

---

**| RESEARCH ARTICLE**

## **Calligraphic Text Recognition by Gemini, Ernie ViLG and Google Translate: A Comparative Study of Arabic, Japanese and Chinese**

**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](mailto:reima.al.jarf@gmail.com)

---

**| ABSTRACT**

This study investigates the ability of Gemini, Ernie ViLG, and Google Translate (GT) to recognize Arabic, Japanese, and Chinese calligraphic text images. Analysis of 15 Arabic, 7 Japanese, and 10 Chinese calligraphic samples shows that Gemini successfully matched 12/15 Arabic calligraphic texts with their correct Qur'anic verses, produced accurate translations for all Japanese samples, and correctly interpreted 9/10 Chinese texts. Ernie ViLG generated incorrect or random Qur'anic matches for Arabic, correctly translated 4/7 Japanese and 3/10 Chinese samples, and frequently defaulted to culturally common themes when unable to decode strokes. GT failed to produce any translations for Arabic calligraphy and rendered partial, fragmented, or incoherent translations for Japanese and Chinese images. Across Arabic, Japanese or Chinese calligraphic texts, the three AI models exhibited distinct strengths and weaknesses due to their differing approaches to calligraphic text recognition. Gemini proved to be the most reliable, leveraging multilingual training, pattern matching, and semantic retrieval to associate stylized characters with canonical works, enabling coherent and culturally grounded interpretations. On the contrary, Ernie ViLG struggled with literal recognition, especially in Arabic, and relied heavily on cultural priors when visual cues were ambiguous. GT was the weakest, as its OCR pipeline is optimized for printed or clean handwritten text and breaks down when confronted with stylized or artistically distorted calligraphy. Modern AI models process calligraphic text through a combination of visual feature extraction and linguistic prediction. Contemporary neural architectures employ convolutional and transformer-based encoders to interpret strokes, curves, and spatial patterns holistically, allowing them to infer characters even when calligraphy departs from standard rules of spacing, baseline alignment, or shape consistency. Arabic, Chinese, and Japanese calligraphic scripts challenge these AI systems because they intentionally distort or stylize characters, requiring both visual recognition and the ability to match ambiguous forms with known verses, idioms, or poetic structures. Overall, the current results highlight how multimodal AI models diverge when confronted with stylization, cultural priors, and incomplete visual information.

**| KEYWORDS**

Calligraphic texts, Arabic calligraphy, Japanese calligraphy, Chinese calligraphy, Gemini, ERNIE-ViLG, Google Translate (GT), OCR (optical character recognition), multimodal AI models, AI models

**| ARTICLE INFORMATION**

**ACCEPTED:** 20 December 2025

**PUBLISHED:** 29 December 2025

**DOI:** 10.32996/jcsts.2025.7.12.54

---

**I. Introduction**

Calligraphy<sup>1</sup> is a visual art that encompasses the design and execution of letterforms in an expressive, harmonious, and skilful manner using tools such as the pen, ink brush, or others. Although most writing systems develop calligraphic or stylized forms, calligraphy exists in specific languages more than others. Some cultures elevate calligraphy into a major artistic tradition. Arabic, Chinese, Japanese and Persian are the "big four" where calligraphy is a central cultural art, not just handwriting. They have deep, formal calligraphic traditions. They treat calligraphy as a major art form, with codified scripts, schools, and centuries of practice.

---

<sup>1</sup> <https://en.wikipedia.org/wiki/Calligraphy>

**Copyright:** © 2025 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (<https://creativecommons.org/licenses/by/4.0/>). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

Arabic<sup>2</sup> includes many styles as Kufic, Thuluth, Naskh, Diwani, Ruq'ah, Nasta'liq, Maghribi, etc. they are used in Qur'ans, architecture, signage, branding. Chinese styles<sup>3</sup> include Seal, Clerical, Regular, Running, Cursive, Wild Cursive. They have 3,000+ years of tradition. Japanese styles<sup>4</sup> include Kaisho, Gyōsho, Sōsho, Kana calligraphy, Edo styles, with strong Zen and artistic influence. Persian styles<sup>5</sup> include Nasta'liq, Shekasteh, Ta'liq which are used in poetry and manuscripts. Languages with recognized but less formal calligraphic traditions as Korean, Greek and Latin scripts, as well as several Southeast Asian scripts such as Thai, Khmer, and Burmese have stylized writing but are not as systematized. Writing systems that were created recently or used mainly for practical purposes have minimal or no traditional calligraphy as in modern alphabets for African languages (Yoruba, Swahili), indigenous scripts created in the 19th–20th centuries (Cherokee, Cree), and Romanized writing systems (e.g., Vietnamese Quốc Ngữ). Languages that are historically oral (e.g., many Indigenous languages) have no calligraphy because they have had no script until recently.

Calligraphy is not only an art form, but it also serves numerous functional purposes: It appears in announcements, shop and street signs, billboards, advertisements, wedding invitations, certificates, religious art, graphic design, commissioned artworks, logos, typography, carved stone inscriptions, memorial documents, and digital font design. It also features in maps, theatrical props, film and television graphics, testimonials, book and magazine covers, and other written media. Recognizing and decoding calligraphic scripts and styles is significant for tourists, shoppers, readers, art students, translators, publishers and other stakeholders.

