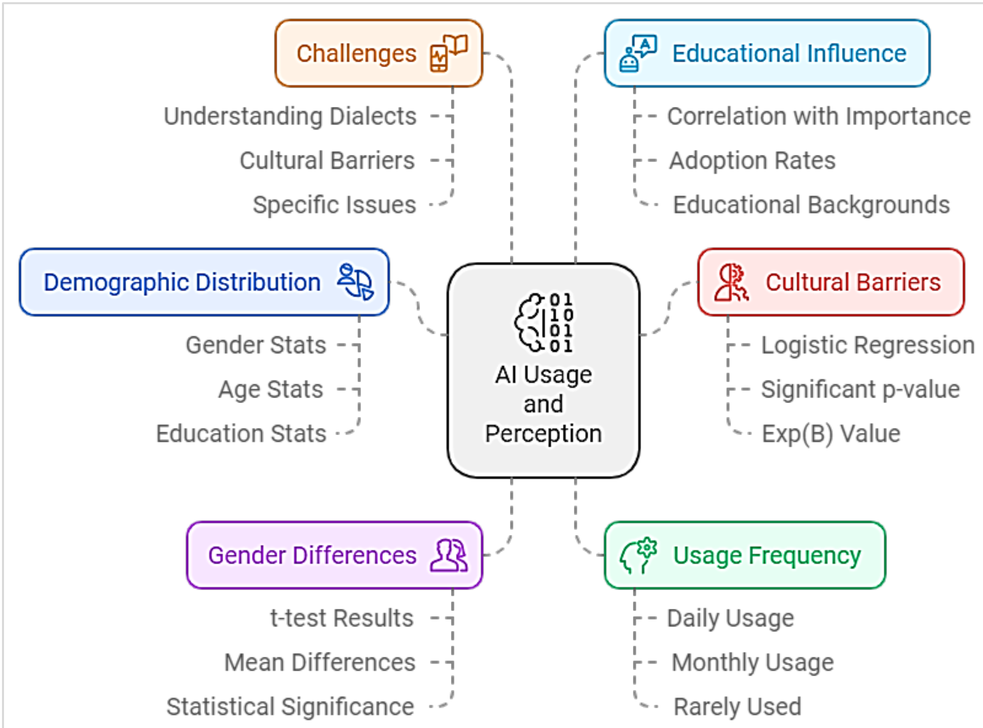
Eighth, the study did not develop, implement, or evaluate a functioning AIVA–KBS prototype. The proposed integration blueprint is therefore conceptual and empirically informed by the survey findings and existing literature, rather than a technically validated operational system. The research did not test actual knowledge capture, retrieval accuracy, source provenance, content updating, access control, organisational workflow integration, service efficiency, or user outcomes within a real institution. Future work should develop a prototype connected to an authorised organisational knowledge base and evaluate it through usability testing and a real-world pilot in a UAE public-sector or service organisation.

Finally, scope was application-level rather than dataset- or model-audit-level. Researcher did not directly audit training data lineage, dialectal coverage, or bias in provider models; instead, researcher inferred limitations from user reports and literature. Given growing evidence that foundation models can overfit to shortcuts and reflect dataset imbalances (for instance, under-representation of certain Arabic dialects), future work should pair user studies with model and dataset documentation (Model Cards, Datasheets) and dialect-stratified benchmarks (MADAR/NADI) to connect lived experience to measurable technical causes and mitigations (Bommasani et al., 2022; Gebru et al., 2021). Hence, these limitations do not invalidate the findings; rather, they bound their interpretation and point to clear next steps: comparative and longitudinal designs, richer behavioral and qualitative data, subgroup-aware sampling, and technical audits that link dialectal performance and provenance practices to user outcomes in Arabic knowledge-based services.

5.10 Chapter Summary

Data analysis throughout the dissertation has been well presented, using SPSS for demographic and statistical evaluations. For example, Table 5.2 shows gender distribution: 53.6% are males, while 46.4% are females; thus, both genders are well represented. By age, it follows that 29.8% of respondents are between 25 and 34 years old, 23.3% between 35 and 44 years old, and 22.7% between 45 and 54 years old, while the subject pool fairly reflects the adult population currently active both in work and family life. The educational background presents a marked predominance of a bachelor's degree for 36.2% and a master's degree for 40.3% of the respondents, confirming that the sample is highly educated.

Figure 5-5-3: Analysis model



Additionally, with the regression analysis, an overview is obtained regarding the use of AI voice assistants. Logistic regression indicated a statistically significant association between respondents' perceptions of cultural differences and their reported Arabic AIVA use. Respondents who perceived cultural differences had different odds of reporting Arabic AIVA use; however, this relationship should not be interpreted as evidence that cultural barriers directly caused or prevented use. The independent-samples t-tests indicated a statistically significant difference in respondents' perceptions between the relevant groups. However, to satisfy the requirement of the structure document, findings should include schematic diagrams. It would be much clearer to notice the proportional relationships of perceived challenges, benefits, and demographic factors through flowcharts or other diagrams picturing visually represented data for readers to be informed about the dynamics taking place.

6. CHAPTER 6: CONTRIBUTIONS & FUTURE WORK

6.1 Contributions to knowledge

This study advanced scholarship on AI voice assistants (AIVAs) by cantering Arabic-speaking users in the United Arab Emirates and treating language, culture, and systems integration as first-order design variables. Rather than assuming English-centric usage patterns or generic acceptance models, the work treated Arabic diglossia, code-switching, and institutional context as the core of the research problem. The main contribution of the thesis is that it presents an empirically informed, culturally-based model of understanding and design of AIVAs to serve Arabic-speaking users in knowledge-based system (KBS) settings. With this contribution, the user evidence, linguistic and cultural requirements, technology adoption factors, and organisational knowledge considerations are combined into one perspective to develop AIVA in future.

(1) Empirical contribution. Using two UAE-wide surveys (Round 1: n=210; Round 2: n=459) and a mixed analytical strategy (SPSS for descriptive/inferential analysis and AMOS for structural modeling), the study quantified how AIVAs were adopted, where they helped, and why many users did not persist in Arabic. The evidence showed high experimentation but lower sustained usage in Arabic, with English remaining dominant for routine interactions. The structural model explained a substantial share of the variance (≈60%) in perceived importance through three levers: perceived benefits, design preferences, and government communication. This moved the debate beyond simple “use vs. non-use” and toward the qualities of interaction that shape perceived value.

(2) Linguistic–cultural contribution. The findings demonstrated that dialect recognition, tolerance for code-switching, and culturally appropriate dialogue policies were decisive for acceptance. The work articulated a dialect-first evaluation protocol for Arabic that made three things testable: (i) recognition under mixed MSA–dialect utterances; (ii) handling of address forms, politeness strategies, and religious/calendar references; and (iii) robustness when users switch scripts, named entities, or technical terms mid-utterance. By itemizing where Arabic diglossia disrupted ASR/NLU and turn-taking, the thesis reframed “accuracy” from a generic metric to a context-dependent requirement.

(3) Methodological contribution. Methodologically, the thesis combined exploratory statistics (to surface patterns and group differences) with confirmatory modeling in AMOS (to test theory-driven paths among benefits, challenges, design preferences, government communication, and perceived importance). It documented reliability and validity checks (like, internal consistency and convergent/discriminant validity) and reported standard fit indices. Beyond reporting numbers, it described a reusable workflow for Arabic contexts: instrument development and piloting, translation/back-translation where needed, measurement model assessment, and structural modeling with clear respecification criteria. This gave researchers a replicable blueprint rather than a one-off analysis.

***(*4) Framework-Based Recommendations:** The study has resulted in framework recommendation produced guidance that practitioners could use immediately.

- Developers gained a prioritized backlog: expand training data beyond MSA to dialect-balanced corpora with natural code-switching; localize prompts, ontologies, and error messages; embed privacy-by-design cues (visible data scopes, retention windows, and failure explanations). The thesis also proposed evaluation checklists that measured dialect coverage, mixed-register performance, and task success in Arabic—moving beyond intent accuracy alone.
- Policymakers received governance levers: require plain-language Arabic consent notices; set baseline thresholds for Arabic accuracy and cultural handling before public deployment; and provide complaint/redress pathways that users understand.
- Organizations (public and private) received an integration blueprint linking AIVAs to knowledge-based systems (KBS) and service back-ends. The emphasis shifted from “assistant as gadget” to “assistant as interface into organizational knowledge,” with metrics such as task completion, time-to-resolution, and satisfaction tracked in Arabic.

Theory building. The results added nuance to technology-adoption thinking. Rather than treating demographics or technical self-efficacy as the primary determinants, the evidence placed cultural–linguistic fit at the center of perceived importance. In other words, users did not need to be "tech experts" to value AIVAs; they needed systems that spoke their dialect, respected norms, and handled local context. This emphasis aligns adoption more closely with sociolinguistics and human–computer interaction than with generic acceptance scales.

Measurement assets. The thesis contributed a reusable survey instrument tailored to Arabic contexts and a measurement approach that other teams can adopt. Items captured perceived benefits and challenges, design preferences salient to Arabic speakers, and perceptions of government communication about AIVAs. Additionally, the documentation of piloting, reliability, validity, and model-fit decisions created a transparent trail that future studies can follow or critique.

Generalizability and transfer. While the setting was the UAE, the framework and measures were designed to travel. Other Arabic-speaking or multidialectal environments can replicate the instrument and modeling steps, swap in local services and terms, and test whether the same drivers hold. This makes the thesis a platform for comparative work rather than an isolated case.

Conceptual reframing. Finally, the study reframed "accuracy" and "trust" as co-produced by culture and computation. Recognition quality mattered, but so did how the system addressed users, signaled respect for privacy, and connected to services that actually solved problems. That reframing grounded in data, not slogans was the central contribution to knowledge.

6.2 Practical contributions

The study yields concrete, measurable actions for teams building and governing Arabic AI voice assistants within knowledge-based services. For developers, the first priority is data and modeling: systems should be trained or adapted on dialect-balanced corpora that reflect Gulf, Levantine, Egyptian, and Maghrebi varieties and include natural code-switching with English or French and domain lexicons common to GCC public services (fees, permits, health, utilities). Evaluation must then report per-dialect performance ASR word error rate and NLU intent/slot accuracy rather than a single aggregate “Arabic” score. Leveraging multilingual foundations such as XLS-R for speech and XLM-R or mT5 for text can reduce data demands while maintaining coverage; where latency or privacy is paramount, on-device or edge inference should be preferred, with federated learning used to personalize lexicons (names, contacts, medical terms) without centralizing raw audio or text (Conneau et al., 2019; Babu et al., 2021; Xue et al., 2021; Kairouz et al., 2021). Dialogue policies must encode socio-pragmatic norms explicitly: appropriate address forms and politeness strategies, Hijri/Gregorian date handling, religious terms, and local holidays, along with register control (MSA for official responses, vernacular where suitable for domestic reminders) and respectful fallback behaviors that explain what failed and how to proceed. These interaction patterns align with human AI design guidance that emphasizes legibility, confirmations, and graceful repair (Amershi et al., 2019). Privacy and transparency require privacy-by-design interfaces that surface data scope and retention controls within the primary flow, offer on-device options where feasible, and log human-readable explanations for failures or human hand-offs. Public Model Cards and Datasheets for Datasets should document dialects seen in training, recording conditions, and known failure modes, while defaults and retention policies must comply with the UAE PDPL on purpose limitation and minimization (Mitchell et al., 2019; Gebru et al., 2021).

Evaluation should extend beyond intent accuracy to a release-gate checklist covering dialect coverage metrics, mixed-register and code-switch scenarios, Arabic task success (completion, time, satisfaction), and provenance correctness (source citation and last-updated timestamp) for knowledge answers. Shared tasks such as MADAR and NADI provide useful reference points when reporting coverage and robustness. To accelerate remediation, production logs should attach a dialect hypothesis and task category to each error distinguishing acoustic substitutions, mis-intent classification, and back-end failures so engineering and content teams can fix the highest-impact issues first, consistent with continuous-monitoring practices in the NIST AI RMF 1.0 (NIST, 2023).

