Frontiers in Computer Science and Artificial Intelligence
DOI: 10.32996/fcsai
Journal Homepage: www.al-kindipublisher.com/index.php/fcsai

FCSAI
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Pronunciation Errors in Arabic YouTube Videos Narrated by AI

Reima Al-Jarf
Full Professor of English and Translation Studies, Riyadh, Saudi Arabia
Corresponding Author: Reima Al-Jarf**, E-mail**: reima.al.jarf@gmail.com

| ABSTRACT

Arabic has three long vowels /a:/, /u:/, /i:/ and three short vowel /a/, /u/, /i/ which are represented by diacritics marked over and under consonant letters. In words that have short vowels, only the consonants are written. Arabs usually read without any diacritics except for the Holy Quran and the Prophet's Hadiths. The absence of short vowel diacritics poses pronunciation difficulties for Artificial Intelligence (AI). Henceforth, this study aims to anlayze a sample of Arabic YouTube videos narrated by AI to find out which Arabic words are mispronounced by AI narrators. It was noted that AI narrators speak with a natural voice, good expression and intonation. They make no grammatical or syntactic errors. But they make pronunciation errors, especially in diacritics and homographs (words that are spelled the same but have different pronunciations and meanings depending on the short vowel diacritics which are not usually shown on written words). This means that AI has difficulty matching the pronunciation of a homophone with the context in which it is used. They confuse short vowel diacritics on the suffix /ta/, /ti/ /tu/ suffix تاء التأنيث when it refers to first, second, or third person, masculine or feminine, imperative and past tense (كَتَبَ كُتِبَ كُتِبَتْ كَتَبَ كَتَبَتْ كَتَبَ كَتَبَتْ كَتَبَ كَتَبَتْ كُتِبَ كُتِبَتْ). This affects comprehension in L2 learners and causes cacophony and distortion for native speakers and non-native speakers of Arabic. The article sheds light on how AI reads Arabic aloud, classification of pronunciation errors in AI narration; variations in the type and frequency of pronunciation errors across videos; why AI makes mistakes in pronouncing Arabic but does not make grammatical or syntactic errors in AI-narrated content; and how AI pulls off realistic intonation. Additionally, the article gives suggestions for improvement and recommendation for students learning Arabic as a foreign language.

| KEYWORDS

AI narrated Arabic videos, AI-generated stories, AI narrators, pronunciation errors, Arabic diacritics, Arabic homophones, text-to-speech, language processing.

| ARTICLE INFORMATION

ACCEPTED: 01 November 2025 **PUBLISHED:** 25 November 2025 **DOI:** 10.32996/jcsts.2025.2.2.1

1. Introduction

The past few years have witnessed a revolution in Artificial Intelligence (AI) and the multitude of tasks that it can perform for linguistic, educational and entertainment purposes including audio and video content production. Stories, advertisements and medical educational videos narrated by AI voices are becoming increasingly common on platforms like YouTube, TikTok, Instagram and Facebook. The rapid advancements in Large Language Models (LLMs) and AI-generated content have accelerated AI-based narration such as *storytelling narration* used in audiobooks, interactive stories, and video games for both children and adults; *documentary narration* used in documentaries or educational videos; *advertising and marketing narration* used in audio ads or promotional *educational narration* found in e-learning lessons, tutorials, and training programs; videos and *interactive narration* found in voice assistants like Siri and Alexa, or in chat applications.

AI narration has received some attention from recent foreign researchers. Numerous studies in the literature explored a range of issues related to AI narration such as assessing student satisfaction using lectures generated by AI (Kallamadugu, Lawal, Burgett, Gajjar, & Bingenheimer, 2025); AI-narrated stories for children using a multi-agent system (Xu et al., 2025); narrative modelling

and archetypes in large AI models (Kabashkin, Zervina & Misnevs, 2025); AI versus human-generated voices and avatars with a focus on user engagement and cognitive load (Zhang, Lucas, Bem-haja, 2025); persuasiveness and engagement in AI-generated narratives (Chu & Liu, 2024); co-creation of digital stories with asylum-seeking children using GenAI tools (Cheung & Shi, 2024); ChatGPT and mind maps for children's storytelling (Chen, Xie, Zou & Wang, 2023); ChatGeppetto for AI storytelling using semiotic relations (De Lima, Feijó, Cassanova & Furtado, 2023); digital storytelling in healthcare communication (Park, Forhan & Jones, 2021); consumer perception of AI-narrated voice-overs vs. journalist-narrated voice-overs in news videos (Smink & Lentz, 2024) and others.

While AI narration continues to evolve in structure and design, it is still facing challenges in expressive delivery, especially in linguistically complex languages like Arabic. The literature revealed numerous Arabic studies that examined the challenges facing Arabic language processing by AI applications. For example, El-Fiki (2024) addressed the linguistic challenges faced by AI in handling Arabic, particularly in text-to-speech conversion. His study highlighted weaknesses of AI speech models in pronouncing emphatic letters and their failure to account for grammatical context in automated speech. His study confirms the existence of pronunciation errors in AI applications and supports the need for analyzing them in AI-narrated stories for phonetic accuracy.

The Arabic Language Researchers website (2024) explored AI's capabilities in processing Arabic. The study discussed challenges in intonation, diacritization, and vocal interaction, noting that current AI models fail to distinguish accurately between dialects and Modern Standard Arabic. This opens the door for analyzing vocal performance in AI-generated stories and highlights the role of diacritics in improving pronunciation.

The impact of AI on Arabic vocabulary through voice applications using Modern Standard Arabic was investigated by HAD Media (2024). The study pointed out poor pronunciation in complex and classical words and recommended training models in educational contexts. It provides practical examples for analysis and supports the idea that AI-generated stories may dilute classical vocabulary usage.

The impact of AI on phonetic and intonational cues in Arabic was analyzed by Jabara & Al-Shukri (2024). Findings indicated that AI models fail to consider the morphological and syntactic context which affects intonation and phonetic cues. The study offered a scientific basis for analyzing phonetic errors in AI-generated stories and supported the use of descriptive analytical methods in evaluating AI-narrated content.

Another group of studies focused on solving the diacritic problem in Arabic language processing. Qubi (2024) examined automatic speech recognition (ASR) systems for Arabic, focusing on linguistic and computational aspects of AI speech processing. His study contributes to the technical understanding of how AI handles Arabic speech and may inform future evaluations of pronunciation accuracy in AI-narrated stories.

In another study, Abed, Alshayeji & Sultan (2019) examined the impact of diacritics on Arabic automatic speech recognition (ASR). They built ASR models using diacritized and non-diacritized corpora across various durations (1–23 hours) and technologies like GMM & DNN. 80 models were created using SRILM and KALDI. Results showed Word Error Rates (WERs) from 4.68% to 42%, with diacritics increasing WER by 0.59% to 3.29%. Despite this, the author recommended the introduction of diacritics for integrated systems like voice-enabled translation, where overall accuracy benefits outweigh the WER increase.

To tackle the difficulty of detecting Arabic AI-generated text (AIGT) due to diacritics, Alshammari & Elleithy (2024) used Transformer-based models as AraELECTRA, AraBERT, XLM-R & mBERT) to train detectors on 9666 examples of human-written and AI-generated texts, both with and without diacritics. Diacritized training improved accuracy to 98.4%, outperforming GPTZero's 62.7%. However, duplicating examples with and without diacritics proved inefficient. Applying a dediacritization filter during evaluation yielded optimal results. Though focused on Arabic, the architecture is adaptable to other languages facing similar challenges.