Given the wide range of cultural, artistic, and functional contexts in which calligraphy appears across different languages, and the growing importance of recognizing calligraphic text by both humans and Artificial Intelligence (AI), a review of the literature revealed many studies that investigated this task AI's ability to process and identify calligraphic scripts. The first group of studies highlighted the variety of computational approaches applied to East Asian calligraphy and demonstrate the growing sophistication of AI in handling stylistic, historical, and multimodal dimensions of handwritten art such as calligraphy character detection by deep convolutional neural network (Peng et al., 2022); the broader transformation of the calligraphic landscape under AI technologies (Jiang et al., 2023); the artistic–technical interplay between traditional ink practices and modern computational models (Lin et al., 2025); technical advances include algorithms for texture-based font recognition (Liu & Zhang, 2024); lightweight models such as PagodaNet for style classification (Zhu & Zhu, 2024); and improved DenseNet architectures for calligraphy font recognition (Wang & Zong, 2023). Additional work has extended calligraphic writer identification to other cultural domains, such as the analysis of Lope de Vega's autographs using AI-based handwriting attribution techniques (González & Cabarrocas, 2025).

Another group of studies examined AI-based recognition of East Asian calligraphy across multiple scripts, particularly Chinese, Japanese, and Korean, such as calligraphic OCR for Chinese calligraphy recognition (Bao et al., 2025); calligraphy Chinese character recognition technology by deep learning and computer-aided technology (Si, 2024); machine learning approaches for identifying handwriting styles (Hu & Wu, 2024); recognition of historical Japanese scripts, particularly Kuzushiji, with surveys and applied studies demonstrating significant progress in deciphering these complex characters (Ueki & Kojima, 2021; Batjargal, 2022); an AI-driven exploration system of emotional and stylistic expressions in calligraphy (Xing, B., et al., 2025); and a machine learning-based calligraphy font recognition system using hog features and support Vector machine (Xu, Liang, & Ling, 2025). Korean calligraphy has received attention through platforms that integrate deep learning to support learners and automate style analysis (Yang, Osman & Sarvghadi, 2025);

A third group of research studies focused on Arabic calligraphy recognition in recent years, reflecting the script's artistic complexity and its diverse cultural applications. These studies demonstrate the breadth of computational approaches applied to Arabic calligraphy and underscored the script's unique visual and structural challenges for AI-based recognition. Several studies have focused on identifying stylistic features in modern contexts, including AI-based analyses of calligraphic characteristics in digital advertisements (Akbar & Hates, 2025); autoencoder models designed to recognize artistic calligraphy types (Al Hmouz, 2020); image and text processing to support the computational reading of Arabic calligraphy (Alsalamah, 2020); texture-based neural networks for style identification (Hussein, 2021); new computational methods for representing and classifying calligraphic styles (Kaoudja, Kherfi & Khaldi, 2021); a systematic review of handwritten Arabic calligraphy generation (Sumayli & Alkaoud, 2025); Deep learning approaches for extracting handwriting from applied calligraphy and historical cursive forms (Zerdoumi et al., 2023); traditional machine-learning and transfer-learning techniques used to analyze calligraphic images (Gürer & Gökbay, 2023); and contemporary styles, soft computing methods for supporting the machine reading of Qur'anic Kufic manuscripts, highlighting the challenges posed by ancient and highly stylized scripts (Zafar & Iqbal, 2022), a handwritten Arabic character recognition technique for machine readers (El-Desouky, et al., 1991); a letters dataset and corresponding corpus of text for machine reading of Arabic

<sup>2</sup> <https://nihad.me/arabic-calligraphy-styles/>

<sup>3</sup> [https://en.wikipedia.org/wiki/Chinese\\_calligraphy](https://en.wikipedia.org/wiki/Chinese_calligraphy)

<sup>4</sup> [https://en.wikipedia.org/wiki/Japanese\\_calligraphy](https://en.wikipedia.org/wiki/Japanese_calligraphy)

<sup>5</sup> [https://en.wikipedia.org/wiki/Persian\\_calligraphy](https://en.wikipedia.org/wiki/Persian_calligraphy)

calligraphy (Salamah, & King, 2018); and the rise and development of the Arabic script from silent stones to the Qur'an's voice (Al Hamad, 2025).

Despite the growing body of research on AI-based recognition of calligraphy across Chinese, Japanese, Korean, and Arabic scripts, the existing literature mostly focuses on a single language, a single script type, or a specific technical approach, with limited attempts to compare challenges and AI model performance across two or more writing systems. Research on East Asian calligraphy tends to emphasize historical scripts and stylistic classification, while studies on Arabic calligraphy often address modern applications, texture-based identification, or ancient manuscripts. This lack of comparative analysis leaves a gap in understanding the performance of AI when confronted with ornamental, non-standard, or highly stylized writing systems. To address this gap, the present study aims to examine AI's performance in recognizing calligraphic texts across Arabic, Japanese, and Chinese. Specifically, the study seeks to (1) examine how Gemini, ERNIE-ViLG and GT decode, transcribe and/or translate Arabic, Japanese, Chinese calligraphic text; and (2) compare the three AI model's performance in terms of accuracy, completeness, error type, consistency and the challenges that each faces in recognizing, decoding and translating Arabic, Japanese and Chinese calligraphic texts.