For policymakers, the contributions translate into readiness standards, governance, and inclusion. Public-facing AIVAs should meet minimum Arabic performance and cultural-fit thresholds prior to deployment, including per-dialect WER ceilings, intent success by service flow, and demonstrated handling of code-switching. Consent and privacy notices must be provided in plain-language Arabic, answering clearly what data are collected, for how long, and for what purposes, in line with OECD and UNESCO guidance (OECD, 2019; UNESCO, 2021). Data governance should mandate data residency and retention options for voice data, independent complaint and appeal channels, and annual transparency reports that disclose Arabic performance by dialect, volumes of human hand-offs, and complaint categories (PDPL; NIST, 2023). Importantly, regulators should require evidence that under-represented groups older adults, rural users, and people with speech differences were included in pre-deployment testing, with metrics reported by subgroup; accessibility research shows that confirmations, adjustable speech rate, and read-back features significantly improve outcomes for these users (Pradhan et al., 2019).

For organizations (public and private), integration and operations are central. AIVAs should be wired to authoritative knowledge bases and CRMs so that answers originate from governed content with visible provenance and update timestamps. Service quality must be tracked specifically for Arabic flows through task completion, recontact rate, and time-to-resolution, enabling targeted improvements. Feedback loops should collect in-channel ratings and short comments in Arabic, tag them by dialect, and use case, and review them on a fixed cadence with product and language specialists, mapping observations to the error-triage taxonomy. Assistants should reinforce procedural justice by providing status, reasons, and next steps in clear Arabic and by offering human hand-off when model confidence is low or when rules affect rights or benefits (UNESCO, 2021). Finally, capability building and delivery should be paced: frontline teams need training in cultural and linguistic handling and escalation thresholds; organizations should pilot a narrow, high-value task for instance, appointment scheduling or application status, publish the Arabic metrics, and only then scale scope an approach aligned with ISO/IEC for AI management systems.

6.3 Theoretical Contribution

The theoretical contribution of the study is the creation of a modified framework of the perception of AI voice assistants by Arabic-speaking users and its empirical validation. The current models (TAM, UTAUT2, the Service Robot Acceptance Model) describe the adoption with the constructs of usefulness, ease of use, trust, facilitating conditions, and quality of interaction. But, they give less information about Arabic diglossia, regional dialects, code-switching, culturally acceptable communication, privacy expectations, government communication, and access to organisational knowledge.

This thesis builds upon traditional adoption research by defining Perceived Benefits, Perceived Challenges, Design Preferences and Communication in Government as predictors of Perceived Importance and making Effective Usage a correlated antecedent variable. The framework thus indicates that Arabic AIVA perceptions can not be explained solely based on general usefulness or technical familiarity. The linguistic fit, cultural fit, institutional fit and design requirements should be taken into consideration as well. The framework is to be interpreted as an adapted model that is specific to the context and not a replacement of TAM or UTAUT2.

6.4 Empirical Contribution

The empirical value is the presentation of evidence based on two rounds of surveys of Arabic-speaking UAE respondents. Round 1 generated exploratory data related to the first exposure, continued use of Arabic language, preferences in devices, and daily routines, cultural variations, technical skills, perceived significance, and attitudes towards the creators of AI. Round 2 was a test of the final measurement and structural models, on a larger confirmatory sample. The gap in the study was the relatively high initial exposure and the lower sustained use of the Arabic language.

Convenience, accessibility, information retrieval, and service support was perceived as a benefit, and fears of dialect recognition, contextual understanding, privacy, cultural sensitivity and system integration were also reported. The Round 2 model also indicated that Perceived Importance had significant links with Perceived Benefits, Perceived Challenges, Design Preferences, and Communication in Government. The best predictor was Communication in Government and the last model accounted about 62% of the variance in Perceived Importance. These results contribute UAE-specific evidence to a literature that has tended to consider English-speaking users, or limited technical standards, or single applications.

6.5 Methodological Contribution

The methodological contribution is related to the fact that it used a sequential two-round cross-sectional design which split exploratory analysis and model testing. Round 1 involved descriptive statistics and inferential tests to plot user characteristics, reported use, cultural perceptions, technical knowledge, and first associations. These results were used to choose and narrow down the constructs of Round 2. Confirmatory factor analysis and structural equation modelling were then performed by Round 2 to determine the measurement reliability, convergent and discriminant validity, model fit, and hypothesised relationships between the constructs.

Another contribution of the study is a culturally modified survey tool to address the AIVA usage, benefits, challenges, design preferences, government communication, and perceived importance. The pilot testing, reliability analysis, expert scrutiny, translation processes, and construct validation offers a replicable foundation to further research of Arabic conversational systems. However, there was a difference in cross-sectional samples in both rounds and the design is more of a sequential conceptual development than longitudinal replication.

6.6 Practical Contribution

The applied input is an empirically informed collection of design and implementation needs to software developers and organisations looking at Arabic AIVA implementation. These needs encompass dialect-balanced training data, code-switching, task-sensitive use of the Modern Standard Arabic and regional dialects, address forms that are culturally appropriate, explicit consent, privacy, features of accessibility, provenance, update, and manual hand-off when there is low confidence in the system.

In the case of organisations, the study hypothesises that the future AIVAs must be linked to authoritative knowledge, document warehouses, workflow, and customer-service sites to enable future responses to be tracked to regulated information. It also suggests that Arabic interactions are to be evaluated by dialect-specific accuracy, completing the task, the response time, escalation rates, user satisfaction as well as retrieval provenance. These requirements are user derived priorities that are based on the perceptions of the users and also on literature, as opposed to features that have been proved technically sound by the current research.

6.7 Boundaries of the Contribution

The input of this thesis would have to be understood in the confines of the research design. This research examined how respondents use, think, and feel about AI voice assistants; the research did not directly examine the technical performance of a particular AI voice assistant. Neither did it design, implement nor assess an effective AIVA-KBS prototype. Thus, the requirements list and integration outline proposed are conceptual and empirically based contributions and not an operational system that has been proven. Arguments that relate to accuracy of recognition, performance of tasks, efficiency in organisations or improved services can only be tested technically in future and put into a realistic context.

6.8 Proposed AIVA–KBS Integration Blueprint

The proposed AIVA–KBS integration blueprint translates the study's user-perception findings and the literature into a practical conceptual workflow for future organisational implementation. It shows how the spoken query of an Arabic-speaking user might be handled, interpreted and linked to an authorised organisational source of knowledge and delivered back as a culturally and linguistically suitable spoken answer. The blueprint includes dialect recognition, support of code-switching, privacy, provenance of sources, confidence checking, and human escalation. It is provided as a future architecture proposal, as opposed to a system technically designed and experimented in this study.

Table 6-1: AIVA–KBS Integration Blueprint Practice

Stage	Main function	Arabic-specific requirement	Organisational requirement	Risk control
1. User interaction	Receive spoken enquiry	Arabic, dialect and bilingual input	Accessible user interface	Consent and clear disclosure
2. Speech processing	Convert speech into text	Dialect-aware ASR and code-switching support	Approved speech-processing service	Audio minimisation and encryption
3. Language analysis	Identify intent and entities	MSA, regional dialects, loanwords and mixed-language terms	Organisational intent catalogue	Confidence threshold
4. Identity and access	Verify user permissions	Arabic authentication prompts	Role-based access control	Data protection and audit logging
5. Query routing	Direct enquiry to correct knowledge domain	Interpret culturally specific phrasing	Connect to relevant department or service	Prevent unauthorised retrieval
6. Knowledge retrieval	Search authorised information	Arabic metadata, synonyms and cross-script terms	Validated policies, procedures and records	Source whitelisting

7. Response validation	Check answer quality	Ensure linguistic and cultural suitability	Confirm source, jurisdiction and update date	Confidence and applicability check
8. Response generation	Provide spoken answer	Preferred dialect, MSA, polite register and accessible speed	Approved organisational wording	Avoid unsupported claims
9. Human escalation	Transfer uncertain cases	Explain transfer clearly in Arabic	Connect to responsible employee	Human oversight
10. Feedback and maintenance	Record outcome and update knowledge	Capture dialect-specific failure patterns	Content review and knowledge updating	Monitoring and periodic audit

6.9 Scope and Status of the Proposed AIVA–KBS Framework

The model and integration map that is introduced in this thesis is a proposed model of the future as opposed to an actual functioning system. The research did not design, deploy or technologically test an operational AI voice assistant. Neither did it build nor tie an AIVA to live organisational knowledge-based system. Hence, no direct measures were undertaken on the accuracy of speech-recognition, dialect recognition, code-switching strength, correctness of responses, latency, retrieval accuracy, execution of tasks, access control, workflow coordination or organisational service results.

The proposed model was based on the reported experiences and perceptions of respondents, Round 1 exploratory results, Round 2 structural model, and literature on the Arabic language processing, culture adaptation, privacy, governance and knowledge-based systems. It must thus be understood as analytically informed conceptual architecture which recognizes requirements and relationships to be implemented in the future. It fails to offer support that the suggested architecture will be effective in automatically enhancing the performance of AI, efficiency of the organisation, quality of services to the population, or user outcomes.

Further study will be necessary to transform this suggested framework into a computer implementation, tie it to an approved organisational repository of knowledge, and test it with controlled technical assessment and usability experiments and an institutional pilot.

6.10 Future work

Future research should broaden the sampling frame to capture dialectal, geographic, and accessibility diversity across the Arab region. Beyond urban, digitally fluent populations, studies ought to purposively recruit rural residents, older adults, low-literacy users, and people with speech or motor impairments, using adaptive instruments (audio prompts, larger fonts, assisted consent). Stratified samples can then support subgroup analyses of error profiles and task success by dialect and capability. This responds to recent evidence that speech/NLP systems exhibit uneven performance across accents and demographics, which can translate into practical exclusion if unmeasured (Koenecke et al., 2020). Panel designs powered for subgroup contrasts would make these disparities visible and tractable.

Methodologically, the survey design should be complemented with multi-method fieldwork that reduces self-report bias and uncovers interactional details. In-situ observation, think-aloud usability sessions, diary studies, and ethnographic interviews can document how users actually speak to assistants, when they abandon voice for touch, and how families or coworkers collaboratively repair recognition errors. Mixed-method human–AI interaction research recommends graceful repair and legible system status; field data can specify what these look like in Arabic contexts (Amershi et al., 2019). For reproducibility, future projects should preregister analysis plans where feasible and share de-identified instruments and code.

Qualitative investigation should also include semi-structured interviews and focus groups with Arabic-speaking users, policymakers, public-sector managers, software developers, knowledge-management specialists, and data-protection professionals. Interviews with users can explore trust, preferred dialects, cultural expectations, privacy concerns, and experiences of failed or successful interaction. Interviews with policymakers and organisational stakeholders can clarify procurement requirements, accountability arrangements, data-governance obligations, content-authorisation processes, and acceptable conditions for human oversight. These perspectives would help refine both the technical architecture and the governance model before real-world implementation.

Comparative and longitudinal designs are also overdue. A multi-site program spanning GCC, Levant, Egypt, and the Maghreb would allow tests of whether the same assistant policy like register switching between MSA and vernacular yields different outcomes by region and service domain (utilities, health, education). Longitudinal panels can track how attitudes and usage evolve as assistants improve (as like, after a dialect-specific ASR update) and as institutions change content governance. Such designs align with international guidance urging continuous monitoring and documented improvement cycles in AI deployments (NIST, 2023; UNESCO, 2021).