To enhance Arabic NLP across modalities and support culturally sensitive AI applications, Aly (2025) introduced a multimodal system for processing Classical Arabic with diacritics, integrating speech, text, and vision. Using the Quran's 20 styles as a dataset, the system included tools like QR-Vision (image diacritic recognition), QRDiaRec (diacritic restoration), and QRSR/DASAM (speech alignment). SemSim analyzes semantic and numeric data. Results revealed 94.2% accuracy in text diacritization, 91.67% in OCR, and up to 96% in audio segmentation.

Ahmad, Haider & Saed (2025) evaluated Arabic speakers' reactions to AI-dubbed versions of *Pride and Prejudice* using ElevenLabs and Dübverse. Findings showed that 80% of respondents found AI dubbing useful for accessibility. ElevenLabs

outperformed Dübverse in voice cloning, handling colloquial language, and managing multiple speakers. However, both lacked lip-synchrony and needed improvements.

While the above studies collectively advance our understanding of AI storytelling across various languages and platforms, they also reveal a persistent gap in addressing the linguistic intricacies of Arabic, particularly in automated speech delivery. Despite growing interest in AI-generated stories, Arabic remains an exceptionally complex challenge in computational processing due to its unique phonological and orthographic characteristics. Few investigations have examined how the Arabic unique phonological and orthographic features affect pronunciation accuracy and listener comprehension. There is an urgent need for a comprehensive study that explores the challenges that the Arabic writing system poses, not only for understanding pronunciation errors existing in AI narration, but also in assessing their impact on comprehension and acceptance by listeners, whether native or non-native speakers of Arabic. Therefore, the current study stems from field observations of several AI-narrated Arabic videos on YouTube and other social media platforms. It aims to analyze, classify, and interpret the patterns of pronunciation errors that exist in them. It also seeks to uncover the technical and linguistic causes behind these errors, the extent to which these challenges the effectiveness of such applications and provide recommendations for improving AI voice performance in educational and narrative contexts. Furthermore, the study seeks to give recommendation for analyzing Arabic texts by addressing essential aspects such as understanding the precise context of words, correcting pronunciation errors, and managing text vocalization. In other words, the current study aims to answer the following questions: (i) Which types of pronunciation errors are most frequent in AI-narrated stories, advertisements and medical educational YouTube videos? (ii) Are these errors due to absence of diacritics or inadequate AI model training? (iii) How do these pronunciation errors affect auditory discrimination and comprehension? (iv) In what ways is AI narration different from human narrators in pronunciation and delivery?

This study is significant because it builds on previous research studies, including the author's published study on AI transliteration of the letter G in foreign words (Al-Jarf, 2025). This study has educational and linguistic significance as AI-narrated videos can be used in teaching Arabic to children and non-native speakers. Correcting pronunciation errors in AI-narrated videos is essential for preserving the Arabic language integrity. Additionally, AI narration is a new field, especially in informal platforms like YouTube. Most existing research focuses on text-to-speech (TTS) systems, natural language generation, or AI in education/media, not specifically on YouTube storytelling or pronunciation accuracy. Arabic pronunciation by AI is underrepresented in global research compared to English or Mandarin.

2. Theoretical Framework

2.1 Derivation in Arabic

In Arabic, there are several word formation processes that are used to coin new vocabulary. The most prominent and most productive of these is derivation (ishtiqāq اشتقاق), whereby words are derived from a triliteral or quadriliteral root, combined with a set of vowels alternating with the root consonants. Different derivational patterns (morphological templates) are employed to form active participles, passive participles, abstract nouns, nouns referring to occupation, nouns of place and time, names of diseases, tools, devices, relational adjectives, the diminutives, and others. For example, many words are derived from the verb kataba ("to write"), such as: كتاب [kiˈtaːb] (book), كتب [ˈkutub] (books), كتيب [kuˈtajˌjib] (small book/booklet), مكتب [ˈmaktab] (office), مكتبة [makˈtaba] (library), كتابة [kiˈtaːba] (writing), كاتب [ˈkaːtib] (writer, masculine), كاتبة [ˈkaːtiba] (writer, feminine), كتابات [kitaːˈbaːt] (writings), مكاتبات [mukaːtaˈbaːt] (correspondences), كتاتيب [kataːˈtiːb] (traditional schools), مكتوب [makˈtuːb] (written/letter), كاتب [ˈkaːtaba] (he corresponded), كويتب [kuˈwajtib] (small author), كتب [ˈkattaba] (he made someone write), يكتب [ˈjaktub] (he writes), اكتب [ˈuktub] (write), استكتب [istaˈktab] (he asked someone to write something academic), استكتاب [istikˈtaːb] (asking someone to write something academic), اكتتب [ikˈtatab] (he subscribed), اكتتاب [iktiˈtaːb] (subscription), انكتب [inˈkatab] (he was registered/written). Verbs can also be derived from nouns, and even from certain abbreviations, through a process known as reverse derivation (اشتقاق عكسي). Examples include: باب [baːb] (door), بوّب [ˈbaw.wab] (he classified), تبويب [tabˈwiːb] (classification), مبوب [muˈbaw.wab] (classified); فهرس [ˈfih.ris] (index), يفهرس [juˈfah.ris] (he indexes), فهرسة [ˈfah.rasa] (indexing), مفهرس [muˈfah.ras] (indexed); كهرباء [kahrabaːʔ] (electricity), كهرب [ˈkahrab] (to electrify), مكهرب [muˈkahrab] (electrified) (Al-Jarf, 2020; Al-Jarf, 1995; Al-Jarf, 1990; Al-Jurf, 1995).

The Orthographic/Phonological System in Arabic

Arabic has three long vowels /a:/, /u:/, /i:/ and three short vowels /a/, /u/, /i/ which are represented by diacritics marked over and below consonant letters. Diacritical marks include: *fatḥa, ḍamma, kasra, sukūn, shadda, madd, tanwīn, the sign of the omitted alif, hamza, ṣila*. In words that have short vowels, only the consonants are written. In words containing short vowels, only the consonants are written with no diacritical marks. Adult readers infer the vowels from context or from diacritical marks if present.

Moreover, Hamzat al-waṣl همزة الوصل and hamzat al-qaṭ همزة القطع appear in forms such as اكتب [ktub] (write!), اختصم [xtasˤam] (he disputed), اعتصم [ʔtasˤam] (he adhered), اعتصام [ʔtisˤaːm] (adherence), استعلم [staʔlam] (he inquired), استبشر

[stabʃar] (he rejoiced), أنظر [ˈunzˤur] (look!), استمع [staˈmaʕ] (listen), with hamzat al-waṣl disappearing in connected speech as in جاء الولد [ˈʒaːʔa lˈwalad] (the boy came).