By bringing together findings from multiple AI models and multiple languages, this study offers a comprehensive perspective on the capabilities and constraints of current AI models when applied to complex calligraphic texts. Studying AI recognition of calligraphic texts is significant for several reasons. First, calligraphy represents some of the most visually complex and culturally meaningful writing systems in the world, and understanding how AI engages with these forms provides insight into the performance of current computational models. Second, improved recognition of calligraphic text has practical implications for cultural preservation, digital archiving, education, and accessibility, particularly for historical manuscripts and artistic works that are difficult for non-experts to interpret. Finally, this study contributes to a broader understanding of how AI interacts with diverse writing systems, highlighting both technological limitations and opportunities for future innovation.

A wide range of stakeholders benefit from this study. Researchers in artificial intelligence can use the findings to better understand the limitations of current models when applied to complex, stylized writing systems. Developers of OCR and handwriting-recognition technologies may draw on the cross-script challenges to design more robust and adaptable systems. Educators, archivists, and cultural institutions working with historical or artistic manuscripts can also benefit from improved recognition methods that enhance accessibility and digital archiving. Scholars in linguistics, digital humanities, and manuscript studies gain insights into how AI tools can support the interpretation of culturally significant calligraphic traditions. Students majoring in art can benefit from a clearer understanding of how AI interprets and sometimes misinterprets artistic scripts, informing both their creative practice and their engagement with digital tools. Translators and translation students also benefit from this study, as improved AI recognition of calligraphic scripts enhances their ability to access, interpret, and translate historical manuscripts, artistic texts, and culturally significant documents. Museum staff may also find value in improved AI-based recognition tools that support the cataloguing, interpretation, and digital exhibition of calligraphic artifacts. Tourists and general audiences visiting museums, heritage sites, or digital collections can benefit indirectly through enhanced accessibility features, clearer interpretive materials, and interactive technologies that make calligraphic traditions more understandable and engaging.

Additionally, this study is significant because it constitutes an addition to a series of studies by the author on the use of AI in translation and education such as: Gaza–Israel war terminology (Al Jarf, 2025b); grammatical terms used metaphorically (Al Jarf, 2025k); zero expressions (Al Jarf, 2025v); Arabic *abu* brand names (Al Jarf, 2025f); denotative and metonymic *abu-* and *umm-* animal and plant folk names (Al Jarf, 2025j); folk medical terms with *om* and *abu* (Al Jarf, 2025t); medical terms (Al Jarf, 2024e; Al Jarf, 2024f); technical terms (Al Jarf, 2021a; Al Jarf, 2016a); human and AI expressions of impossibility (Al Jarf, 2025s); human vs AI translation of chemical compound names (Al Jarf, 2025m); educational polysemes in AI translation of Arabic research articles (Al Jarf, 2025a); Copilot's English translation of contrastive emphatic negation in Arabic discourse (Al-Jarf, 2025i); translations from five languages into English and Arabic by Google Translate (2012–2025) (Al-Jarf, 2025l); AI decoding and interpreting of encrypted Arabic on Facebook and YouTube to evade algorithmic moderation (Al-Jarf, 2025e); Arabic transliteration of borrowed English nouns with /g/ (Al Jarf, 2025c); pronunciation errors in Arabic YouTube videos (Al Jarf, 2025h; Al Jarf, 2025o; Al Jarf, 2025p); editors' perspectives on the publication of AI-generated research articles (Al Jarf, 2025r); Arab instructors' views on AI-generated student assignments (Al Jarf, 2024a); encrypted Arabic on Facebook and YouTube (Al Jarf, 2024c); "sleep" terms (Al Jarf, 2025u); specific linguistic questions that Artificial Intelligence cannot answer accurately (Al-Jarf, 2025q); translations from five languages into English and Arabic by Google Translate (2012–2025) (Al-Jarf, 2025k); electronic translation between Arabic and European languages (Al-Jarf, 2012). Together, these studies illustrate recurring weaknesses in AI's handling of linguistic, cultural, and scholarly tasks, reinforcing the diagnostic framework adopted in this paper.

Furthermore, this study is significant because it constitutes an addition to a series of studies by the author on the transliteration, decoding and translation of shop signs, street signages and linguistic landscapes as pan Arab linguistic and translation errors and

strategies in bilingual linguistic landscapes (Al-Jarf, 2025n); whether Arabic product names as judged by student translators should be definite or indefinite (Al-Jarf, 2024a); whether Arabic and foreign shop names should be translated or not (Al-Jarf, 2024d); semantic and syntactic anomalies of Arabic-transliterated compound shop names (Al-Jarf, 2023); deviant Arabic transliterations of foreign shop names in Saudi Arabia and decoding problems among shoppers (Al-Jarf, 2022a); English language representation in Korean linguistic landscapes (Al-Jarf, 2024b); promotional, sociocultural and globalization issues contributing to the dominance of foreign shop names over Arabic names in Saudi Arabia (Al-Jarf, 2022b); linguistic-cultural characteristics of hotel names in Makkah, Madinah and Riyadh (Al-Jarf, 2021b); teaching English with linguistic landscapes to Saudi students studying abroad (Al-Jarf, 2021c).