On the technical side, future work should move beyond generic "Arabic support" toward dialect-and code-switch-robust pipelines. Promising directions include: (i) fine-tuning self-supervised speech encoders such as XLS-R on Gulf/Levantine/Egyptian/Maghrebi audio to lower ASR word-error rates; (ii) leveraging weakly supervised models like Whisper while instituting human-in-the-loop review to mitigate hallucinations in sensitive answers; (iii) integrating diacritization for clearer TTS and name pronunciation; and (iv) adopting multilingual NLU (such as, XLM-R, mT5) that handles mixed-script entities and loanwords (Babu et al., 2021; Radford et al., 2022; Conneau et al., 2019; Xue et al., 2021). Critically, research should report dialect-stratified metrics and stress-tests on code-switch corpora, using community benchmarks (such as, MADAR, NADI) as external checks.

Sector-specific evaluations would deepen practical relevance. In healthcare, for example, studies could compare AIVA-supported triage or appointment scheduling against existing channels on task completion, time-to-resolution, and recontact rates, while auditing safety (correct medication names, unambiguous dates) and equity (no dialect group disproportionately fails recognition). In education, experiments could test whether voice-first access to learning resources improves study adherence for learners with reading difficulties, following accessibility findings that confirmations and read-back features reduce cognitive load (Pradhan et al., 2019). In digital government, randomized pilots can assess whether provenance cues (source and last-updated timestamp) increase trust and compliance relative to generic answers—an operationalization of transparency recommended by UNESCO (2021).

A real-world pilot is then to be carried out in an organisation in the UAE that is in the public sector like a municipality, utility company, university, healthcare company or a government service centre. The test pilot is supposed to attach the prototype to a supervised and approved body of knowledge and test performance in the real service environment. The outcomes that would be relevant would be the retrieval accuracy, completion of tasks, response time, rates of escalation, user satisfaction, accessibility, staff workload and the uniformity in information delivered among the Arabic dialect groups. Before and after or controlled comparative design would be useful in determining whether or not the prototype helps in service delivery and no assumption that improvement will automatically be realised in the organisation should be made.

The AIVA-KBS framework proposed should be compared with other Arabic voice-assistants studies and implementations in future research. As an example, Salah (2023) showcased good Arabic command-recognition results with the Arabic Personalized Control of Things system, yet the analysis was based primarily on the predetermined home-automation tasks and technical recognition results. Likewise, Elmedany (2024) wrote about the implementation of the smart helper Ali in the government services in Bahrain, whereas Al Shamsi (2021) addressed the issue of AIVA use among students of the UAE university. These studies present valuable evidence, though they are rather limited to technical, institutional or user contexts.

Finally, governance-oriented research should prototype and evaluate operational safeguards. This includes privacy-by-design experiments (on-device processing for short commands; federated learning for personalization without centralizing raw audio), Model Cards and Datasheets tailored to speech/NLP that disclose dialect coverage and known failure modes, and redress mechanisms that route low-confidence or contested outputs to humans (Mitchell et al., 2019; Gebru et al., 2021; Kairouz et al., 2021). Given UAE PDPL requirements on purpose limitation and retention, future studies should compare user comprehension and trust under alternative consent and retention UIs written in plain-language Arabic. Together, these lines of inquiry would advance AIVAs from generic tools to accountable, culturally adept knowledge mediators across Arabic-speaking societies.

6.11 Chapter summary

This chapter consolidated the thesis's academic and practical contributions, acknowledged its limits, and set a focused agenda for subsequent studies. Evidence from two UAE-wide surveys showed that cultural–linguistic fit more than demographics or technical self-efficacy shaped how important users perceived AIVAs to be. The practical guidance targets developers, policymakers, and organizations to turn those insights into deployable systems and accountable governance. Future work should broaden samples, add longitudinal and mixed methods, establish dialect-aware benchmarks, and validate outcomes in real services. Together, these steps can produce Arabic-capable AIVAs that are accurate, trustworthy, and genuinely useful.

REFERENCES

Abdelraouf, E. (2024). Examining AI Integration by Audio-Visual Media Platforms in Oman: A Qualitative Analysis of Media Professionals' Perspectives. *Arab Media & Society*, (37).

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Adetayo, A. J., Babalola, O. O., & Olukayode, V. A. (2025). The Convergence of AI and Librarianship: Evaluating the Real-World Application of ChatGPT, Gemini, and Copilot. In *Real-World Applications of AI Innovation* (pp. 407-424). IGI Global Scientific Publishing.

AI, N. (2023). Artificial intelligence risk management framework (AI RMF 1.0).

Al Ali, F. A., & Khalil, E. (2025). The Use of Artificial Intelligence in Government Communication in the UAE. *practice*, *7*(02).

Al Shamsi, J. H. (2021). *Students behaviour toward voice assistant technology in the UAE* (Master's thesis, The British University in Dubai).

AlAfnan, M. A. (2025). Artificial Intelligence and language: Bridging Arabic and English with technology. *Journal of Ecohumanism*, *4*(1), 240-256.

Albirini, A. (2016). *Modern Arabic sociolinguistics: Diglossia, variation, codeswitching, attitudes and identity*. Routledge.

Alford, J. (2002). Defining the client in the public sector: A social-exchange perspective. *Public administration review*, *62*(3), 337-346.

Al-Fraihat, D., Alzaidi, M., & Joy, M. (2023). Why do consumers adopt smart voice assistants for shopping purposes? A perspective from complexity theory. *Intelligent Systems with Applications*, *18*, 200230.

Ali, A., Vogel, S., & Renals, S. (2017, December). Speech recognition challenge in the wild: Arabic MGB-3. In *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* (pp. 316-322). IEEE.

Alkarkhi, D. A., Al-Majedy, Z., & Abbas, M. R. (2024). Development of an Arabic-language virtual assistant for public services to improve accessibility of government services for Iraqis. International Journal of Artificial Intelligence.

Alshehhi, K., Cheaitou, A., & Rashid, H. (2024). Procurement of Artificial Intelligence Systems in UAE Public Sectors: An Interpretive Structural Modeling of Critical Success Factors. *Sustainability*, *16*(17), 7724.

Alsulami, M., & Mehmood, R. (2018, March). Sentiment analysis model for arabic tweets to detect users' opinions about government services in Saudi Arabia: ministry of education as a case study. In *Al Yamamah Information and Communication Technology Forum* (pp. 1-8). Riyadh: Al Yamamah Information and Communication Technology Forum.

Alwash, L. M. (2025). Pragmatic Differences in Standard Arabic–English Translation: Implications for Translators with Special Reference to Other Arabic Local Languages: Case study. *Al-Satil Journal*, *20*(44), 1-22.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019, May). Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1-13).

Antoun, R. T. (1968). On the Modesty of Women in Arab Muslim Villages: A Study in the Accommodation of Traditions 1. *American Anthropologist*, *70*(4), 671-697.

Antoun, W., Baly, F., & Hajj, H. (2020). Arabert: Transformer-based model for arabic language understanding. *arXiv preprint arXiv:2003.00104*.

Anwar, M. R., & Ahyarudin, H. A. (2023). Ai-powered arabic language education in the era of society 5.0. *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, *5*(1), 50-57. *ARABnet | Arabic speaking chatbot 'Arabot' raises $1M in seed funding, eyes regional expansion*.

Babu, A., Wang, C., Tjandra, A., Lakhotia, K., Xu, Q., Goyal, N., ... & Auli, M. (2021). XLS-R: Self-supervised cross-lingual speech representation learning at scale. *arXiv preprint arXiv:2111.09296*.

Badran, E. and Seker, H. (2024) 'Studying Factors Affecting Attitudes on Artificial Intelligence Voice Assistants in the United Arab Emirates', *Proceedings of the ISICS2024 International Conference on Intelligent Computing Systems*.

Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., ... & Kaplan, J. (2022). Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*.

Ballout, S. (2024). People of Arab Heritage. In *Handbook for Culturally Competent Care* (pp. 97-137). Cham: Springer International Publishing.

Bella, G., Helm, P., Koch, G., & Giunchiglia, F. (2024). Tackling language modelling bias in support of linguistic diversity. ACM Digital Library.

Benjamin, M. (2025). The Evolution of Voice Assistants in Sales Support.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, *88*(3), 588.

Besdouri, F. Z., Zribi, I., & Belguith, L. H. (2024). Arabic automatic speech recognition: challenges and progress. *Speech Communication*, *163*, 103110.

Bhatt, S., Dev, S., Dossou, B., Ghosal, T., Haddad, H., Lepp, H., ... & Xu, W. (2022, December). Proceedings of the Sixth Widening NLP Workshop (WiNLP). In *Proceedings of the Sixth Widening NLP Workshop (WiNLP)*.

Blum-Kulka, S., House, J., & Kasper, G. (1989). Cross-cultural pragmatics: Requests and apologies.

Bollen, K. A. (1990). Overall fit in covariance structure models: Two types of sample size effects. *Psychological bulletin*, *107*(2), 256.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2022). *On the opportunities and risks of foundation models (arXiv: 2108.07258). arXiv*.

Bouamor, H., Habash, N., Salameh, M., Zaghouani, W., Rambow, O., Abdulrahim, D., ... & Oflazer, K. (2018, May). The MADAR Arabic dialect corpus and lexicon. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*.

Bouzid, M. (2023). The future of artificial intelligence in the Arab world The experience of some Arab countries. *IJEP*, *6*(1), 257-271.

Bozeman, B. (2019). Public values: Citizens' perspective. *Public Management Review*, *21*(6), 817-838.

Boztemir, Y., & Çalışkan, N. (2024). Analyzing and mitigating cultural hallucinations of commercial language models in Turkish. TechRxiv.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

Bureau, N. (2025, September 10). *Abu Dhabi ALC, Amazon Alexa partner to enrich Arabic content*. Emirati Times. https://emiratitimes.com/abu-dhabi-alc-amazon-alexa-partnership/

Byrne, B. M. (2013). *Structural equation modeling with EQS: Basic concepts, applications, and programming*. Routledge.

Cao, H. (2024). Interaction in virtual worlds: a comparative study of the virtual assistants Siri and XiaoIce.

Chen, Y., & Wei, L. (2024). Effects of artificial intelligence learning environments on engineers' intercultural communication competence: Mediating role of learning adaptability. *Interactive Learning Environments*.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., ... & Stoyanov, V. (2019). Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Dahabreh, F. M. (2023). *The continued usage of artificial intelligence in the United Arab Emirates public sector organisations: An extended information system success model* (Doctoral dissertation, University of Northumbria at Newcastle (United Kingdom)).

Djanibekov, A., Toyin, H. O., Alshalan, R., Alitr, A., & Aldarmaki, H. (2024). Dialectal Coverage And Generalization in Arabic Speech Recognition. *arXiv preprint arXiv:2411.05872*.

Dorodnykh, N. O., Nikolaychuk, O. A., Stolbov, A. B., & Yurin, A. Y. (2022, November). Towards knowledge-based virtual assistant development with the aid of ontology transformations. In *2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)* (pp. 830-834). IEEE.

Dwivedi, A., Sahu, S., & Srivastava, A. (2024, December). AI-Powered Real-Time Text Editor with Multilingual Translation and Speech Recognition. In *2024 IEEE 16th*

International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 1335-1340). IEEE.

Egbemhenghe, A. U., Ojeyemi, T., Iwuozor, K. O., Emenike, E. C., Ogunsanya, T. I., Anidiobi, S. U., & Adeniyi, A. G. (2023). Revolutionizing water treatment, conservation, and management: Harnessing the power of AI-driven ChatGPT solutions. *Environmental Challenges*, *13*, 100782.

El Zahraa, F. (2025). Leveraging Artificial Intelligence and Digital Technologies to Enhance Sociolinguistic Competence and Arabic Language Skills. In *Proceeding of the International Conference on Religious Education and Cross-Cultural Understanding* (Vol. 1, No. 1, pp. 33-49).

Elmagarmid, A., & Saoudi, A. (2021). Building a preeminent research lab in the Arab region: The case of QCRI. *Communications of the ACM*, *64*(4), 50-53.

Elmedany, W. M., AL-Abbasi, L. M., & Hewahi, N. M. (2024). Secure Intelligent Agent for Bahrain's e-Government Call Center. *Authorea Preprints*.

Elsheikh, A. (2023). *Enhancing the efficacy of assistive technologies through localization: A comprehensive analysis with a focus on the Arab region*. Nafath.