Solar and lunar *lām* (Arabic definite article) are illustrated in الولد [alˈwalad] (the boy), الثعلب [alˈθaʕlab] (the fox). The hamza of "al-" is not pronounced in connected speech, as in كتب الدرس [ˈkataba dars] (he wrote the lesson), or when attached to a single letter: كالولد [kalwalad] (like the boy), بالولد [bilwalad] (with the boy), والولد [walwalad] (and the boy), فالولد [falwalad] (so the boy). Feminine *tāʾ* تاء التأنيث appears in verbs and nouns such as أكلت [ˈakalat] (she ate), فاطمة [ˈfaːtˤima] (Fatima), حياة [ħaˈjaːt] (life). The tied *tāʾ* التاء المربوطة is pronounced *tāʾ* in connected speech and *hāʾ* in pause, as in السيارة سوداء [asˤajˈjaːra suːˈdaːʔ] (the car is black), جاءت السيارة [ˈʒaːʔat asˤajˈjaːra] (the car came). It is pronounced as *tāʾ* ة and dotted when used for femininity or reciprocity, as in السيارة [asˤajˈjaːra] (the car), المشاركة [almuˈʃaːraka] (participation), and pronounced as *hāʾ* ه and undotted when used for pronouns or structure, as in له وجيه [ˈlahu waˈʒiːh] (he has prestige). The short vowel /i/ is inserted after تاء التأنيث الساكنة before a consonant, as in جاءت البنات [ˈʒaːʔat alˈbanaːt] (the girls came).

Verb forms include past, present, and imperative, with distinctions for first, second, and third person, masculine and feminine, active and passive voice, transitive and intransitive verbs. Imperative verbs may be reduced to a single letter with *hāʾ*, as in عه [ʕih] (keep promise!) and فه [fih] (understand!).

Attached pronouns include forms such as يتعاونّ [jataʕaːwanna] (they cooperate, feminine), أعنّي [ʔinni] (help me), آمنّا [ˈʔaːmanna] (we believed).

The alif of tanwīn in the accusative case appears in كتابًا [kiˈtaːban] (a book), فتىً [faˈtaːʔan] (a boy), عصًا [ʔaˈsˤaːʔan] (a stick). Adverbials connected with "iðin" include وقتئذٍ [waqtʔiðin] (at that time), حينئذٍ [ħiːnʔiðin] (then), يومئذٍ [jawmʔiðin] (on that day).

The *wāw* occurs in the middle of words such as أولاء [ʔuˈlaːʔi] (those), أولئك [ʔuˈlaːʔika] (those people), أولو [ˈuːluː] (possessors), أولات [ˈuːlaːt] (women possessing), and at the end of words such as عمرو [ʕamr] (Amr).

In the the first three grades, Arab children practice and acquire decoding skills using diacritical marks. Later, they learn to decode Arabic written words without diacritics, relying on the context to understand meaning and correct pronunciation associated with it (Al-Jarf, 2007). Words that have the same constituent letters but differ in pronunciation and meaning due to the use of different diacritics on two or more consonants (homophones) pose word identification/recognition difficulties for AI. For native speakers of Arabic, the correct pronunciation of homophones is inferred from context (Al-Jarf, 1999). In the case of homophones such as كَتَبَ [ˈkataba] (he wrote), كُتِب [ˈkutiba] (it was written), كُتب [ˈkutub] (books), كتّب [ˈkatːaba] (he made someone write), كُتُب [ˈkutub] (books), كتبتُ [kaˈtabtu] (I wrote), كُتِبَتْ [ˈkutibat] (it was written), the correct pronunciation is usually inferred from the context of the text.

In addition, Arabic contains case endings (حركات الاعراب) that determine the syntactic function of the word in the sentence and its position as subject, object, or genitive, adding another layer of phonological and orthographic complexity.

In the case of homophones - words that are spelled the same, but have different pronunciations and meanings depending on the short vowel diacritics as كَتَبَ كُتِب كُتب كتّب كُتُب كتبتَ كتبتُ كتبتِ كُتِبَتْ. The correct pronunciation of a homophone is usually inferred from context. Additionally, Arabic has inflectional endings (declension) to show the function of a word in a sentence.

1.2. *How AI Reads Aloud*

AI narration uses text-to-speech (TTS) technology, where an AI takes a written text and converts it into spoken words. This involves complex linguistic models, phonetic databases, and voice synthesis techniques to generate speech that sounds natural. Examples of AI tools that can generate Arabic narration for stories & other content are:

- ***ElevenLabs*** which offers Arabic text-to-speech (TTS) with customizable voices, accents, and emotional tone. It aims to produce natural-sounding Arabic speech with contextual awareness.
- ***Narakeet*** which provides Arabic voice-over generation, allowing users to convert text into lifelike Arabic narration for audiobooks, videos, and educational content.
- ***Google's WaveNet*** which is a neural network model developed by DeepMind that generates highly realistic human speech by directly modelling raw audio waveforms.

1.3. *Types of Arabic AI-generated and narrated story videos*

The first type of Arabic story videos uses human-written scripts but rely on AI for narration. AI handles both the storytelling and voice, with varying degrees of human supervision. In the **second** type, especially on YouTube and TikTok, the stories are translated from English sources such as public domain texts, Reddit posts, blogs, into Arabic using tools like Google Translate,

GPT-4, DeepL, or platforms like Araby.ai and Katteb.com for more stylistic refinement. Creators may manually adjust the translation for tone, dialect, or clarity, or rely on AI to adapt the style (e.g., Classical vs. colloquial Arabic). The finalized script is then narrated using text-to-speech engines such as Alstudio, ElevenLabs, Google's WaveNet, or Araby.ai, which offer expressive, customizable Arabic voiceovers in terms of gender, emotion, and dialect. In a third type, the script is generated by AI using large language models (LLMs) like GPT-4, SILMA, mBERT, XLM-R, AraBERT,or Katteb, as well as platforms such as Alstudio and Araby.ai. These tools are fine-tuned or prompted to produce culturally relevant, diacritic-aware narratives. In many cases today, both the story and the narration are created by AI. The creator provides a prompt or idea (e.g., "a suspense story set in old Damascus"). An AI language model (like GPT-4, Katteb or SILMA,) generates the full Arabic script.

The use of AI in story generation and narration is characterized by speed, no cost, ad scalability. The entire story video can be produced in minutes. There is no need to hire translators or voice actors. Creators can publish daily content across multiple languages.

2.4 *Average Number of Arabic Words Read in AI-Narrated Stories*

The number of words read can be estimated based on the typical reading speed of either an AI narrator or a human reader, as follows: (i) the average slow reading speed is about 100 words per minute (common in meditation or deep learning videos). (ii) the average medium reading speed is about 130–150 words per minute (most common in educational videos). (iii) the average fast reading speed is about 180–200 words per minute (used in news or promotional content).