## **2. Methodology**

### **2.1 Sample of calligraphic text images**

Three sets of 15 Arabic, 7 Japanese and 10 Chinese calligraphic text images were collected. The Arabic sample consisted of Islamic calligraphy art, mostly verses of the Quran, collected from Google Images. They cover a range of calligraphic styles (as stylized Thuluth, diwani & Kufi), from simple decorative scripts to more complex artistic forms (superimposed and overlapping words, syllables and letters), including circular, oval, rectangular, and freeform layouts, and diverse colors and decorative backgrounds to reflect the diversity of Arabic calligraphy encountered in real-world contexts. The Japanese and Chinese samples were collected from the author's decorative gift boxes and from Google Images. The sample includes a range of writing styles, from formal poetic inscriptions to highly ornamental brushwork. Calligraphic texts include classical Chinese poetry, stylized Japanese kanji, and packaging elements with traditional symbols such as torii gates and seal stamps. The images were randomly selected to reflect real-world contexts in which calligraphy appears outside of formal documents, emphasizing the need for robust recognition systems that can handle non-standard input. All images were sourced from publicly available materials, non-copyrighted decorative items and the author's personal collection of 1 Chinese and 5 Japanese souvenirs. No personal data or sensitive content was used. The study focuses exclusively on AI model performance and does not evaluate or critique cultural or artistic traditions.

### **2.2 Sample of AI Models**

This study adopts a comparative, cross-script design to examine how three AI tools - Gemini, ERNIE-ViLG and Google Translate (GT) - handle calligraphic text in Arabic, Japanese, and Chinese. Gemini was selected for its strong multimodal capabilities and its ability to recognize and interpret calligraphic text across multiple writing systems. Its multilingual training and integrated vision-language architecture enable it to decode stylized Arabic, Japanese, and Chinese characters with high consistency. Using Gemini provides a unified baseline for cross-script comparison and strengthens the methodological rigor of the study. ERNIE-ViLG was selected because it is a Chinese multimodal AI model trained on large-scale native datasets, giving it superior ability to interpret handwritten and stylized Chinese characters within images. Its architecture integrates visual and linguistic understanding, making it more reliable for image-based Chinese text extraction than general-purpose models. GT was selected because it is a widely used, general-purpose translation system with integrated OCR capabilities. Its inclusion provides a practical baseline for evaluating how mainstream tools handle calligraphic Arabic, Japanese, and Chinese text images. Comparing GT with advanced multimodal models highlights the strengths and limitations of everyday AI systems when confronted with stylized or handwritten scripts.

### **2.3 Procedures**

Each calligraphic text image was uploaded individually to Gemini, Ernie and GT. For the Arabic images, Gemini and Ernie, were asked to transcribe the calligraphic text in Arabic script. For the Japanese and Chinese images, Gemini and Ernie were asked to translate the texts in the image to English. Arabic, Japanese and Chinese calligraphic text images were uploaded to GT one by one which was asked to translate the images to English. The same prompts were given to Gemini and Ernie and were used across all images and the three languages.

### **2.4 Evaluation Criteria**

The transcription and/or translations of the images generated by the three AI models were assessed by the author for accuracy (the correct orthographic form for Arabic and whether the translation matches the intended meaning for Chinese/Japanese); completeness, i.e., whether the tool captured the entire text or omitted elements; error type, i.e., hallucinated content, misrecognition of Japanese and Chinese characters, incorrect segmentation, or stylistic misinterpretation); and consistency, i.e., whether the tool produced similar errors across images or scripts. A simple categorical marking system was used (accurate / partially accurate / inaccurate), accompanied by qualitative notes describing the nature of the errors.

Although the author does not speak Chinese or Japanese, the analysis does not rely on linguistic proficiency. Instead, the author functioned as an observational analyst rather than a translator. She compared model behavior, characters and translations, cultural explanations, error types, and hallucinations across the AI models, examining how they interpreted the same visual input, how their translations converged or diverged, and what these patterns reveal about underlying model capabilities. This approach allows for a rigorous evaluation of multimodal AI performance without requiring direct knowledge of the source languages, demonstrating that meaningful cross-linguistic analysis can be achieved through structured comparison rather than linguistic expertise.