Expedia Group Inc. (2024) Romie.

Expedia Group. (2024, May 14). Put your trip on autopilot: Expedia Group introduces new innovations at EXPLORE to take the stress out of travel and enhance partner experience. *Expedia Group News Releases*.

Fang, C. M., Samaradivakara, Y., Maes, P., & Nanayakkara, S. (2025, April). Mirai: A Wearable Proactive AI" Inner-Voice" for Contextual Nudging. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1-9).

Ferguson, C. A. (1959). Diglossia. *word*, *15*(2), 325-340.

Fernandez, K. J. (2020, May 19). *Why making Arabic voice apps is so tricky | WIRED Middle East*. WIRED Middle East.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, *18*(1), 39-50.

Freire, S. K., Panicker, S. S., Ruiz-Arenas, S., Rusák, Z., & Niforatos, E. (2022). A cognitive assistant for operators: Ai-powered knowledge sharing on complex systems. *IEEE Pervasive Computing*, *22*(1), 50-58.

Gallegos, I. O., Rossi, R. A., Barrow, J., & Tanjim, M. M. (2024). Bias and fairness in large language models: A survey. MIT Press.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, *64*(12), 86-92.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. *Contexts of accommodation: Developments in applied sociolinguistics*, *1*, 1-68.

Gnana Rajesh, D., Tamilarasi, G., & Khan, M. E. (2023). Voice and text-based virtual assistant for academic advising using knowledge-based intelligent decision support expert system. In *Sentiment Analysis and Deep Learning: Proceedings of ICSADL 2022* (pp. 483-491). Singapore: Springer Nature Singapore.

Government of Dubai. (2021). *Dubai Electricity & Water Authority (DEWA) | DEWA 2021 strategy updated for government of the future and 4th Industrial Revolution. (2021). Dewa.gov.ae; DEWA.*

Grimsley, M., & Meehan, A. (2007). e-Government information systems: Evaluation-led design for public value and client trust. *European Journal of Information Systems*, *16*(2), 134-148.

Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California management review*, *61*(4), 5-14.

Harrat, S., Meftouh, K., & Smaïli, K. (2018). *Maghrebi Arabic dialect processing: An overview*. HAL Open Science.

Horesh, U., & Cotter, W. (2016). *Current research on linguistic variation in the Arabic-speaking world*. Wiley Online Library.

Huang, M. (2024). *Cross-cultural communication in the digital era: Insights from social media interactions*. EWADirect.

Huehnergard, J., 2017. Arabic in its Semitic context. In *Arabic in context* (pp. 3-34). Brill.

International Organization for Standardization. (2023). *ISO/IEC 42001:2023*. ISO. https://www.iso.org/standard/42001

Jaradat, G. A., Alzubaidi, M. A., & Otoom, M. (2022). A novel human-vehicle interaction assistive device for Arab drivers using speech recognition. *IEEE Access, 10*, 127514-127529.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and trends® in machine learning*, *14*(1–2), 1-210.

Kassis-Henderson, J., Cohen, L., & McCulloch, R. (2018). Boundary crossing and reflexivity: Navigating the complexity of cultural and linguistic identity. *Business and Professional Communication Quarterly*, *81*(3), 304-327.

Kendall, L., Chaudhuri, B., & Bhalla, A. (2020). Understanding technology as situated practice: everyday use of voice user interfaces among diverse groups of users in urban India. *Information Systems Frontiers*, *22*(3), 585-605.

Khader, M. M., Al-Sharafi, A., Al-Sioufy, M. H., Zaghouani, W., & Al-Zawqari, A. (2024, May). Munazarat 1.0: A corpus of Arabic competitive debates. In *Proceedings of the 6th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT) with Shared Tasks on Arabic LLMs Hallucination and Dialect to MSA Machine Translation@ LREC-COLING 2024* (pp. 20-30).

Khaleej Times. (2023). *Dubai: Dewa announces pilot use of ChatGPT to boost capabilities of its virtual employee*. (2023). Khaleej Times. https://www.khaleejtimes.com/business/tech/dubai-dewa-announces-pilot-use-of-chatgpt-to-boost-capabilities-of-its-virtual-employee

Khalilia, M., Malaysha, S., Suwaileh, R., Jarrar, M., Aljabari, A., Elsayed, T., & Zitouni, I. (2024). Arabicnlu 2024: The first arabic natural language understanding shared task. *arXiv preprint arXiv:2407.20663*.

Khan, I., Rajput, J., & Jeyasekar, A. (2025). AI based Voice Assistant System. *MASK International Journal of Science and Technology*, *4*(1).

Kheir, Y. E., Khnaisser, F., Chowdhury, S. A., Mubarak, H., Afzal, S., & Ali, A. (2023). QVoice: Arabic Speech Pronunciation Learning Application. *arXiv preprint arXiv:2305.07445*.

Kim, D. (2020). Learning language, learning culture: Teaching language to the whole student. *ECNU Review of Education*, *3*(3), 519-541.

Kingaby, S. A. (2021). *Data-Driven Alexa Skills*. Apress.

Kline, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford publications.

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., ... & Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the national academy of sciences*, *117*(14), 7684-7689.

Kumar, A., Bala, P. K., Chakraborty, S., & Behera, R. K. (2024). Exploring antecedents impacting user satisfaction with voice assistant app: A text mining-based analysis on Alexa services. *Journal of Retailing and Consumer Services*, *76*, 103586.

Kumar, V., Dixit, A., Javalgi, R. R. G., & Dass, M. (2016). Research framework, strategies, and applications of intelligent agent technologies (IATs) in marketing. *Journal of the Academy of Marketing Science*, *44*(1), 24-45.

Leblebici, D. (2024). “You are Apple, why are you speaking to me in Turkish?”: the role of English in voice assistant interactions. *Multilingua*, *43*(4), 455-485.

Leonidou, E., & Kypros, Z. (2024). Consumer Trust and Voice Marketing in the Age of AI and Robotics. *Technology, Employment, Change Management and Human Well-Being (TECH)*.

Leopard, A. S. (2025, August 14). *Global AI Governance: Five Key Frameworks Explained*. Bradley.com.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, *33*, 9459-9474.

Li, X. (2023). “There’s no data like more data” automatic speech recognition and the making of algorithmic culture. *Osiris*, *38*(1), 165-182.

Litt, E., Zhao, S., Kraut, R., & Burke, M. (2020). *What are meaningful social interactions in today’s media landscape? A cross-cultural survey. Social Media + Society*, **6**(3), 1-14.

Luger, E., & Sellen, A. (2016, May). " Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of* *the 2016 CHI conference on human factors in computing systems* (pp. 5286-5297).

Madhi, Z. K. (2025). The Role of Digital Language in Shaping Social Identity among Arab Youth: ALinguistic Analysis of Social Media Discourse. *Journal of The Iraqi University*, *73*(1).

Mahlknecht, A., Engl, A., Piccoliori, G., & Wiedermann, C. J. (2023). Supporting primary care through symptom checking artificial intelligence: a study of patient and physician attitudes in Italian general practice. *BMC Primary Care*, *24*(1), 174.

Mahmoud Alhilal, D. (2024). The Linguistic Dichotomy: A Comparative Study of Standard Arabic and Colloquial Arabic. *10*(19), 347-367.

Mahmudulhassan, M. (2024). Exploring the essence, importance, and distinctive attributes of Islamic culture: An in-depth cultural analysis. *Bulletin of Islamic Research*, *2*(2), 311-326.

Manolescu, V. D., AlZu'bi, H., & Secco, E. L. (2025). Interactive Conversational AI with IoT Devices for Enhanced Human-Robot Interaction. *Journal of Intelligent Communication*.

Marconi, U. G. VOICE-FIRST: Supporting Human Assistants with Real-time Voice Understanding.

May, S. (2013). *Language and minority rights: Ethnicity, nationalism and the politics of language*. Routledge.

McCormack, L., Bendechache, M., Lewis, D., & Huyskes, D. (2025). Trust and transparency in AI: industry voices on data, ethics, and compliance. *AI & SOCIETY*, 1-29.

McCracken, K., & Hogan, S. S. (2021). Community First: Indigenous Community-Based Archival Provenance. *Across the Disciplines*, *18*(22–32).

McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa… examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in human behavior*, *99*, 28-37.

Miao, X., & McLoughlin, I. (2019). Lstm-tdnn with convolutional front-end for dialect identification in the 2019 multi-genre broadcast challenge. *arXiv preprint arXiv:1912.09003*.

Mirsagatova, M. A. (2023). Semantics Of Words Expressing Respect In Arabic. *European International Journal of Philological Sciences*, *3*(11), 32-38.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019, January). Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 220-229).

Mohan, V., & Paramesh, A. S. (2024, February). Knowledge based organizational personal assistant. In *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)* (pp. 1-7). IEEE.

Monroy-Osorio, J. C. (2024). Assessing the impact of digital service innovation (DSI) on business performance: The mediating effect of artificial intelligence (AI). *Journal of Enterprise Information Management*.

Muminovna, M. M. (2024). Importance Of Linguistics. *Texas Journal of Philology, Culture and History*, *29*, 16-19.

Nasr, S., Duwairi, R., & Quwaider, M. (2023). End-to-end speech recognition for arabic dialects. *Arabian Journal for Science and Engineering*, *48*(8), 10617-10633.

Navigli, R., Conia, S., & Ross, B. (2023). Biases in large language models: Origins, inventory, and discussion. ACM Digital Library.

NIST. (2023). AI Risk Management Framework. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, *1*(1).

Noe, K. K. (2021). *A Framework for Equitable Design in Extended Reality Cultural Heritage Exhibitions* (Master's thesis, University of Hawai'i at Manoa).

OECD. (2019). *The OECD Artificial Intelligence (AI) Principles*. Oecd.ai; OECD. https://oecd.ai/en/ai-principles

Olawade, D. B., Bolarinwa, O. A., Adebisi, Y. A., & Shongwe, S. (2025). The role of artificial intelligence in enhancing healthcare for people with disabilities. *Social Science & Medicine*, *364*, 117560.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, *35*, 27730-27744.

Pasupuleti, M. K. (2025). Human-in-the-Loop AI: Enhancing Transparency and Accountability. *International Journal of Academic and Industrial Research Innovations(IJAIRI)*, *05*(05), 574–585.

Pawar, A., Patil, N., Pawar, A. G., & Patil, N. V. (2024). Comparative study of techniques for spoken language dialect identification. *Int. J. Comput. Appl*, *186*(7), 975-8887.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, *88*(5), 879.

Pradhan, A., Lazar, A., & Findlater, L. (2020). Use of intelligent voice assistants by older adults with low technology use. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *27*(4), 1-27.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023, July). Robust speech recognition via large-scale weak supervision. In *International conference on machine learning* (pp. 28492-28518). PMLR.

Rahman, A., Kabir, M. M., Mridha, M. F., Alatiyyah, M., Alhasson, H. F., & Alharbi, S. S. (2024). Arabic speech recognition: Advancement and challenges. *IEEE Access*, *12*, 39689-39716.

Rath, S., Nagayach, O., Boddu, A., & Krishna, R. J. (2024, October). Multilingual Sentiment Analysis over Real-Time. In *Cyber Warfare, Security and Space Computing: Second International Conference on Cyber Warfare, Security and Space Computing, SpacSec 2024, Jaipur, India, February 22–23, 2024, Proceedings* (Vol. 2195, p. 238). Springer Nature.

Rawat, R., Chakrawarti, R.K., Sarangi, S.K., Rajavat, A., Alamanda, M.S., Srividya, K. and Sankaran, K.S. eds., (2024). *Conversational Artificial Intelligence*. John Wiley & Sons.

Raza, A. (2022). *Capitalizing knowledge using blockchain: multi blockchain based framework for rule-based knowledge base systems (KBS)* (Doctoral dissertation, University of Tasmania).

RoboMatic (2025). *Introduction*. Robomatic.ai.