3. Methodology

3.1 Sample and Data Collection

A sample of 25 Arabic YouTube videos narrated by AI were collected. The sample included 17 stories, some originally produced in Arabic and some translated from English and Chinese, alongside one psychological video, three health/ medical videos and one advertisement. The sample of videos included the following:

- قصة إسلام ملياردير ياباني بعد لحظة لم يتوقعها... نصف موزة غيّرت مصيره! (19:16minutes)
- تحذير إيلون ماسك: الحقيقة التي يخاف الجميع من سماعها عن انهيار النظام القادم (12:46 minutes)
- قصة مؤثرة تلامس القلب || "عاقبة المحبة"|رواية عربية كاملة || أفضل رواية|روايات رومانسية|روايات حب (16:24 minutes)
- في يوم الأب، إبني أهدي لحماته سيارة فاخرة وأنا أهداني محرد حذاء رخيص وبعد أسبوع إتصل بي 23 مرة... (36minutes)
- بيت الزمان من أجمل القصص قبل النوم ضع السماعات واغلق عينيك واستمع #قصة (3:27 minutes)
- الملبونير المتغطرس طرد مربية الأطفال... وفي اليوم التالي صُدم بما رأى (31:44 minutes)
- البناتُ عُددنَ تَحْسًا، و لكن...قصة مؤثرة | قصة عاطفية من إبداع قناة قصة مؤثرة (38:49 minutes)
- "قاموا بطردها وضربوها بعد أن هددوها... ظنّوها امرأة عادية، ولم يدركوا من تكون في الحقيقة!"
- سيدة أمريكية تسيء معاملة عاملة عربية في مطعم راقٍ... والنتيجة كانت مفاجئة!
- بر الوالدين# قصص واقعية#🧐الذي ضرب أمه بحذاء زوجته إرضاءً لها، فماذا حدث له قبل أن يموت؟ قصة الشاب#
- غرور الأخت وظلم الزوج"|قصة مؤثرة | قصة عاطفية من إبداع قناة قصة مؤثرة "
- "مليونير أخذ ابنته البكماء إلى مقهى — وتحمّد عندما استخدمت نادلة عربية لغة الإشارة"
- بعد أن أصبحت أرملة، بدأت أعيل نفسي من خلال الزواج المؤقت نكاح المتعة
- عامل نظافة نفخ فقاعات لتهدئة فتاة مصابة بالتوحد – دون أن يعلم أن والدتها الملياردرة كانت تراقب.
- كان لديّ مطعم شاحنات، وفي أحد الأيام ترك شابٌ والدته العجوز عندي ورحل.|صوت عطفي|
- سيدي، ارحوك خذ اختي الصغيره، انها تجوع — فالتفت الرجل وتجمد من الصدمه!
- قامت بإهانة حارسة الأمن فعلمتها الرئيسة معنى الإحترام ♡ #shorts #ytshorts
- رجل أراد إختبار مدى أمانة هذا الصبي 😁 #shorts
- هذا الاب قام باختبار خنان أبنائه عليه (صينية)
- الحقيقة عن النساء اللواتي يعشن بمفردهن بدون رجل | كارل يونغ
- أخطاء في تناول الأفوكادو قد تؤثر على صحتك بعد سن الستين 5
- كبار السن، 5 فواكه حلوة وشائعة تساعد في مكافحة السكري | نصائح صحية لكبار السن
- يكشف طبب أسنان عن الغسول الذي يعالج اللثة ورائحة الفم الكريهة. (11:58 minutes)
- وجدت إبنها بعد 20 سنة من الغياب♡ #shorts #ytshorts
- إستقالة من الوظيفة لأنه لم يتحمل إهانات المدير له

The videos in the sample ranged in length between 3 and 40 minutes. All selected videos were narrated in Modern Standard Arabic, while those produced in colloquial dialects were excluded. In some cases, the text appeared on screen simultaneously with the narration. Pronunciation errors analyzed in this study do not include mistakes in inflectional endings حركات الإعراب. Furthermore, the analysis does not address cultural or social dimensions, nor does it evaluate whether the thematic content of the videos is appropriate to Arab and Islamic cultural contexts, or how the content is foreignized or localized.

3.2 Data Analysis

Each video was watched and errors in pronouncing Arabic words were noted and recorded in a compounds or short stretches of discourse showing the faulty word and its correct equivalent. The pronunciation errors differed from video to video. In total, more than 700 pronunciation errors were identified across 24 videos; however, the results focus on a representative sample of 105 errors selected to reflect variation in type and context, in order to enable qualitative analysis and avoid excessive length in the article. The errors encompassed changes in diacritical marks, mistakes in tanwīn and the addition of extraneous *nūn*, inaccuracies in pronoun, gender, and number agreement, unintelligible or uninterpretable words, confusion between two meanings of the same orthographic form due to failure to match pronunciation with context, misidentification of the correct pronunciation of a word or the invention of a pronunciation not used in Arabic, failure to apply correct vocalization for jussive verbs, confusion between verb and noun forms as well as between masculine and feminine, dual and singular distinctions, errors in derived forms such as active and passive participles, mismatches in suffix agreement caused by mispronunciation of the open *tāʾ* التاء المفتوحة, confusion among verb forms in first, second, and third person, errors in selecting between active and passive voice or transitive and intransitive verbs, unnatural pauses within phrases, mispronunciation of the hamzat al-qaṭʿ همزة القطع , deviations in pronouncing loanwords or Arabized foreign terms, failure to pronounce foreign proper names according to their original language pronunciation, omission of translation for certain foreign words, and mismatches in agreement between adjectives or verbs and their referent nouns. Finally, the rate of pronunciation errors was calculated across the videos.

4. Results

4.1 *Rate of Pronunciation Errors in the Videos*

Pronunciation errors vary from one video to another in both quantity and type, depending on the AI model used and the text-to-speech app used. The number of words read in a 19-minute video was estimated based on the typical reading speed of either an AI narrator or a human narrator mentioned above. For example, with 45 mispronounced words (without repetition) in a 19-minute video, the average error rate would be 2% or less. This is relatively acceptable in automated speech models, particularly when the words are difficult or uncommon. If pronunciation errors are distributed across non-repeated words, this would indicate that the model requires improvement in handling linguistic diversity or phonetic context.

4.2 *Classification of Pronunciation Errors in AI Narration*

Data analysis has shown that AI narrators in all the videos speak/read Standard Arabic with a natural voice, good expression and intonation. AI-generated voices sound human-like, fluid and expressive. AI narration is convincing, and the voices do not sound artificial. Rhythm, stress, and tonal shifts are human-like. Arabic phonemes like ح /ħ/, خ /x/, ص /sˤ/, ض /dˤ/, ط /tˤ/, ظ /ðˤ/, ع /ʕ/, غ /ɣ/, ق /q/ are pronounced correctly. No grammatical or syntactic errors are found in the stories, medical, or advertisement videos. Some mistakes in inflectional endings حركات الاعراب were noted depending on word position in sentences, but these errors are not the focus of the current study.

Results of the video analysis in this study showed that AI does makes certain pronunciation errors, particularly in diacritical marks and in words composed of the same letters but differing in pronunciation and meaning depending on the vocalization of one or more letters (homophones). Since such diacritics are usually absent in printed texts, this demonstrates that speech synthesis systems still face major challenges in handling the fine nuances of Arabic grapheme-phoneme correspondence, especially in narrative contexts that require deep understanding of meaning, grammar, morphology and orthography. Changes in diacritical marks lead to ambiguity, altered or distorted meaning or making the word even meaningless. The following is a detailed account of pronunciation errors by type:

(1) Pronunciation errors due to changes in diacritical marks leading to ambiguity, altered meaning or nonsense words. These errors account for 32%, as in the following examples:

- اجبوني [?ajibu:ni] instead of [?aji:bu:ni] (answer me).
- أجلوا [?ajalu:] instead of [?ajja;u:] (postponed).
- أزل الساق برفق [azal] instead of [?azil] (slipped).
- اضطرابات pronounced /ʔitˤtˤːuraːbaːt/ instead of /ʔitˤtˤːiraːbaːt/ (disturbances).
- ألين [?layan] instead of [?ali:n] (more leaient).
- آمن [?a:minu] instead of ?aman] (he believed).
- أياد [?ayya:d] instead of [?ya:d] (a proper noun).
- بذرة [baðra] instead of [biðra] (seed).
- تخبز [taxbuz] instead of [taxbiz] (bake).
- تنفرني [tanfiruni] instead of [tunaffirni] (repel away).
- الجبال [?aljaba:l] instead of [?aljiba:l] (mountains).
- Gemination vs. lack of gemination as in جدا (very) pronounced /jadan/ instead of /jiddan/ without shadda, or مطبقة /muṭbaqa/ without gemination on b, altering meaning.