To verify Japanese/ Chinese translations to English without knowing Chinese and Japanese, the author performed the following (i) a cross-model convergence which means that If 2 or 3 models independently give the same meaning, this means the translation is very likely correct. For example: if Gemini gives "To know enough is to be always joyful" & Ernie gives "Contentment brings constant happiness". These can be treated as the correct translation. (ii) Identifying the "odd one out". When an AI model gives a translation that is structurally different, thematically unrelated, or introduces new concepts not present in others ... it is likely incorrect. Example: If three models say "Wealth, Prosperity, Longevity, Happiness, Blessing" and one model says "Decision / Last year I wrote a poem," you immediately know which one is wrong. (iii) Use of external verification of famous Chinese phrases, idioms, or poems, using Google search, Wiktionary, Chinese idiom databases or Classical poetry databases. (iv) Internal linguistic consistency checking such as whether the characters rendered match those in the image and whether they are identical for the AI models. If the calligraphy has 4 big characters, it is likely a chengyu (idiom). If the translation is a long paragraph, it is suspicious. If the image looks like a poem, but one model gives a business slogan, then it is wrong. If the image looks like a blessing scroll, but one model gives a martial arts maxim, it is wrong. (iv) Looking for expressions of uncertainty as may be, probably, most likely, or. If the model says "the calligraphy is abstract and highly cursive, making it difficult to read" and offers uncertain interpretations depending on whether the phrase is a standalone artistic piece or a known phrase and translated phrases as a corrupted or stylized version of something and if the model says "for a precise translation, a clearer or more standard calligraphy sample would be needed and do you have additional context, that would help, then the translation is most likely inaccurate as these are signs of semantic projection, not recognition.

For validity and reliability checks, Microsoft Copilot (MC) was used as an independent analytical tool to compare the translations produced by Gemini and ERNIE ViLG for Japanese and Chinese. Although MC is itself an AI model, its role in this study was not to generate translations, but to evaluate and compare the responses of Gemini and Ernie. The author's long experience using MC supports its use as a reliable tool for identifying semantic differences, translation errors, and structural divergences in the interpretation of Chinese and Japanese calligraphic text images. MC was not used to assess the accuracy of Arabic calligraphic transcriptions, as the author is a native speaker of Arabic and is fully familiar with Qur'anic verses and classical calligraphic forms. Arabic accuracy was therefore evaluated directly by the author, while Japanese and Chinese translations were assessed through structured cross-model comparison supplemented by MC.

Furthermore, to evaluate reliability, each image was submitted to the same AI model multiple times in different sessions. Across repeated trials, responses were sometimes identical and sometimes different; however, in all cases the responses remained incorrect, whether in transcription or translation. The models produced different character readings and translations, indicating that OCR interpretation of stylized calligraphy is unsure. This variability was recorded as part of the error analysis.

**2.5 Data Analysis**


The total number of correct responses rendered by Gemini, Ernie and GT out of the 15 Arabic, 7 Japanese and 10 Chinese set calligraphic text images was computed. Additionally, the accuracy of each response to each image by the three AI models was described qualitatively together with MC's comparative analysis and note.








**3. Results**

**3.1 Arabic Calligraphic Text Recognition and Transcription/Translation by Gemini, ERNIE-ViLG, and GT**



The 15 Arabic images show calligraphic representations of 15 verses from the Holy Quran. Data analysis showed that Gemini could correctly recognize and transcribe 12/15 images, Ernie 0/15 and GT .5/15. The 3 AI models varied considerably in their behavior. For each image, Gemini gave the Arabic verse, the source (the āyah and sura number where it is located), English translation, decoding details as artistic composition, visual balance, specific Script style as Thuluth, Jali Thuluth and kufi, artistic composition, decorative border, zoomorphic composition, modern aesthetics) and so on. Although these additions were beyond the scope of the task, they demonstrated Gemini's strong internal alignment between visual recognition and Qur'anic textual knowledge. Ernie failed to decode all the images, and GT gave 1 partially correct response. The following Table, a summary of each model's response to the Arabic images 1-15 are presented.

**Table 1: Gemini, Ernie-ViLG and GT's Responses to the Arabic Calligraphic text Images.**

 <p><b>Arabic Image 1</b></p>	<p>For Arabic Image 1, Gemini correctly identified the verse <i>يَسْمُ اللّٰهُ الرَّحْمٰنُ الرَّحِيْمُ .. اَلَمْ نَشْرَحْ لَكَ .. صَدْرَكَ وَوَضَعْنَا عَنكَ وَرِزْقَكَ الَّذِي اَنْقَضَ ظَهْرَكَ وَرَفَعْنَا لَكَ ذِكْرَكَ فَاِنَّ مَعَ الْعُسْرِ يُسْرًا فَاِذَا مَا اَلْرَّحْمٰنُ الرَّحِيْمُ . اَلْحَمْدُ لِلّٰهِ رَبِّ الْعٰلَمِيْنَ . يَسْمُ اللّٰهُ الرَّحْمٰنُ الرَّحِيْمُ . صِرَاطَ الَّذِيْنَ اَنْعَمْتَ عَلَيْهِمْ غَيْرِ . اِهْدِنَا الصِّرَاطَ الْمُسْتَقِيْمَ . اِيَّاكَ نَعْبُدُ وَاِيَّاكَ نَسْتَعِيْنُ . مَلِكِ يَوْمِ الدِّيْنِ صِرَاطَ الَّذِيْنَ اَنْعَمْتَ عَلَيْهِمْ وَلَا الضَّالِّيْنَ</i>. Likewise, GT produced a nonsensical and incoherent translation</p>
--	--