Roig, A. (2017). Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR). *European Journal of Law and Technology*, *8*(3).

Roman, G. (2024). A Multifaceted Approach to Linguistic and Cultural Inclusion for Migrants and Refugees in Europe. In *Transformative Intercultural Global Education* (pp. 370-384). IGI Global Scientific Publishing.

Salah, A., Adel, G., Mohamed, H., Baghdady, Y., & Moussa, S. M. (2023). Towards personalized control of things using Arabic voice commands for elderly and with disabilities people. *International Journal of Information Technology*, 1-22.

Salehan, M., Kim, D. J., & Lee, J.-N. (2018). Are there any relationships between technology and cultural values? A country-level trend study of the association between information communication technology and cultural values. Telematics and Informatics, **35**(5), 1372-1388.

Salhab, M., & Harmanani, H. (2024). AraSpot: Arabic Spoken Command Spotting. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *23*(7), 1-12.

Salhab, M., Elghitany, M., Sait, S., Ullah, S. S., Abusheikh, M., & Abusheikh, H. (2025). Advancing Arabic Speech Recognition Through Large-Scale Weakly Supervised Learning. *arXiv preprint arXiv:2504.12254*.

Salih, L., Tarhini, A., & Acikgoz, F. (2025). AI-Enabled service continuance: roles of trust and privacy risk. *Journal of Computer Information Systems*, 1-16.

Scott, C. A. (Ed.). (2016). *Museums and public value: creating sustainable futures*. Routledge.

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press.

Shin, D. (2023). Embodying algorithms, enactive artificial intelligence and the extended cognition: You can see as much as you know about algorithm. *Journal of Information Science*, *49*(1), 18-31.

Siemund, P., Al-Issa, A., Rahbari, S., & Leimgruber, J. R. (2021). Multilingualism and the Role of English in the United Arab Emirates, with views from Singapore and Hong Kong. *Research developments in world Englishes*, 95-120.

Smairi, H. (2023). The digital synergy: Startups as cultural bridge and catalysts in the digital cross cultural communication experience.

Soofastaei, A. (Ed.). (2021). *Virtual Assistant*. BoD–Books on Demand.

SoundHound AI Inc. (2024, June 17). How Mercedes-Benz's MBUX voice control revolutionizes the user experience. *SoundHound Voice AI Blog*.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate behavioral research*, *25*(2), 173-180.

Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government information quarterly*, *36*(2), 368-383.

Sung, E. C., Bae, S., Han, D. I. D., & Kwon, O. (2021). Consumer engagement via interactive artificial intelligence and mixed reality. *International journal of information management*, *60*, 102382.

Susanto, C., & Hendarman, A. F. (2025). Bridging Cultural and Linguistic Gaps with AI-Based Knowledge Management: A Solution for Language and Cultural Barriers at PT. STI. *Eduvest-Journal of Universal Studies*, *5*(9), 11420-11434.

Team, F., Abbas, U., Ahmad, M. S., Alam, F., Altinisik, E., Asgari, E., ... & Ruan, C. (2025). Fanar: An arabic-centric multimodal generative ai platform. *arXiv preprint arXiv:2501.13944*.

Team, G., Anil, R., Borgeaud, S., Alayrac, J. B., Yu, J., Soricut, R., ... & Blanco, L. (2023). Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Thouvenin, F., Fruh, A., & Henseler, S. (2022). Article 22 GDPR on Automated Individual Decision-Making: Prohibition or Data Subject Right?. *Eur. Data Prot. L. Rev.*, *8*, 183.

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1-10.

UAE Government. (2023, November 7). *UAE Strategy for Artificial Intelligence - The Official Portal of the UAE Government*. U.ae.

UAE. (2024, May 13). *Data protection laws - The Official Portal of the UAE Government*. U.ae.

UNESCO. (2023, May 16). *Recommendation on the Ethics of Artificial Intelligence | UNESCO*. Www.unesco.org.

Vinti, C. (2021). An Evaluation of the "Designation" of Products, Sectors and Industry for Local Production and Content under the Preferential Procurement Policy Framework Act 5 of 2000. *Journal of Law, Society and Development*, *8*, 20-pages.

Waheed, A., Talafha, B., Sullivan, P., Elmadany, A., & Abdul-Mageed, M. (2023). Voxarabica: A robust dialect-aware arabic speech recognition system. *arXiv preprint arXiv:2310.11069*.

Wang, C., Teo, T. S., & Janssen, M. (2021). Public and private value creation using artificial intelligence: An empirical study of AI voice robot users in Chinese public sector. *International Journal of Information Management*, *61*, 102401.

Wirtz, B. W., Langer, P. F., & Fenner, C. (2021). Artificial intelligence in the public sector-a research agenda. *International Journal of Public Administration*, *44*(13), 1103-1128.

Xia, Y., Shin, S. Y., & Kim, J. C. (2024). Cross-cultural intelligent language learning system (CILS): Leveraging AI to facilitate language learning strategies in cross-cultural communication. *Applied Sciences, 14*(13), 5651.

Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., ... & Raffel, C. (2020). mT5: A massively multilingual pre-trained text-to-text transformer. *arXiv preprint arXiv:2010.11934*.

Zerilli, J., Bhatt, U., & Weller, A. (2022). How transparency modulates trust in artificial intelligence. *Patterns*, *3*(4).

Zhai, C., Wibowo, S., & Li, L. D. (2024). Evaluating the AI dialogue System's intercultural, humorous, and empathetic dimensions in English language learning: A case study. *Computers and Education: Artificial Intelligence*, *7*, 100262.

Zobeidi, L., & Si Merabet, M. (2023). *The Salient Factors That Contribute to Language Shift and Language Death* (Doctoral dissertation, Université IBN KHALDOUN-Tiaret)

Zou, B., Wang, C., Yan, Y., Du, X., & Ji, Y. (2025). Exploring English as a foreign language learners' adoption and utilisation of ChatGPT for speaking practice through an extended technology acceptance model. *International Journal of Applied Linguistics*, *35*(2), 689-704.

APPENDICES

Appendix A: Statistics of the Respondents

Statistics						
		Gender	**Age**	**Education**	**Employment**	**Work sector**
N	Valid	210	210	210	210	210
	Missing	0	0	0	0	0

Appendix B. Initial Exposure and Current Usage of AI Voice Assistants

Do you use or have you ever tried to use any AI voice assistants?			
		Frequency	**%**
Valid	No	13	6.2
	Yes	197	93.8
	Total	210	100.0
Do you currently use an AI voice assistant in Arabic?			
Valid	No	85	40.5
	Yes	125	59.5
	Total	210	100.0
If you answered yes to question 7, how often do you use the AI voice assistant?			
Valid	Daily	8	3.8
	Monthly	79	37.6
	Rarely	12	5.7
	Weekly	26	12.4
	.	85	40.5
	Total	210	100.0
If you answered yes to question 7, for how long have you been using the AI voice assistant?			

Valid	About 1 to 3 years	41	19.5
	About 3 to 5 years	73	34.8
	Less than 6 months	11	5.2
	.	85	40.5
	Total	210	100.0

Appendix C: Detailed Responses to "What Is the Primary Purpose for Using Your AI Voice Assistant?"

Table 0-1. Primary purposes of using AI voice assistant

Purpose for AI usage		Frequency	%
Valid	Entertainment (music, movies, etc.)	19	9.0
	Entertainment	1	.5
	Entertainment, home automation (controlling lights, thermostats, etc.)	1	.5
	Entertainment, home automation, shopping	12	5.7
	Entertainment, home automation, work purposes.	2	1.0
	Entertainment, home automation, work purposes, shopping	6	2.9
	Entertainment, information retrieval (web searching, getting news updates, etc.)	2	1.0
	Entertainment, information retrieval, home automat	1	.5

Entertainment, information retrieval, home automation	2	1.0
Entertainment, information retrieval, home automat	3	1.4
Entertainment, information retrieval, shopping.	7	3.3
Entertainment, information retrieval, work purpose	2	1.0
Entertainment, information retrieval, work purpose	6	2.9
Entertainment, productivity (setting reminders, scheduling events, etc.)	1	.5
Entertainment, productivity, home automation	1	.5
Entertainment, productivity, home automation	1	.5
Entertainment, productivity, home automation	4	1.9
Entertainment, productivity, information retrieval	1	.5
Entertainment, productivity, information retrieval	3	1.4
Entertainment, productivity, Information retrieval	1	.5
Entertainment, productivity, Information retrieval	2	1.0
Entertainment, productivity, shopping	6	2.9
Entertainment, productivity, work purposes, shopping	2	1.0
Entertainment, shopping	34	16.2
Entertainment, work purposes	6	2.9

Entertainment, work purposes, shopping	29	13.8
Home automation	1	.5
Home automation, shopping	1	.5
Information retrieval	12	5.7
Information retrieval, shopping	1	.5
Information retrieval, shopping	1	.5
Not applicable	1	.5
Not using	1	.5
Productivity	3	1.4
Productivity, home automation, shopping	1	.5
Productivity, home automation, work purposes	2	1.0
Productivity, information retrieval	4	1.9
Productivity, information retrieval	2	1.0
Productivity, information retrieval	1	.5
Productivity, shopping	6	2.9
Productivity, work purposes	1	.5
Productivity, work purposes., shopping	3	1.4
Research	1	.5
Shopping	6	2.9
Work purposes	4	1.9
Work purposes, shopping	3	1.4
Total	210	100

Table 0-2. Language Related Responses

In which languages have you used AI voice assistants? Select maximum 2			
	Language	**Frequency**	**%**
Valid	Arabic	5	2.4

	English	205	97.6
	Total	210	100.0

If you have not used an AI voice assistant, which languages would you prefer?

	Language		
Valid	Arabic	36	17.1
	Arabic, French	1	.5
	Arabic, German	3	1.4
	English	59	28.1
	English, Arabic	80	38.1
	English, Arabic, Chinese	1	.5
	English, Arabic, Chinese, French, Urdu	1	.5
	English, Arabic, French	4	1.9
	English, Arabic, German	1	.5
	English, Arabic, German, Chinese, French, Urdu	1	.5
	English, Arabic, German, Urdu	1	.5
	English, Arabic, Urdu	5	2.4
	English, Chinese	1	.5
	English, Chinese, French	2	1.0
	English, Chinese, French, Urdu	1	.5
	English, Chinese, Urdu	1	.5
	English, French	3	1.4
	English, German	3	1.4
	English, German, French	2	1.0

	English, Urdu	3	1.4
	Not applicable	1	.5
	Total	210	100.0

In which Areas do AI voice assistants in Arabic need an improvement?			
	Improvement Area		
Valid	All of the above	1	.5
	I do not speak Arabic, so I haven't used it	1	.5
	Seamless interaction with other devices & applications	22	10.5
	Not applicable	1	.5
	Providing accurate & reliable information	34	16.2
	Understanding context	15	7.1
	Understanding natural language & accents	136	64.8
	Total	***210***	***100.0***

Table 0-3. Challenges, benefits, technical knowledge and communication responses

In which Operating system (OS) have you used while using AI voice assistant in Arabic?