- حرفي [ħurafi:] instead of [ħirafi:] (artisan).
- حزن الاب [ħazana] instead of [ħazina] (the father was sad).
- حماك [ħammaːk] instead of [ħamaːk] (father-in-law).
- حمضي [ħamḍi:] instead of [ħimḍi:] (acidic).
- دوما [du:ma:] instead of [dawman] (always).
- سرية [sarya] instead of [siriyya] (confidential).
- صمتي [sˤamati:] instead of [sˤamti:] (my silence).
- عندها [?undaha:] instead of [ʔindaha:] (she has).
- فتلين [fatalayyan] instead of [fatali:n] (mellow down).
- فسالتها [fasa?alit-ha:] instead of fasa?alat-ha:] (I asked her).
- رائحة الروث **كفيت** [kafayyat] instead of kufi:t] (was spared the smell of manure).
- حذرا **كن** [kunna ħaðiran] instead of [kun ħaðiran] (be careful).
- هل **كنت** [kunnat] instead of [kunta] (were you?).
- صائد افوكادو ذكيا **كن** [kunna] instead of [kun] (be).
- كيس **ورقي** [ki:s waruqiyy] instead of [ki:s waraqi:] (paper bag).
- كيف **دمرت** [kayfa dammrit] instead of [dammarat] (how she destroyed).
- لا **تتردد** [la: tatardu] instead of [la: tararadad] (do not hesitate).
- **لأحضري** [liʔaħḍuri::] instead of [liʔaħḍura] (to bring).
- هل **مررت** يوما بلحظة [marrarta] instead of [marart] (have you been through).
- عمل **ورش** [warʃ] instead of [wuraʃ] (workshops).
- **يرزقني** [yurzaqani] instead of [yarzuqni] (give me sustenance).
- استخدمه آباؤكم [?istaxdumuhu] (your forefathers used it) instead of [?istaxdamahu].
- ألقيت [?alqayat] (cast) instead of [?alqaytu].
- توفيت [tuwafayt] (passed away) instead of [tuwufiyat] (passed away.

Further examples that show changes in diacritical marks leading to ambiguity or meaningless words. Such errors distort meaning. They result from the AI model's inability to identify the correct pronunciation of a word and the invention of pronunciations not used in Arabic. They indicate weakness in distinguishing diacritics that alter meaning, and the system's inability to link words to their grammatical or semantic context.

(2) Failure to select a pronunciation consistent with the context, such as confusing masculine and feminine forms by changing the short vowel diacritic on verb suffixes تاء مربوطة due to differences in the vocalization of التاء المفتوحة and confusion among verb forms in the first, second, and third person. This category of errors accounts for 29% of the sample, as illustrated in the following examples:

- Confusing the past tense & imperative forms as in تخيل (imagine)referring to third person masculine singular "he" instead of the imperative تخيل with sukun diacritic on /l/; شارك قصتك (share your story) referring to third person masculine singular "he" instead of the imperative شارك; واشترك (subscribe) referring to third person masculine singular "he" instead of the imperative اشترك with sukun.
- Confusing the referent of the pronoun suffix التاء المفتوحة such as: تصفحت وسائل التواصل (browsed social media) referring to *she* instead of I; اخذت اول حافلة (took the first bus) referring to she instead of I; اخذت اول حافلة referring to I instead of she; أصبحت (أنا) زوجتي منزعجة (my wife was annoyed) with أصبحت referring to I instead of she; حملتها referring to you instead of her; دخلتreferring to I instead of her; رايت (saw)referring to I instead of you (singular); شارك في التعليقات (make comments) referring to he instead of you; علمت (I knew) referring to she instead of I; فاستغليت (took advantage of) referring to I instead of you; (أنا) وقلت (هي) فَضحكت (I laughed and said) referring to she and وقلت referring to I instead referring to I in both verbs; لاحظت (you noticed) referring to she instead of you (sing); ((what I saw) ما رايته referring to you instead of I; فتحت (opened) referring to I instead of you; نزلت (came down) referring to she instead of I; هربت زوجتك for you instead of she; واجهت referring to she instead of you; ورايت referring to I instead of you; وضعت كيس المال with the verb referring to I instead of she; وقعت referring to I instead of she; وقفت هناك referring to she instead of I.

(3) Failure to apply correct vocalization for jussive verbs, confusion between verb and noun forms, masculine and feminine, dual and singular, as well as confusion between derived forms such as the active and passive participles اسم الفاعل واسم المفعول, as illustrated in the following examples:

- Failure to mark the jussive verb with the proper diacritical sign as in لم تختف [lam taxtaf] instead of [lam taxtafi] (did not disappear).

- Confusing masculine and feminine لا أعرفك كإنسانة (I do not know you as a human being) pronounced in the masculine rather than the feminine form, or لصحتك أنت (for your health) which incorrectly shifted from masculine [liSihhatika] to feminine [liSihhatiki].
- Confusing singular and dual forms: بين وجهي [bayna wajhayy] (dual) instead of بين وجهي [bayna wajhi:] (sing) *my face*.
- Confusing two derivatives (form referring to the doer of the action & receiver of the action): بعد ان كان محبطا ويائسا [muHbiTan]) doer of the action اسم فاعل) instead of محبطا [muHbaTan] (receiver of the action اسم مفعول) *frustrated*.

(4) Errors in agreement between the adjective + referent noun, pronoun and referent noun or verb + referent noun:
- (ذو الألوان الغريبة) "من بدلا الغريبة ذات الألوان الافوكادو **ذات** الألوان الغريبة الأفوكادو /ʔalʔafoka:do: **ða**:**t** alʔalwa:n alɣari:ba/ (feminine) instead of الألوان الغريبة **ذو** /**ðu**: alʔalwa:n alɣari:ba/ (masculine).
- العَفَن **تُنتِجُها** /tun.ti.d͡ʒu.ha: al.ʕa.fan/ (feminine verb) instead **of** العَفَن **يُنتِجُها** /jun.ti.d͡ʒu.ha: al.ʕa.fan/ (masculine verb) (produced by rot).

Such errors reflect weakness in tracking pronouns and understanding speaker or addressee gender, which is critical in Arabic.