	<p>that is not even related to any verse: <i>"The ten who are silent and we are not like those who are absent, and I remember you, for now with the fragrance, with the fragrance"</i>.</p>
 <p><b>Arabic Image 2</b></p>	<p>For Arabic Image 2, Gemini correctly identified the verse as <i>فَاتَّهَا لَا تَعْمَى الْأَبْصَارُ وَلَكِنْ تَعْمَى</i>; Ernie-ViLG failed to decode the verse and rendered unrelated Qur'anic excerpts <i>الْقُلُوبُ الَّتِي فِي الصُّدُورِ</i>; <i>بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ</i>; and GT did not recognize it and did not give any response.</p>
 <p><b>Arabic Image 3</b></p>	<p>For Arabic Image 3, Gemini correctly identified the verse in the large font in the middle and in the small fonts in the circle: <i>اللَّهُ لَا إِلَهَ إِلَّا هُوَ الْحَيُّ الْقَيُّومُ لَا تَأْخُذُهُ سِنَّةٌ وَلَا نَوْمٌ لَهُ مَا فِي السَّمَاوَاتِ وَمَا فِي الْأَرْضِ مَنْ ذَا الَّذِي يَشْفَعُ عِنْدَهُ إِلَّا بِإِذْنِهِ يَعْلَمُ مَا بَيْنَ أَيْدِيهِمْ وَمَا خَلْفَهُمْ وَلَا يُحِيطُونَ بِشَيْءٍ مِنْ عِلْمِهِ إِلَّا بِمَا شَاءَ وَسِعَ كُرْسِيُّهُ السَّمَاوَاتِ وَالْأَرْضَ وَلَا يَئُودُهُ حِفْظُهُمَا وَهُوَ الْعَلِيُّ الْعَظِيمُ</i>. Ernie-ViLG failed to recognize the actual verse and gave <i>بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ</i> instead, and GT did not recognize, nor translate the image, producing nothing.</p>
 <p><b>Arabic Image 4</b></p>	<p>For Arabic Image 4, Gemini correctly identified the verse as <i>وَكَيْفَ بِاللَّهِ وَكَيْلًا</i>; Ernie-ViLG failed to recognize the actual verse and instead hallucinated a sequence of unrelated religious phrases: <i>بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ</i>; <i>بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ</i>; <i>بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ</i>; <i>بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ</i>; <i>بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ</i>; and GT produced a nonsensical and incoherent response <i>"God, for God, and me, with a night S"</i>.</p>
 <p><b>Arabic Image 5</b></p>	<p>For Arabic Image 5, Gemini could correctly decode the verse as <i>يَا أَيُّهَا النَّبِيُّ إِنَّا أَرْسَلْنَاكَ وَإِلَيْهَا رُجُوعُكَ وَمَا نَكُودُكَ</i>. On the contrary, Ernie-ViLG failed to recognize the verse in the image and rendered <i>بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ</i> followed by <i>قُلْ هُوَ اللَّهُ أَحَدٌ</i> and GT could not detect the text and rendered nothing.</p>
 <p><b>Arabic Image 6</b></p>	<p>For Arabic Image 6, Gemini could accurately recognize and decode the calligraphic Qur'anic verse as <i>وَمَا أَرْسَلْنَاكَ إِلَّا رَحْمَةً لِّلْعَالَمِينَ</i>, whereas Ernie-ViLG could not recognize the calligraphic Qur'anic verse and gave a random verse <i>ذَاتِ الْبُرُوجِ</i> and GT did not recognize it and rendered nothing.</p>
 <p><b>Arabic Image 7</b></p>	<p>For Arabic Image 7, Gemini correctly recognized and decoded the verse in the image as <i>وَاللَّهُ عَظِيمٌ</i>, but Ernie-ViLG failed to recognize the verse in the image and rendered an unrelated Islamic expression <i>لَعَلَّيْ خُلِقَ عَظِيمٌ</i> and GT did not recognize the verse and rendered nothing.</p>
 <p><b>Arabic Image 8</b></p>	<p>For Arabic Image 8, Gemini successfully recognized and decoded the verse in the big and small fonts as <i>فَتَوَكَّلْ عَلَى اللَّهِ إِنَّكَ عَلَى الْحَقِّ الْمُبِينِ</i>, whereas Ernie-ViLG gave a nonsensical equivalent <i>فَتَكَاثُرُ الْمُلُوكِ</i> and then said: This is a partial Qur'anic verse from Surah At-Takathur (Chapter 102, Verse 1-2), which begins: <i>حَتَّى زُرْتُمُ الْمَقَابِرَ</i>. GT gave a partially correct translation "Trust in God" and a faulty response "God is with the dreamers" for <i>إنك على الحق المبين</i>.</p>