	OS	Frequency	%
Valid	Android	98	46.7
	Both	19	9.0
	iOS	93	44.3
	Total	210	100.0

In your opinion, what is the most beneficial aspect of using an AI voice assistant?			
	AI Benefit		
Valid	Ability to multitask	21	10.0
	Convenience	115	54.8
	Hands-free operation	32	15.2
	Time-saving	42	20.0
	Total	210	100.0
In your opinion, what is the most challenging aspect of using an AI voice assistant?			
	AI Challenge		
Valid	Difficulty in understanding commands	92	43.8
	Limited capabilities	47	22.4
	Privacy concerns	59	28.1
	Reliance on internet connectivity	12	5.7
	Total	210	100.0

Table 0-4: AI Voice Assistant Preferences and Devices

If you answered yes to question 7, which AI voice assistant(s) do you use?				
AI Voice Assistants	**Frequency**	**%**	**Valid %**	**Cumulative %**
Alexa by Amazon	3	1.4	1.4	1.4
Alexa, Bixby by Samsung	3	1.4	1.4	2.9
Alexa, Bixby, Blu (voice Bot by telecoms company Du)	2	1.0	1.0	3.8
Alexa, Bixby, RASHID - SMART CITY ASSISTANT	2	1.0	1.0	4.8
Alexa, Blu	5	2.4	2.4	7.1
Alexa, Cortana by MS	3	1.4	1.4	8.6
Alexa, Cortana, Bixby, Blu	1	.5	.5	9.0
Alexa, Cortana, Bixby, RafiQ (Understands colloquial Arabic)	1	.5	.5	9.5
Alexa, Cortana, Blu	3	1.4	1.4	11.0
Alexa, Cortana, RafiQ	1	.5	.5	11.4
Alexa, Cortana, RASHID	1	.5	.5	11.9
Alexa, Cortana, RASHID, Blu	2	1.0	1.0	12.9
Alexa, Cortana, Salma (Arabic AIVA by Mawdoo3)	1	.5	.5	13.3
Alexa, Cortana, Salma, RASHID	1	.5	.5	13.8
Alexa, Google Now	2	1.0	1.0	14.8

Alexa, Google Now, RASHID, Blu	1	.5	.5	15.2
Alexa, RafiQ	1	.5	.5	15.7
Alexa, RASHID	1	.5	.5	16.2
Alexa, RASHID, Blu	3	1.4	1.4	17.6
Alexa, Salma	4	1.9	1.9	19.5
Alexa, Salma, Blu	2	1.0	1.0	20.5
Alexa, Salma, RASHID	2	1.0	1.0	21.4
Alexa, Salma, RASHID	2	1.0	1.0	22.4
Alexa, Siri	8	3.8	3.8	26.2
Alexa, Siri, Bixby, Blu	1	.5	.5	26.7
Alexa, Siri, Bixby, RafiQ	1	.5	.5	27.1
Alexa, Siri, Bixby, RafiQ	1	.5	.5	27.6
Alexa, Siri, Bixby, RASHID, Blu	1	.5	.5	28.1
Alexa, Siri, Bixby, Salma	3	1.4	1.4	29.5
Alexa, Siri, Bixby, Salma	1	.5	.5	30.0
Alexa, Siri, Blu	17	8.1	8.1	38.1
Alexa, Siri, Cortana	6	2.9	2.9	41.0
Alexa, Siri, Cortana, Blu	1	.5	.5	41.4
Alexa, Siri, Cortana, RafiQ	2	1.0	1.0	42.4
Alexa, Siri, Cortana, RASHID	2	1.0	1.0	43.3
Alexa, Siri, Cortana, Salma	1	.5	.5	43.8
Alexa, Siri, Cortana, Salma	1	.5	.5	44.3
Alexa, Siri, Cortana, Salma	2	1.0	1.0	45.2
Alexa, Siri, Google Now	3	1.4	1.4	46.7
Alexa, Siri, Google Now, RASHID	2	1.0	1.0	47.6
Alexa, Google Now, RASHID, Blu	2	1.0	1.0	48.6

Alexa, Siri, Google Now, Salma	1	.5	.5	49.0
Alexa, Siri, RafiQ	7	3.3	3.3	52.4
Alexa, Siri, Rammas (Developed by DEWA)	1	.5	.5	52.9
Alexa, Siri, RASHID	4	1.9	1.9	54.8
Alexa, Siri, RASHID, Blu	14	6.7	6.7	61.4
Alexa, Siri, Salma, Blu	4	1.9	1.9	63.3
Alexa, Siri, Salma, RafiQ	3	1.4	1.4	64.8
Alexa, Siri, Salma, RafiQ	4	1.9	1.9	66.7
Alexa, Siri, RASHID	1	.5	.5	67.1
Alexa, Siri, Salma, RASHID	6	2.9	2.9	70.0
Bixby, Blu	3	1.4	1.4	71.4
Blu	1	.5	.5	71.9
Cortana, Blu	1	.5	.5	72.4
Google Now	9	4.3	4.3	76.7
Google Now, Bixby	2	1.0	1.0	77.6
Rammas	5	2.4	2.4	80.0
RASHID, Blu	1	.5	.5	80.5
Salma	3	1.4	1.4	81.9
Salma, RASHID, Blu	11	5.2	5.2	87.1
Siri	4	1.9	1.9	89.0
Siri, Blu	4	1.9	1.9	91.0
Siri, Cortana	2	1.0	1.0	91.9
Siri, Cortana, Google Now	7	3.3	3.3	95.2
Siri, Google Now, Bixby	6	2.9	2.9	98.1
Siri, RASHID	1	.5	.5	98.6
Siri, Salma, Blu	1	.5	.5	99.0
Siri, Salma, Rammas	2	1.0	1.0	100.0
Total	***210***	***10***	***100.0***	

Do you believe that AI voice assistants can effectively communicate with Arabic speakers?			
	Frequency		
Valid	No	8	3.8
	Not sure	17	8.1
	Yes	185	88.1
	Total	210	100.0

How important is it for AI voice assistants to be able to communicate in Arabic other than other languages (e.g., English)?			
Valid	Extremely important	175	83.3
	Neutral	3	1.4
	Somewhat important	32	15.2
	Total	210	100.0

Appendix D: Cultural differences response description

Cultural Differences		**Frequency**	**%**
Valid	Accessibility	43	20.5
	Age	3	1.4
	Analysis of the requirement	4	1.9
	Availability	2	1.0
	Communication	10	4.8
	Communication	2	1.0
	Communication	8	3.8
	Contents	1	.5
	Cultural awareness	1	.5
	Cultural sensitivities	1	.5
	Dialect variations	1	.5

Dialectics	1	.5
Dialectics	2	1.0
Dialects	1	.5
Diversity	1	.5
Gender roles	1	.5
Gender roles and bias	2	1.0
Gender roles and language	1	.5
Imagine saying (يعطيك العافية) to a Moroccan guy 😅	1	.5
Issues with understanding different Arabic accents	1	.5
Language	1	.5
Language	6	2.9
Language & communication	64	30.5
Language dialytic	1	.5
Languages	1	.5
Languages	1	.5
Limited capabilities	6	2.9
Localized content & services	1	.5
Arabic AIVA software poor performance	1	.5
Not understanding commands	1	.5
Perception of Time	1	.5
Policies	1	.5
Privacy	2	1.0
Privacy	1	.5

Privacy and trust	12	5.7
Privacy concern	3	1.4
Privacy data	1	.5
Security	1	.5
Security	7	3.3
Social norms and etiquette	8	3.8
Issues with understanding different Arabic accents	1	.5
No policies & regulation for using AI tools	1	.5
Users trust	1	.5
Total	210	100.0

If you answered yes to question 21, please describe the cultural differences you have experienced.

Table. Cultural differences identified

		Frequency	**%**
Valid	Ability to understand different accents	2	1.0
	Ability to understand Arabic accents & Arabic words	1	.5
	Accent	1	.5
	Accent & may be dialects	1	.5
	Accent & the Arabic context is not easy	1	.5
	Accessibility	2	1.0
	accessibility	1	.5
	Accessibility	2	1.0

Add English language	1	.5
Age limited	2	1.0
Age Limited	1	.5
Availability	8	3.8
Awareness	1	.5
Awareness	1	.5
Able to understand different dialects	1	.5
Big dataset for training	1	.5
Communication	3	1.4
Context of use, & commands with accent	1	.5
Continuous Improvement	1	.5
Cultural Sensitivity & Contextual Awareness	1	.5
Data Availability	2	1.0
Data Quality	1	.5
different dialects can be huge challenge	1	.5
difficult to used	1	.5
Environmentally friendly	1	.5
Fast response	1	.5
Government controlling	1	.5
Government policies	1	.5
Government policies and regulations	1	.5
Include different Arabic dialects	1	.5
Language	1	.5
Language	11	5.2
Languages	1	.5

Local dialects, some ancient words, & the conjugations of some linguistic verbs	1	.5
Loss of individual freedom and autonomy as a result of continuous machine monitoring. –	1	.5
More maturity needed	1	.5
More practice	1	.5
more training set in my opinion	1	.5
Na	1	.5
No Policies and regulation	1	.5
No policy & regulation implemented from the government	1	.5
Policies and regulation	1	.5
privacy	6	2.9
Privacy	73	34.8
Privacy and Age	1	.5
Privacy and Security	2	1.0
Privacy concern	2	1.0
privacy data	1	.5
Privacy data	1	.5
Rules	1	.5
Safety	1	.5
Security	30	14.3
Security and privacy	1	.5
Security information	1	.5

Semantic Understanding and Context	1	.5
Short introduction video in how and why to use	1	.5
Speed	1	.5
The accent	1	.5
The different dialects and pronunciation.	1	.5
The main challenges in improving AI voice assistants for Arabic-speaking applications include the complexity of the lang	1	.5
Arabic accents	1	.5
Trained on different Arabic language accents	1	.5
Translation services	1	.5
Understandable the deferred Arabic dialects	1	.5
Understanding	1	.5
Understanding accents	1	.5
Understanding accent	1	.5
Understanding different accents	1	.5
understanding different Arabic accents	1	.5
understanding spoken accents	1	.5
Understanding the different accents	1	.5

	User Engagement and Feedback	1	.5
	User Trust and Acceptance	2	1.0
	Users trust	2	1.0
	Using good Arabic Grammar (e.g. sentence structure, verb conjunctions	1	.5
	Voice Recognition	1	.5
	Total	210	100.0

In your opinion, what are some of the main challenges that need to be addressed in order to improve the functionality and accuracy of AI voice assistants in Arabic-speaking applications?

Table. Challenges identified by respondents for improvements in AIVAs

		Frequency	%
Valid	Dialect Recognition	1	.5
	Access	3	1.4
	Accessibility	4	1.9
	accurate information	1	.5
	adding more secure	1	.5
	More training for accurate datasets	1	.5
	Age	47	22.4
	Age limitation	1	.5
	Age limited	1	.5
	ANA	1	.5
	Applications services	1	.5
	Availability	1	.5

By improving the understanding of difference Arabic accents.	1	.5
By providing different dialects and pronunciation.	1	.5
centre data	1	.5
Centralized	1	.5
Centralized data	2	1.0
Centralized	1	.5
Clear policies	1	.5
communication	4	1.9
Communication	1	.5
compile and store enough data on Arabic content	1	.5
Content and Services	3	1.4
Contextual Understanding	5	2.4
Continuous Feedback and Iterative Improvement	1	.5
Cultural Awareness	4	1.9
Cultural Context Awareness	1	.5
Data Privacy	1	.5
Data Privacy and Security	5	2.4
Dialect and Accent Adaptation	1	.5
Dialect Support	2	1.0
Do more practical methods	1	.5
Doing multitasking	1	.5
Filter commands and offer alternative selections	1	.5
Gender	2	1.0

Gender and Addressing Options	4	1.9
Gender Roles	1	.5
Good service	1	.5
Government Policy and regulation	1	.5
I think most in services for old person as an personal assistant	1	.5
if it understands many Arabic dialects	1	.5
Implemented a clear government polices and regulation	1	.5
Improve language	1	.5
Insert street language	1	.5
Iterative Improvement	1	.5
Iterative Improvements	1	.5
language	1	.5
Language	2	1.0
Language and Dialect Recognition	3	1.4
Large Arabic dataset	1	.5
Localization	2	1.0
Localization and Cultural Sensitivity	1	.5
More	1	.5
More natural use	1	.5
more tailored	1	.5
More training on Arabic accents	1	.5

Most of the Arab speaking are in the delicate language and each country has its own wording Plus the Arabic content in t	1	.5
multitask	1	.5
N/A	1	.5
na	1	.5
Na	1	.5
No idea	1	.5
No ideas	1	.5
Not applicable	1	.5
Open sources	1	.5
Paying attention to the elderly who cannot read and write and suffer from visual problems that prevent them from using t	1	.5
Personalization and User Profiling	1	.5
Personalizing courses for students can do the same for teachers by analysing students' learning abilities and educational	1	.5
Privacy	1	.5
Privacy and Trust	2	1.0
Provide customized settings	1	.5
Safety	2	1.0
Save time and it offers convenience	1	.5

Security	17	8.1
security and privacy	1	.5
services	1	.5
Service providing	1	.5
services	10	4.8
Services	8	3.8
services and communication	1	.5
Short sentences	1	.5
Skills	1	.5
Main challenges in improving AIVAs for Arabic-speaking applications include language complexity	1	.5
Capable of recognizing different Accents	1	.5
Contextual in its design	1	.5
Trust	2	1.0
User Feedback	4	1.9
User Feedback and Iterative Improvements	1	.5
Users service	2	1.0
Users trust	1	.5
Using AI	1	.5
when its quickly accessible	1	.5
wider system	1	.5
Yes/No questions	1	.5
Total	210	100.0

Appendix E: In which area of IT your specialism is?