(5) Confusion between two meanings of the same orthographic form due to failure to match the pronunciation with the context in which the word appears. These errors account for 10%, as in the following examples:
- بعض الوثائق **لاحضر** [li?aHDur] which means to attend instead of [?uHDir] *bring*.
- ولا **يفوت** الخطأ الأخير [yafu:t] which means enter instead of يفوت [yufawwit] *to let go, overlook*.
- عميقا اخذت **نفسا** [nafsan] meaning soul instead of [nafasan] *breath*.نفسا pronounced /nafsan/ (soul) instead of /nafasan/ (breath), shifting the meaning from "breath" to "soul.
- على الخبز الكامل **دهنه** [duhnihi] meaning its fat instead of [dahnihi] *spread it on bread*.
- تذكر [tathakur] which is meaningless instead of [tathakkar] *(remembered)*.
- الذكر [thikr] meaning supplicating Allah instead of [thakar] meaning (*male)*.
- فصلتها من العمل [faSSAlatha minial ʔamal] (tailor it), instead of [faSalatha] *(fired her)*.
- على الجميع **مطبقة** القوانين [muTbaqa] instead of [muTabbaqa] *applied to*.
- (يتواجد) زجاجات الحليب كيف [yaHDur zujaja:t alHaleeb] meaning to attend instead of [yuHaDDIr zujaja:t alHaleeb] *preparing the milk formula*.
- اليوم **من** [man] meaning who *today instead of [min] *from today*.
- بعد **الظهر** [thahr] after* back instead of [Thuhr] *after noon*.

These errors reflect a lack of contextual understanding, as the system is unable to distinguish between the multiple meanings of a single word depending on its context.

(6) Failure to account for context in selecting the passive and active voice, or intransitive and transitive verb forms, as illustrated in the following examples:
- بنيت ثروتي (passive voice, referring to third person her) instead of active voice referring to first person I) *I built my*. wealth تنجب (passive voice) instead of (active voice) *give birth*.
- ولدت فقيرا (passive voice, refers to he), instead of ولدت (passive voice, referring to I) *born poor*.
- فتحت (passive voice) instead of (active voice) *open*.
- يستحمهم (intransitive verb form) (*take them a bath) instead of يحممهم (transitive verb form) *(give them a bath)*.

(7) Errors in tanwīn and extraneous *nūn*, such as ليلا (at night) pronounced as /laylanan/ instead of /laylan/; ملابسه بسيطة (his clothes are simple) /mala:bisuhu basī:ṭa-tun/ instead of and ملامحه متعبة /malāmiḥuhu mutʕaba-tun/ with inappropriate tanwīn added before a pause. These errors are common in models trained on unvocalized texts, where the system struggles to distinguish between genuine tanwīn and extraneous *nūn*.

(8) **Few errors in hamzat al-qaṭʔ همزة القطع (glottal stop hamza), i.e.,** only a small number of mistakes were observed in the correct use or pronunciation of ***hamzat al-qaṭʔ*** همزة القطع the written and pronounced hamza that always appears at the beginning of a word as in saying يا إبنتي (dear daughter) with hamzat al-qat**ʔ** Instead of hanzat a l-wasl.

(9) **Unnatural pauses within phrases,** whereby the AI does not read the text in coherent thought groups, as illustrated in the following examples:
- من اللازم (pause) يعمل اكثر (work more than enough).
- وسبعون مليار (pause) أربعة (74 billion).
- سارة (pause) شَعَرْتُ (Sara felt).

- سيدي (pause) حاولت (I tried, Sir).

(10) Errors in pronouncing foreign words that have been Arabized or borrowed into Arabic in a way different from the conventional Arabic pronunciation, or failure to pronounce foreign proper names as they are pronounced in the original language. This type of error constitutes 10%. For example:
- سواريس Suárez [swa:ri:s) instead of [swa:ris].
- الافوكادو avocado [uvuka:do:] instead of [avocado:].
- آلان Alan [?ala:n] instead of [?alan].
- آيسكريم ice-cream [ayiskari:m] instead of [ayskri:m].
- بيجية [bi:jiya] instead of [beyjiyya] (beige).
- السلمون [salmu:n] instead of [salamon] (salmon).
- صويا. [su:ya] instead of [soya] (soy beans).
- الفولات [fawla:t] instead of [fo:la:t] (folate).
- كرن [karan] instead of [ka:ren] (Karen, a proper name).
- كوالا لمبور [kawa:la: limipu:r] instead of [kwala: lumpur] (Kuala Lumpur].
- اللب الأخضر الكريمي [al-kari:mi] instead of [kreymi] (creamy green pulp).

(11) Failure to translate some foreign words, i.e., certain foreign words were left untranslated such as:
- مسجد جامك [masjid jamek] instead of [masjid jameʔ] (Grand Mosque).
- desperation يبكون عليها ب instead of يبكون عليها بيأس (They weep over her in desperation).

(12) Errors in pronouncing unintelligible or uninterpretable words, such as الغواك مولي اللذيذ /al-ghuwāk mūlī al-ladhīdh/, which is illogical or incomprehensible. Such errors may result from flaws in text-to-speech conversion, misinterpretation of syntactic structure, or training on noisy data.

Most of the errors mentioned above are attributable to the absence of diacritical marks in the narrative texts, which forces the AI model to rely on guesswork. There is also a weakness in understanding narrative context and an inability to track references to specific names or characters in the story, as well as to match verbs, adjectives, or pronouns with their proper referents. This results in inappropriate pronunciation of words that share the same orthographic form but carry multiple meanings. Moreover, certain errors recur across different AI models, indicating a pattern that can be classified and analyzed statistically.

4.3 *Arab Viewers'/Listeners' Comments on AI-Narrated Videos in Arabic*

The pronunciation confusions made by AI narrators affect auditory discrimination and listening comprehension in students learning Arabic as a foreign or second language (AFL). It causes cacophony and distortion of the spoken text for native speakers of Arabic. Spontaneous comments posted by Arab viewers and listeners on the videos in the sample revealed lack of tolerance, dissatisfaction, and critique of errors. The following are examples of Arab viewers/listeners' comments:
- *The story is touching, but the reading is full of mistakes.*
- *The pronunciation is wrong. It should be "adh-dhakar" with a fatḥa, not "adh-dhikr" with a kasra.*
- *The meaning changes: "adh-dhakar" with a fatḥa means 'boy', while "adh-dhikr" with a kasra means 'remembrance of God'.*
- *Major linguistic errors in narrating this beautiful story, which you sadly distorted.*
- *You have catastrophic linguistic mistakes.*
- *"Ḥaythu shiʔt", the meaning changes. "Ḥaythu shiʔti", major mistakes. I won't follow you anymore.*
- *Oh God… dear commenter… I learned Arabic from you. Please, before you narrate a story, seek help from a teacher or professor in Arabic. The narration is full of mistakes.*
- *The reading contains major mistakes. Try to improve your pronunciation.*
- *You paused. That's a mistake. The meaning changes.*
- *You have catastrophic and severe mistakes.*
- *A painful story—and what hurt me most was your cruelty to the Arabic language. Have mercy on it with proper pronunciation, please!*
- *A beautiful story, but the pronunciation mistakes in the narration are awful. Shame on you.*
- *What is this incorrect reading?*
- *Why is everything done by AI? Why not a human narrating the story? There are so many mistakes in the speech. Why did you change the Arabic language?*
- *From how the story was delivered, it seems you didn't finish elementary school, so many mistakes. A piece of advice: before publishing the story, practice reading it.*