 <p><b>Arabic Image 14</b></p>	<p>For Arabic Image 14, Gemini failed to decode the proverb in the image and gave 3 unmatching verses in three attempts “ بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ ”, “ وَاِنَّكَ لَعَلٰی خُلِقَ عَظِیْمٌ ”, and “ اِنَّ یَتَصَرَّكُمُ اللّٰهُ فَاَلَا ” إن یتصركم الله فلا “ إن یتصركم الله فلا ”. Ernie-ViLG could not recognize the proverb in the image and defaulted to parts of Surat Al-Fatiha “ بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ ”. GT hallucinated and gave the non-sensical equivalent “beard”.</p>
 <p><b>Arabic Image 15</b></p>	<p>For Arabic Image 15, Gemini failed to decode the verse in the image and read it as “ فَاِنَّ مَعَ ” فَأِنَّ مَعَ “ مَعَ الْعُسْرِ يُسْرًا اِنَّ مَعَ الْعُسْرِ يُسْرًا اِنَّ مَعَ الْعُسْرِ يُسْرًا فَإِذَا فَرَغْتَ فَانصَبْ وَإِلَىٰ رَبِّكَ فَارْجَب ” which does not match the verse in the image. Similarly, Ernie-ViLG could not decode the verse and rendered the unrelated verse “ وَمَا أَرْسَلْنَاكَ إِلَّا رَحْمَةً لِّلْعَالَمِیْنَ ”. Here again, GT failed to recognize the verse in the image and returned nothing.</p>

Gemini was able to accurately decode the Arabic calligraphic images because when it processes Arabic calligraphy, it does not rely on standard OCR, which converts pixels into text based on linear patterns, but processes the visual layout and linguistic meaning simultaneously. Gemini explained that it does not simply “see” lines; it understands the artistic geometry of the script and is capable of “mentally unstacking” letters in styles like Jali Thuluth, where characters are layered to fill circular or oval compositions. Even in highly stylized art, Arabic letters follow a specific “skeleton” (Rasm), and it looks for tall, straight strokes of Alif (ا) and Lam (ل) as physical markers to define vertical rhythm and word boundaries, the tails of Noon (ن) or the hooks of Seen (س), which remain constant even when the rest of the letter is stretched (Mashq). In addition, Gemini relies on semantic and contextual probability, because these images are almost always Qur’anic. Gemini reported that once it identifies even two or three distinct words - for example, “خَلَقْنَاكُمْ” - it cross-references them with its internal database of the Qur’an, using predictive completion to verify distorted or compressed letters and effectively solving the calligraphy like a puzzle whose answer key is already known. Its script-style knowledge and noise-filtering abilities that allow Gemini to distinguish between Tashkeel (essential vowel marks) and decorative fillers, and to recognize iconic templates such as the “Ship of Salvation” or the “Circular Medal,” which indicate where the primary text begins and how secondary text wraps around it. Gemini’s 3/15 errors stem from identifiable limitations: extreme compression in circular pieces, where the “jigsaw effect” disrupts the expected logical flow, and abstract textures - such as wood-grain backgrounds - that introduce visual noise and obscure the clean edges of the letter skeleton, leading to misinterpretation.

It is noteworthy to say that Gemini’s strategy in spotting 2 or three words within the composition and then guessing the whole vers is similar to how skilled human readers approach ornate Arabic calligraphy. Many Arab readers struggle to decode highly stylized Qur’anic calligraphy unless they already know the Qur’anic verse by heart. This parallel between human and model behavior highlights that successful decoding of complex calligraphy often depends not only on visual recognition but also on prior knowledge with Qur’anic verses or canonical texts.

On the other hand, Ernie ViLG failed on all Arabic calligraphy images because it lacks any meaningful capacity to decode Arabic script and is not trained on Arabic calligraphic forms. Its responses show a consistent pattern of hallucination: when the visual layout appears religious, it defaults to memorized Qur’anic verses; when it cannot anchor the strokes, it produces generic moral statements; when the calligraphy is highly stylized, it renders nonsensical strings; and across repeated submissions, it varies the hallucination even for the same image. This behavior reflects Ernie’s reliance on Chinese-centric visual priors rather than true character-level recognition. Arabic calligraphy differs fundamentally from Chinese script—its cursive nature, positional letter shapes, diacritics, and artistic styles such as Thuluth, Diwani, and Naskh require specialized OCR training and linguistic grounding that Ernie does not possess. As a result, Ernie cannot parse the rasm, distinguish ornamentation from text, or reconstruct the intended sequence of letters. Instead, it projects familiar religious or cultural content based on superficial visual cues. Unlike Gemini,





which benefits from broader multilingual training, Ernie’s architecture and training data are overwhelmingly optimized for Chinese language and culture, leaving it unable to generalize across scripts with radically different stroke logic. Consequently, Ernie is expressive within its own cultural domain but functions as a non-reader of Arabic calligraphy, which explains its score of 0/15 on the Arabic dataset.

Similarly, GT’s responses reflect its inability and limitations to process stylized Arabic calligraphy from images, especially when the text is ornate or embedded in decorative design.