Table. Respondents' field of expertise

		Frequency	%	
Valid	Student	17	8.1	
	Administrator	2	1.0	
	Business analyst	2	1.0	
	CIS faculty	2	1.0	
	Cloud	1	.5	
	College student	18	8.6	8.6
	Cybersecurity	1	.5	
	DevOps	1	.5	
	Fintech	8	3.8	
	Infrastructure	1	.5	
	IS	1	.5	
	IT	1	.5	
	IT student	19	9.0	
	IT student	1	.5	
	IT Student	5	2.4	
	Linux	4	1.9	
	Networking & Security	2	1.0	
	Non	1	.5	
	Non-IT professional	1	.5	
	NA	1	.5	
	Non-IT professional	1	.5	
	Non-IT professional	1	.5	
	Non-IT professional	1	.5	
	Not working	1	.5	
	Programmer	1	.5	

	Security	1	.5
	Software Engineer	2	1.0
	Student	106	50.5
	Teaching	6	2.9
	Teaching	1	.5
	Total	**210**	**100.0**

Table. Respondents' technical knowledge

Technical knowledge		**Frequency**	**%**
Valid	18yrs of industry, academic and R&D in ICT solutions and Cybersecurity GRC	1	.5
	Academic	7	3.3
	IT project manager	7	3.3
	Analysis	8	3.8
	Analytics	6	2.9
	Application developer	10	4.8
	Application development major	3	1.4
	Applications development	1	.5
	Apps developer	1	.5
	AWS AND ORACLE OCI	1	.5
	Big Data	1	.5
	Big data analytics	1	.5
	Bigdata Analytics	3	1.4
	Business solution	1	.5
	Cloud analytics	1	.5
	Cloud computing analytics	2	1.0
	Cloud security	1	.5
	Computer Science & Software Engg	2	1.0

Cyber security	2	1.0
Cybersecurity, IOT, Networks, modelling & simulation, formal methods	1	.5
data analytics, text analytics	1	.5
Data collection	1	.5
developer	1	.5
Developer	2	1.0
Developer apps	2	1.0
Experience in developing voice recognition applications for user with disabilities	1	.5
Health care	1	.5
PhD software engineering with > 20 years of R&D experience.	1	.5
I don’t have	1	.5
ICT business analysis	1	.5
Information system & physical security	2	1.0
IS	4	1.9
IS analysis	17	8.1
IS Analytics	1	.5
IT	4	1.9
IT Desing	10	4.8
IT infrastructure	1	.5
ITA	1	.5
Java developer	2	1.0
Limited	1	.5
Major of Business solution	4	1.9
Medium	1	.5
Mobile apps	3	1.4

Mobile apps	3	1.4
Mobile apps analytics	2	1.0
Mobile apps developer	1	.5
Mobile developer	1	.5
My terminal degree is on CIS	2	1.0
N/A	1	.5
Na	4	1.9
NA	1	.5
Networking & security	2	1.0
Nil	1	.5
Non	1	.5
Non-IT	1	.5
Not applicable	1	.5
PhD	1	.5
Programmer	1	.5
Programmer	1	.5
Programing	2	1.0
programmer	1	.5
Programmer	1	.5
programmer, web developer, mobile apps, game developer, programming with LLM	1	.5
programming	1	.5
Programming	1	.5
security	11	5.2
Security	2	1.0
Security Major	4	1.9
Software	1	.5
software design	1	.5
Software developer	1	.5
Student	2	1.0

Student level	15	7.1
System Analysis	1	.5
System analytics	1	.5
System Analytics	1	.5
System Dependability Analysis	4	1.9
Teacher with long experience	1	.5
technical	1	.5
Technical	1	.5
technical analytics	5	2.4
University lecturer specialized in CS, teaching programming, web development & data base	1	.5
Very good level. teaching the topics	1	.5
Web design	1	.5
WEB developer	1	.5
Total	*210*	*10 0*

Table. Suggestions for AI developers

AI_Technical_KBS			
		Frequency	**%**
Valid	No	86	41.0
	Yes	124	59.0
	Total	*210*	*100.0*
AI Developers			
Valid	Extremely important	3	1.4
	NA	181	86.2
	Neutral	12	5.7
	Somewhat important	14	6.7
	Total	*210*	*100.0*

AI_Developer_KBS_Function			
Valid	Extremely important	3	1.4
	NA	186	88.6
	Neutral	14	6.7
	Somewhat important	7	3.3
	Total	*210*	*100.0*

Appendix F: First round of the survey.

I agree to participate in this research study: *

- Yes
- No

Section A: Demographic Survey

1. What is your gender? *

- Male
- Female
- Prefer not to answer

2. What is your age range? *

- ◯ 18-24
- ◯ 25-34
- ◯ 35-44
- ◯ 45-54
- ◯ 55 and above
- ◯ Prefer not to disclose

3. What is your highest level of education? *

- ◯ High school diploma or equivalent
- ◯ Bachelor's degree
- ◯ Master's degree
- ◯ Doctoral degree
- ◯ Other...

4. What is your employment status?

- ◯ Student
- ◯ Full- time employed
- ◯ Part-time employed
- ◯ Unemployed
- ◯ Retired
- ◯ Homemaker
- ◯ Self-employed
- ◯ Unable to work

5. In which sector you work. *

- ◯ Information Technology sector
- ◯ Real estate and development sector
- ◯ Food sector
- ◯ Education sector
- ◯ Banking Sector
- ◯ Health care Sector
- ◯ Government sector
- ◯ Transportation sector
- ◯ Other, (specify please).

Section B: Voice Assistant Usage

6. Do you use or have you ever tried to use any AI voice assistants? *

- ◯ Yes
- ◯ No

7. Do you currently use an AI voice assistants in **Arabic**? *

- ◯ Yes
- ◯ No, please go to the question 10

8. If you answered **yes** to question 7, how often do you use the AI voice assistants in **Arabic**?

- ◯ Daily
- ◯ Weekly
- ◯ Monthly
- ◯ Rarely

9. If you answered **yes** to question 7, for how long have you been using the AI voice assistant(s) in **Arabic**?

- ◯ Less than 6 months
- ◯ 6 months to 1 year
- ◯ About 1 to 3 years
- ◯ About 3 to 5 years
- ◯ 5 years or more

10. Which AI voice assistant(s) do you use? *

(Select all that apply)

- [ ] Alexa by Amazon
- [ ] Siri by Apple
- [ ] Cortana by Microsoft
- [ ] Google Now by Google
- [ ] Bixby by Samsung
- [ ] Salma (The first Arabic personal digital assistant developed by Mawdoo3)
- [ ] RafiQ (is a virtual assistant who can understand colloquial Arabic voice commands with localize...

11. What devices do you normally use for AI voice assistant(s)? *

(Select all that apply)

- [ ] Mobile phone Apps.
- [ ] Smart Speaker (e. g. Google Home, Amazon Echo)
- [ ] Laptop or computer
- [ ] Smart watch
- [ ] Smart TV
- [ ] Other...

12. What is the primary purpose for using your AI voice assistants?

(Select all that apply)

- [ ] Entertainment (music, movies, etc.)
- [ ] Productivity (setting reminders, scheduling events, etc.)
- [ ] Information retrieval (searching the web, getting news updates, etc.)
- [ ] Home automation (controlling lights, thermostats, etc.)
- [ ] Work purposes.
- [ ] Shopping.
- [ ] Other...

13. In which languages have you used AI voice assistants? Select maximum 2 *

(1 is the most frequent , 2 is the least used

	English	Arabic	German	Chinese	French	Urdu
1	○	○	○	○	○	○
2	○	○	○	○	○	○

14. If you have not used an AI voice assistants, which languages would you prefer? *

- [ ] English
- [ ] Arabic
- [ ] German
- [ ] Chinese
- [ ] French
- [ ] Urdu
- [ ] Other...

15. In which Area do AI voice assistants in **Arabic** need an improvement?

- ◯ Understanding natural language and accents
- ◯ Providing accurate and reliable information
- ◯ Interacting with other devices and applications seamlessly
- ◯ Understanding context and intent
- ◯ Other...

16. In which Operating system have you used while using AI voice assistants in **Arabic**? *

- ◯ iOS
- ◯ Android
- ◯ Both

Section C: Specific Aspect of AI Voice Assistant(s)

17. In your opinion, what is the most beneficial aspect of using an AI voice assistants? *

- ○ Convenience
- ○ Time-saving
- ○ Hands-free operation
- ○ Ability to multitask
- ○ Other...

18. In your opinion, what is the most challenging aspect of using an AI voice assistants? *

- ○ Difficulty in understanding commands
- ○ Limited capabilities
- ○ Privacy concerns
- ○ Reliance on internet connectivity
- ○ Other...

19. Do you believe that AI voice assistants can effectively communicate with **Arabic speakers**? *

- ○ Yes
- ○ No
- ○ Not sure

20. How important is it for AI voice assistants to be able to communicate in **Arabic** other than other languages (e.g., English)? *

- ◯ Extremely important
- ◯ Somewhat important
- ◯ Neutral
- ◯ Not very important
- ◯ Not important at all

21. Have you noticed any cultural barriers when using AI voice assistants in **Arabic**? *

- ◯ Yes
- ◯ No, please go to the question 23

22. If you answered **yes** to question 21, please describe the cultural barriers you have experienced .

Short answer text

Section D: Challenges of AI voice assistant while using in Arabic specking application.

23. In your opinion, what are some of the main challenges that need to be addressed in order to improve the functionality and accuracy of AI voice assistants in **Arabic-speaking applications**? *

Short answer text

24. How do you think AI voice assistants could be better tailored to meet the needs and preferences of **<u>Arabic-speaking users</u>**? *

Short answer text

25. Are there any specific features or capabilities that you would like to see in future AI voice assistants designed for **<u>Arabic-speaking users</u>**. *

Short answer text

Section E: IT specific field of AI voice assistants

If you are IT Professional, please answering the below questions otherwise submit the survey.

26. In which area of IT your specialism is?

- ◯ Software Engineer
- ◯ Cloud
- ◯ Linux
- ◯ DevOps
- ◯ Other...

27. Please describe your technical knowledge or background below.

Short answer text

28. Are you familiar with the technical aspects of AI voice assistants and their underlying knowledge-based systems?

- ◯ Yes
- ◯ No

29. In your opinion, how important is it for developers of AI voice assistants to have a deep understanding of ***Arabic language*** and culture?

- ◯ Extremely important
- ◯ Somewhat important
- ◯ Neutral
- ◯ Not very important
- ◯ Not important at all
- ◯ Other...

30. How important is it for developers of AI voice assistants to have a deep understanding of the knowledge-based systems that power their functionality?

- ◯ Extremely important
- ◯ Somewhat important
- ◯ Neutral
- ◯ Not very important
- ◯ Not important at all
- ◯ Other...

Appendix G: Second round of the Survey.

Demographic information

Q1) What is your gender?

A. Male
B. Female

Q2) What is your age range?