- *The story is beautiful and moving, but the narration is full of language mistakes. Arabic has been lost and is now a thing of the past. Alas for the next generation that won't be able to know proper Arabic.*
- *The avalanche of errors during narration robs the listener of the joy of completing the story. I've started lowering the volume and reading it myself to enjoy the story, because the avalanche of linguistic mistakes is unbearable—especially since we love our Arabic language and don't want its beauty distorted. The channel owner is a respectable professor, but he specializes in French and has no connection to Arabic or its literature. I thank the channel. If the owner likes my voice, he can contact me—and anyone who knows me.*
- *The abundance of errors in the narration burdens the listener and ruins the joy of finishing the story. The voice has become annoying, so I read it myself to enjoy the story without mistakes. The narrator doesn't master Arabic and doesn't love what he's narrating. He doesn't know where to place emphasis in a sentence or how to pronounce letters. If he wants to improve, he should listen to narrators who love and master Arabic. I hope the story's author replaces the narrator with someone knowledgeable.*

5. Discussions

5.1 Variations in type and frequency of pronunciation errors across videos

Results of the current study revealed variations in the types and amount of pronunciation errors across the AI-narrated videos in the sample due to differences in the AI models or software used. This can also be attributed to the type of AI model used for voice generation, language model quality and its ability to process context, diacritic awareness, and differences in text-to-speech (TTS) engine which include voice model training, prosody and intonation and error types, script quality and preprocessing (manual vs. AI-generated scripts, presence of diacritics) and customization and tuning. Some platforms allow creators to adjust pitch, speed, and emotion, which can either enhance or distort pronunciation depending on how it is used. This means that pronunciation accuracy in AI-generated Arabic content is the outcome of a complex interaction between the language model, the speech engine, the nature of the text, and the level of voice customization. This underscores the need to develop more integrated models with greater awareness of Arabic linguistic context, particularly for educational applications.

5.2 Why AI makes mistakes in pronouncing Arabic

AI narrators in YouTube videos make mistakes in pronouncing Arabic for reasons such as lack of proper diacritics and training data limitations. Arabic writing often omits diacritical marks, making it harder for AI to determine the correct pronunciation of words. Some AI voice models might not be trained on a wide enough Arabic dataset, causing pronunciation errors. These tools use AI voice synthesis to create realistic speech, but pronunciation accuracy is still a challenge, especially when dealing with diacritics and homographs in Arabic. Arabic narration still presents challenges because AI must determine context-dependent pronunciation without vocalization marks. However, as Arabic AI speech models improve, better contextual analysis and phonetic adaptation will make Arabic narration more accurate over time.

On the contrary, AI-narrated English stories on YouTube make no pronunciation errors for reasons such as standardized spelling, predictable phonetics, massive training data, and advanced AI models for English. English words generally have a fixed spelling and pronunciation that remains consistent, unlike Arabic, where missing diacritics can change how a word is pronounced. Although English has exceptions, its phonetic rules are simpler compared to Arabic's complex system of diacritical marks and homographs. AI has been trained on vast amounts of English speech and text, giving it deep exposure to correct pronunciation patterns. Since AI development has largely focused on English-speaking markets, English voice synthesis has been optimized much more than other languages.

5.3 Why AI does not make grammatical and syntactic errors in AI-narrated Arabic content

This comes down to how AI processes language. AI translation models are trained on structured grammar rules, ensuring proper sentence construction and accurate syntax. Arabic grammar and sentence structure is well-defined and follows clear linguistic rules, making them easier for AI to handle, but Arabic pronunciation can vary due to missing diacritics. AI understands sentence syntax through machine learning, so the word order and grammar stay accurate, even if pronunciation struggles with homographs. Pronunciation challenges stem from unvocalized text. AI mispronounces words when vowel markings (حركات) are missing, but the underlying grammar remains correct because AI knows word relationships and sentence structure. Arabic syntax is relatively stable, while pronunciation depends on phonetic interpretation, which is where AI struggles.

5.4 How AI pulls off realistic Arabic intonation

AI can achieve natural-sounding voices and intonation through the following: Advanced speech synthesis techniques, particularly neural text-to-speech (TTS) models. Unlike older robotic-sounding TTS systems, AI now uses neural networks for speech and deep learning to study how humans speak, including rhythm, stress, and tonal shifts. This allows AI-generated voices to sound more fluid and expressive. AI learns prosody patterns (changes in pitch, stress, and rhythm in spoken language) by analyzing massive datasets of human speech, enabling it to mimic natural intonation, pauses, and emphasis. AI models do not just read

words, they analyze sentence meaning and structure to apply the correct intonation. For example, a question is spoken differently from a statement, and AI adjusts tone accordingly. Some advanced AI narrators allow users to tweak the tone, mood, or expressiveness of a voice. For instance, making it sound more excited, serious, or calm depending on the context. Google's WaveNet, ElevenLabs, other AI Voice Models and AI-powered TTS systems use real human voice recordings to synthesize speech, allowing AI to generate voices that sound natural instead of robotic.

6. Recommendations

Over time, AI speech synthesis had greatly improved as developers continued to refine training data and fine-tune pronunciation models. Still, further improvements in AI pronunciation of Arabic should be made through several advancements in speech synthesis and language processing. For better accuracy, and to improve AI's ability to match pronunciation with context AI speech models need the following: (i) Enhanced training data and exposure to diverse Arabic pronunciations, including different dialects, regional accents, and phonetic complexities. A richer dataset helps AI recognize and reproduce accurate speech patterns. (ii) Contextual understanding: Arabic script often omits diacritical marks (حركات), making pronunciation ambiguous. AI models trained to infer missing vocalization based on context could improve speech accuracy. (iii) Contextual language models and deeper natural language understanding (NLP) models that analyze the entire sentence structure and meaning before selecting the correct pronunciation. This would work similarly to how humans infer meaning from context when reading unvocalized Arabic text. (iv) Better neural speech synthesis, deep learning models and neural networks to simulate natural speech flow can make AI-generated voices sound more human-like and fluid. (ii) Diacritic restoration: Training AI with massive datasets where diacritical marks are applied correctly can help teach it to predict the missing vowel marks based on surrounding words and grammar rules. (v) Semantic AI integration: Combining speech synthesis with semantic analysis to help AI determine meaning first, then applying the correct pronunciation. For example, understanding whether علم means *science* (*ʿilm*) or *flag* (*ʿalam*) based on sentence context before vocalizing it. (vi) Human-AI hybrid correction: AI-generated voiceovers could go through a quick human-assisted verification process where native speakers correct any pronunciation mistakes, helping refine AI models over time. (vii) Arabic AI voice technology is progressing, but tackling the diacritic challenge requires a mix of linguistic refinement, deeper AI learning, and improved contextual recognition.

To expand research in AI-based speech narration in Arabic, future investigations that illuminate pronunciation challenges and their pedagogical implications can include the following: (A) Experimental studies comparing novice and advanced learners' comprehension when exposed to AI-generated speech, with particular focus on errors arising from incorrect diacritics. (B) Perceptual analyses of learners' attitudes toward AI voices and the impact on engagement in learning processes. (C) Technical–linguistic evaluations of speech engines' performance in pronouncing homonyms and expressive intonation. (D) Corrective modelling: Design hybrid frameworks combining human oversight with automated verification to enhance pronunciation quality in educational contexts. (E) Corpus development: Establish a dedicated, fully vocalized Arabic database to train AI on contextual pronunciation, a critical step for resolving ambiguities in homographic words. (F) Child-centered evaluation: Assess pronunciation and expressiveness in AI-driven storytelling for children, alongside analysis of linguistic errors in on-screen texts during educational and narrative videos (including overlays and kinetic typography).

In conclusion, the above recommendations will extend ongoing scholarship in AI-driven Arabic narration and open new avenues for the development of pedagogical tools that are both linguistically precise and pedagogically effective, particularly for non-native learners.

Conflicts of Interest: The author declares no conflict of interest.

ORCID ID: https://orcid.org/0000-0002-6255-1305

References

[1] Abed, S., Alshayeji, M., & Sultan, S. (2019). Diacritics effect on Arabic speech recognition. *Arabian Journal for Science and Engineering*, *44* (11), 9043-9056.

[2] Ahmad, M., Haider, A. & Saed, H. (2025). Assessing AI-driven dubbing websites: Reactions of Arabic native speakers to AI-dubbed English videos in Arabic. *Humanities*, *6* (1).

[3] Al-Jarf, R. (2025). Arabic transliteration of borrowed English nouns with /g/ by Artificial Intelligence (AI). *Journal of Computer Science and Technology Studies*, *7*(9), 245-252. https://doi.org/10.32996/jcsts.2025.7.9.29. Google Scholar

[4] Al-Jarf, R. (2020). Issues in Translating English and Arabic Plurals. The Journal of Linguistic and Intercultural Education - *JoLIE, 13*(1), 7-28. https://doi.org/10.29302/jolie.2020.13.1. Google Scholar

[5] Al-Jarf, R. (2018). First, Second and Third Grade Students' Word Identification Difficulties. Eurasian Arabic Studies Арабистика Евразии, 8, 22-93. Google Scholar

[6] Al-Jarf, R. (2015). *English and Arabic writing systems for translation students*. Academia.edu . Google Scholar

[7] Al-Jarf, R. (2007). Developing reading and literacy in Saudi Arabia. ERIC ED497944. Google Scholar

[8] Al-Jarf, R. (1995). *An Arabic word identification diagnostic test for the first three grades. Center for Educational Research.* College of Education. King Saud University. Google Scholar

[9] Al-Jarf, R. (1994). Analysis of Saudi first, second and third grade students' errors in word identification. *Journal of Contemporary Education, Cairo, 9* (61), 88-147. Google Scholar

[10] Al-Jarf, R. (1992). *Classification of word identification exercises in elementary school reading textbooks in Saudi Arabia*. Third Yearbook of the Saudi Educational and Psychological Association. Google Scholar

[11] Al-Jarf, R. (1989). *Classification of reading comprehension questions in elementary basal readers in Saudi Arabia.* Center for Educational Research, King Saud University. https://www.researchgate.net/profile/Reima-Al-Jarf/publication/352665671 Google Scholar

[12] Al-Jarf, R. (1995). Contrastive Analysis for Translation Students. 2nd Edition. King Saud University. Google Scholar https://www.researchgate.net/publication/268274970.

[13] Al-Jarf, R. (1994). English and Arabic Phonology for Translation Students. https://www.researchgate.net/publication/281003181. Google Scholar

[14] Al-Jarf, R. (1990). *A Contrastive analysis of English and Arabic morphology.* https://www.researchgate.net/publication/31 2193999. Google Scholar

[15] Al-Jurf, R. S. (2002). A Contrastive Analysis of English and Arabic for Translation Students. King Saud University. Google Scholar

[16] Al-Jurf, R. (1995). *A Contrastive analysis for translation students*. Riyadh: King Saud University. https:/www.researchgate.net/publication/268274970, Google Scholar

[17] Al-Jurf, R. (1994). Contrastive phonetics for translation students. https://www.researchgate.net/publication/281003427. Google Scholar

[18] Alshammari, H., & Elleithy, K. (2024). Toward robust Arabic AI-generated text detection: Tackling diacritics challenges. *Information*, *15* (7), 419.

[19] Aly, A. (2025). *A support system for diacritic-aware classical Arabic language processing: Integration of speech, text, and vision modalities*. Doctoral dissertation. University of Washington.

[20] Bahetho Arabia. (n.d.). Artificial intelligence and the Arabic language: Prospects and challenges. *Bahetho Arabia*. https://bahethoarabia.com/aiaralan. (In Arabic).

[21] Chen, X., Xie, H., Zou, D., & Wang, F. (2023, October). ChatGPT for generating stories and mind-maps in storytelling. In *2023 10th international conference on behavioural and social computing (BESC)* (pp. 1-8). IEEE.

[22] Cheung, L. & Shi, H. (2024). Co-creating stories with generative AI: Reflections from undergraduate students of a storytelling service-learning subject in Hong Kong. Australian Review of Applied Linguistics, (47)3, 259 - 288. DOI: https://doi.org/10.1075/aral.24101.che.

[23] Chu, H., & Liu, S. (2024). Can AI tell good stories? Narrative transportation and persuasion with ChatGPT. *Journal of Communication*, *74* (5), 347-358.

[24] De Lima, E., Feijó, B., Cassanova, M. & Furtado, A. (2023). ChatGeppetto-an AI-powered Storyteller. *22nd Brazilian Symposium on Games and Digital Entertainment,* 28-37.

[25] El-Faki, M. A. (2025). Arabic Language Processing in Artificial Intelligence Applications Linguistic Challenges and Proposed Solution. *Proceedings of the Arabic Language Forum: Between the Challenges of the Digital Age and International Requirements*, 42, 2, 45–88.

[26] Igor Kabashkin, O. & Misnevs, B. (2025). AI narrative modelling: How machines' intelligence reproduces archetypal storytelling. *Information 16* (4), 319. Doi: 10.3390/info16040319.

[27] Jabara, M. B. A., & Al-Shukri, B. M. Y. (2024). Artificial intelligence and some verbal indicators in the Arabic language. *Journal of Research and Studies in New Media, 5*(3), 7–19. (In Arabic).

[28] Kallamadugu, A., Lawal, N., Burgett, J., Gajjar, D. & Bingenheimer, K. (2025). Assessing Student Satisfaction Using Lectures Generated by Artificial Intelligence (AI). Research Square. 1-35. DOI: https://doi.org/10.21203/rs.3.rs-7394155/v1.

[29] Lim, S. (2025). *Exploring the Effects of AI-Generated Pedagogical Agents in Instructional Videos on Learning* (Doctoral dissertation, Columbia University).

[30] Park, E., Forhan, M. & Jones, C. (2021). The use of digital storytelling of patients' stories as an approach to translating knowledge: a scoping review. *Res Involv Engagem* 7, 58. Doi: 10.1186/s40900-021-00305-x.

[31] Qubi, A. (2024). Artificial intelligence and automatic processing of the Arabic language: A study on automatic speech recognition systems. *Scientific Research Journal, 1*(2), 8. ojs.ahumi.edu.vn (In Arabic).

[32] Smink, A. & Lentz, T. (2024). Consumer Perception of AI-narrated Voice-Overs vs. Journalist-narrated Voice-Overs in News Videos. Master's Thesis. Tilburg University, Netherlands.

[33] Xu, X., Mei, J., Li, C., Wu, Y., Yan, M., Lai, S., & Wu, M. (2025). Mm-storyagent: Immersive narrated storybook video generation with a multi-agent paradigm across text, image and audio. arXiv preprint arXiv:2503.05242.

[34] Zhang, Y., Lucas, M., Bem-haja, P. et al. (2025). AI versus human-generated voices and avatars: rethinking user engagement and cognitive load. Education and Information Technologies. https://doi.org/10.1007/s10639-025-13654-x.