A. 18-24
B. 25-34
C. 35-44
D. 45-54
E. 55 and above

Q3) What is your highest level of Education?

A. High school
B. Bachelor's degree
C. Master's degree
D. Doctoral degree

Q4) In which sector you work?

A. Government
B. Private
C. Health care
D. Education

1. Indicate (✓) your level of agreement with the following statements.

SD: Strongly Disagree, **DA:** Disagree, **N:** Neutral, **A:** Agree, **SA:** Strongly Agree

Sr. No	Statement	SD	DA	N	A	SA
	Effective usage of AI Voice Assistant					
Q1	AI voice assistants is an effective tool to communicate with the government.					
Q2	I have used AI voice assistants to verify information with the government.					

Q3	The AI voice assistants facilitate the direct communication between the government & me.					
Q4	I have great dialogs with the government in AI voice assistants.					
	Perceived challenges:					
Q1	Using the AIVA enables me to perform tasks more quickly.					
Q2	Using the AIVA saves me time and effort in performing tasks.					
Q3	Using the AIVA improves my task performance.					
Q4	The AIVA is useful in performing my task.					
	Perceived benefits					
Q1	Spending time using the AIVA is exciting.					
Q2	Spending time using the AIVA is pleasant.					
Q3	Spending time using the AIVA is interesting.					
	Design Preferences					
Q1	Customizable language settings are important for AIVA					
Q2	AIVAs should support local dialects.					
Q3	The cultural relevance of responses is crucial for the adoption of AIVAs					
Q4	AIVAs should offer features tailored to the needs of the Arabic-speaking communities.					

	Communication in Government					
Q1	AI Voice Assistants provide accurate information during conversations					
Q2	AI Voice Assistants are useful for managing and organizing communication tasks					
Q3	Using AI Voice Assistants enhances my ability to communicate in different languages.					
Q4	I feel comfortable interacting with the AIVA since it generally fulfills its duties efficiently					
	Perceived public service					
Q1	Compared to the effort citizens need to put in, the use of the AI voice assistant is beneficial to me and other citizens.					
Q2	Compared to the time citizens need to spend, the use of the AI voice assistant is worthwhile to me and other citizens.					
Q3	Compared to the risk citizens need to take, the use of the AI voice assistant offers value for me and other citizens.					
Q4	Overall, the use of AI voice assistant gives good value to me and other citizens.					

Appendix H: Voice Assistants Tools

AI Tool	Platform	Key Features	Languages
Siri	Standard on all Apple devices Developed by Apple, Inc., in 2011	Voice-activated AI allows users to send text messages, make schedules & calls, play music & videos. Automation shortcuts Personalized with machine learning Secure & private	English (Australia, Canada, India, NZ, UK, US), Spanish, French, German, Italian, Japanese, Korean, Mandarin, Norwegian, Cantonese, Swedish, Danish, Dutch, Russian, Turkish, Thai, & Portuguese.
Bixby	Standard on all Samsung devices	Allows voice commands Bixby quick commands Use Bixby to translate Convert images to text Find photos Post photos on social media Find & send location	English (US, British, Indian) French, Korean, Chinese, German, Spanish, Italian, & Portuguese (Brazilian).
Google Assistant	iOS & Android Launched in 2012	Control electronic & smart home devices Access calendar & other personal information Gets online information from restaurant	English, French, Hindi, Hindi, Indonesian, Italian, Japanese, Korean, Portuguese (Brazilian),

		bookings to direction, weather, & news Play Chromecast or other devices Run timers & reminders	Spanish, Swedish, Vietnamese. .
Amazon Alexa	Amazon Apple store Google Play	Voice interaction, making to-do lists, setting alarms, streaming content Playing audiobooks Provides weather, sports, & other real-time information	English, French, German, Hindi, Italian, Japanese, Portuguese (Brazilian), Spanish Features
Cortana	Standard on all MS devices Apple store Google play Launched in 2014	Manage calendar & MS Teams meetings Create & manage lists Set reminders & alarms Find facts & information Open apps on computer	English, Portuguese, French, German, Italian, Spanish, Chinese, & Japanese
Facebook M	Developed by Facebook	Stickers, payments, polls, reminders, rides, and location sharing. (Slowly incorporated in Facebook Messenger)	English
Black-berry Assistant	Developed by Blackberry Limited /RIM	Set reminders launch apps Send BBM messages	Čeština, Deutsch, English, Español. Français, Hrvatski

		Search through e-mail & calendars Listen to recent e-mails Set the ringer to “phone calls only”	Magyar, Bahasa Indonesia Italian, Bahasa Malaya, Nederland, Polski, Português, Romana
Braina	Developed by Brainasoft	Dictation Custom Commands, replies & hotkeys. Play Songs and Videos Text to Speech Mathematics Search Information Dictionary and Thesaurus	African (South Africa), Indonesian, English (UK, US, Indian, New Zealand, Philippines), Spanish (Argentina, Chile, Colombia)
Teneo	Developed by Artificial Solutions	API Reporting/Analytics Activity Dashboard Multi- Natural Language Processing AI/Machine Learning Speech Recognition Natural Language Search	80 languages
Speak-toit Assistant	Developed by Speaktoit	Answers questions in natural language Setting reminders	(Cantonese & Mandarin) English French German Italian Japanese Korean Portuguese

		Customize services according to user preferences	(European & Brazilian) Russian Spanish
Hound	Apple store Google play	Get weather updates Makes calls & sends texts Find hotel that matches user preferences Address navigation/Uber booking Play music and games	English
DataBot	Windows Chrome web store Apple store Google play	Customize itself based on user preferences Provides daily horoscope Discover tests	Japanese, Chinese, Korean, Turkish, Polish, & Arabic
Salma	IOS & Android App	Shares weather forecasts Currency exchange rates Sets alarm & play music Makes call Answer general knowledge questions Gives price checks such as plane tickets. Also performs voice-based tasks	Arabic dialects (Egyptian & Saudi Arabic)

RafiQ	Google Play	Opening YouTube and Facebook & Sets alarm Booking Careem or Uber Pulling up menus of nearby restaurants	Egyptian Arabic
Rammas	IOS & Android App DEWA, first government organisation to launch it to serve users	Provides two ways of interaction, either by menu selection or direct questions to live agent Continue learning and understanding customers’ needs based on their queries Offer transactional and informational services	Arabic, English

Appendix I: Ethics approval

STAFFORDSHIRE UNIVERSITY

School of Digital, Technologies and Arts

ETHICAL APPROVAL FEEDBACK

Researcher name:	Eslam Badran
Title of Study:	SU_21_197 'Investigation of the role of AI Voice Assistants in Arabic Speaking citizen within the context of knoweldge based system.'
Award Pathway:	PGR
Status of approval:	Approved

Your project ***proposal has been approved*** by the Ethics Panel and you may commence the implementation phase of your study. You should note that any divergence from the approved procedures and research method will invalidate any insurance and liability cover from the University. You should, therefore, notify the Panel of any significant divergence from this approved proposal. This approval is only valid for as long as you are registered as a student at the University.

You should arrange to meet with your supervisor for support during the process of completing your study and writing your dissertation.

When your study is complete, please send the ethics committee an end of study report. A template can be found on the ethics BlackBoard site.

The Ethics Committee wish you well with your research.

Signed: **Date: 22.02.2023**

E. [signature]

Prof. Elhadj Benkhelifa

Chair of the Digital Technologies Ethics Panel

Appendix J: Conference paper contribution

ISICS2024 International Conference on Intelligent Computing Systems.

Studying Factors Affecting Attitudes on Artificial Intelligence Voice Assistants in the United Arab Emirates

Eslam Badran[1] and Huseyin Seker[2]

[1]Higher Colleges of Technology, Computer Information Science, UAE. ebadran@hct.ac.ae
[2]The University of Sharjah, College of Computing and Informatics, UAE. hseker@sharjah.ac.ae

Abstract: Artificial intelligence (AI) is an entangling force impacting the scope of businesses in areas such as health, finance, education, and even consumer technology. Against this background, we sought to explore the influence of demographic factors on attitudes towards AI among Arabic speakers. A total of 210 participants were selected from the Arab speaking community in the United Arab Emirates, followed by a hypothesis testing using an independent t-test of the data collected through the survey that contained 30 questions divided into five different categories including demographics, voice assistant usage, specific aspect of AI voice assistants, challenges of AI voice assistants and IT specific field of AI voice assistants. Results demonstrated that the variable of gender shows a significant influence on the perceived importance of AI, where females think AI to be more important than males. There was a significant difference: that is, those who did not perceive cultural barriers rated AI as more important than those who did. This outcome indicates that the cognitive representation of AI depends on the cultural context. Likewise, no significant difference was found across the question of whether technical knowledge is perceived to be a barrier, or not. This might imply that technical knowledge may raise awareness, though not necessarily related to being translated to a higher perceived importance of AI. The study sheds light on how demographic factors interact with each other in a complex way and puts emphasis on the importance of demographic-specific approaches in implementing AI technology to boost integration and acceptance effectively.

Keywords: Artificial Intelligence, Demographic Factors, Gender Differences, Cultural Impact, Technical knowledge

INVESTIGATING AI VOICE ASSISTANTS AMONG ARABIC-SPEAKING CITIZENS IN PUBLIC AND PRIVATE SECTORS: A QUANTITATIVE STUDY FROM THE UAE WITH STAKEHOLDER RECOMMENDATIONS

Eslam Badran[1] and Huseyin Seker[2]

ABSTRACT

AI voice assistants such as Google Assistant, Amazon Alexa, and Apple Siri are increasingly used to access information and complete tasks via natural language. This study examines adoption and impact among Arabic-speaking users in the United Arab Emirates (UAE) using a single-method quantitative survey ($N = 459$) and covariance-based SEM across six constructs: Effective Usage (EUAIVA), Perceived Challenges (PC), Perceived Benefits (PB), Design Preferences (DP), Communication in Government (CIG), and Perceived Importance of AI Services (PIAIS). Because the UAE has moved early on digital government and AI, it serves as a regional model for Arabic AI deployment in the Middle East. Effective usage positively predicts challenges, benefits, design preferences, and government communication; in turn, challenges, benefits, design preferences, and government communication significantly predict the perceived importance of AI services. The structural model shows excellent fit ($\chi^2/df = 1.175$, RMSEA = 0.020, SRMR = 0.027, NFI/IFI/TLI/CFI $\geq$ 0.95) and explains 61.8% of the variance in PIAIS. The contribution is a structured recommendations framework with stakeholder-specific guidance for AI developers, policymakers, and organizational users to deliver culturally and linguistically optimized Arabic AIVAs. Personalization, cultural compatibility, and linguistic flexibility are treated as foundational design principles for Arabic AIVAs and underpin the stakeholder recommendations proposed in this study.

KEYWORDS

AI voice assistants, Arabic-speaking societies, Recognized benefits and challenges, Personalization in AI

1. INTRODUCTION

AI voice assistants, such as Google Assistant, Amazon's Alexa, and Apple's Siri, have become a prime part of life since they facilitate information access, making the human-machine interaction easy [1][2]. Globally, these technologies have gained much momentum because of the potential they hold for the improvement of service delivery and enhancement of user experiences; yet uptake in Arabic-speaking societies, especially within the United Arab Emirates, has not been sufficiently examined. However, Arabic has special linguistic complexities in its culture and dialects [3], and that creates an obstacle for optimal deployment and usage of these digital tools [4]. While a lot of work has been done in the area of NLP, the voice assistants, in their attempts to adapt to the requirements of Arabs' linguistics and culture, find it challenging.

The limited deployment of AI voice assistants in the Arab world reflects both technical challenges, like the lack of linguistic support for unique features of the Arabic script, and the complexity of Arabic dialects that poses challenges to speech recognition [5]. Existing systems fail to fully support the Arabic dialects, cultural sensitivity, and local requirements, and thus can absorb limited benefits associated with AI-related advantages in terms of knowledge management and organizational or decision-making activities. Current literature focused on multilingual support for mainstream