**3.2 Japanese Calligraphic Texts Recognition and Translation by Gemini, ERNIE-ViLG, and GT**

Data analysis showed that out of 7 Japanese calligraphic text images, Gemini correctly translate all images; Ernie Ernie-ViLG 4/7 and GT failed to render a response and gave 6 partial, fragmented incoherent responses. Gemini’s responses are characterized by defining the characters that constitute the whole text and explaining the meaning of each, providing the translation, cultural explanation, long linguistic breakdowns, object descriptions, symbolic interpretation, historical background, script/style commentary and philosophical commentary. In other words: Gemini behaved like a cultural encyclopedia, not just a translation model. Ernie-ViLG’s responses consist of partial readings, uncertain guesses, and culturally-plausible hallucinations. It often misreads characters, offers multiple alternative interpretations, defaults to Buddhist or Chinese-centric phrasing, and reconstructs meaning based on visual or thematic cues rather than accurate character recognition. Its responses mix fragments of real text with invented explanations, showing semantic projection rather than transcription. GT’s responses consist of literal, surface-level word extractions with no cultural context, no genre awareness, and frequent OCR errors. It produces fragmented, incoherent phrases, often missing structure or meaning entirely. GT behaves like a raw dictionary lookup tool, giving isolated words rather than translations, and failing completely when the calligraphy is stylized or artistic. For more details, responses of the three AI models are reported in Table 2.

**Table 2: Gemini, Ernie-ViLG and GT’s translation of Japanese Calligraphic Text Images**

 <p><b>Japanese Image 1</b></p>	<p>For Japanese Image 1, Gemini accurately identified all the characters the image and identified the object as a Torii gate. It translated 開運 as “good fortune,” 厄除 as “warding off evil,” and おみくじ as “fortune slip.” Like Gemini, Ernie-ViLG gave similar translations with slight phrasing differences. GT gave a fragmented word-for word translations “torii. Kaie. Luck. Remove. Good luck. Comes with a fortune slip” without coherence.</p> <p>MC noted full semantic convergence with complementary framing between Gemini and Ernie, highlighting how multimodal AI models can align in meaning while offering distinct interpretive lenses.</p>
 <p><b>Japanese Image 2</b></p>	<p>For Japanese Image 2, Gemini translated the phrases correctly as follows: (千栄作 -Made by Chie (artist’s name); 通商産業大臣賞受賞 - Ministry of International Trade and Industry Award, and 創作こけし- Creative Kokeshi. Ernie-ViLG explained the text on the object. It often showed uncertainty by offering multiple possible readings and translations as: 日本通商産業大臣 is an older or formal term, but in modern Japanese, it would be 日本通商産業大臣 meaning “Minister of International Trade and Industry of Japan.” However, on consumer products, this is often part of a certification or approval mark indicating that the product was approved or recognized by the relevant Japanese authorities. “千栄作” –a brand or maker’s name. It can also be read as “Sen-eisaku” in Chinese, but the meaning is not immediately clear. Then Ernie-ViLG gave the following full Translation/Interpretation: This is a Japanese-made product, possibly with some form of official recognition or approval. The brand name is 千栄作, and it was manufactured in Japan. GT gave a fragmented, incoherent and partially correct translation “Minister of International Trade and Industry, Mr. Kishi. Creation Award-winning author”.</p> <p>MC noted that both Gemini and Ernie-ViLG recognized the label’s connection to traditional craftsmanship and official certification. However, both models misread the central maker’s name 千察作, interpreting it as 千栄作 or 千榮作, likely due to stylization. Gemini framed the item as a modern artistic Kokeshi with a ministry award, while Ernie-ViLG interpreted it as a vintage consumer product label. This example illustrates contextual convergence with textual</p>

	<p>drift, where models agree on cultural and institutional framing but fail in precise character recognition under stylized conditions.</p>
 <p>Japanese Image 3</p>	<p>For Japanese Image 3, Gemini identified the image as a Japanese Creative Kokeshi doll and translated the phrases 近代創作こけし Modern Creative Kokeshi; 紅花 Safflower 通商産業大臣賞 Minister of International Trade and Industry Award; 農林水産大臣賞 Minister of Agriculture, Forestry and Fisheries Award; 受賞作家 Award-winning artist; Similarly, Ernie-ViLG identified the image as a traditional Japanese or Chinese calligraphic piece and translated the phrases as follows: 近代創作 Modern creative work; 紅花 Safflower / Red Flower; 通商産業大臣賞 – Minister of International Trade and Industry Award; 農林水産大臣賞 Minister of Agriculture, Forestry and Fisheries Award; 受賞作家 Award-winning artist; Ernie translated the phrases accurately but framed the label as a literary or exhibition piece rather than a Kokeshi doll tag. GT correctly translated the Japanese phrases as follows: 近代創作こけし Modern creative kokeshi; 紅花 Safflower; 通商産業大臣賞 Minister of International Trade and Industry Award; 農林水産大臣賞 Minister of Agriculture, Forestry and Fisheries Award; 受賞作家 Award-winning artist. But GT’s translations are fragmented and incoherent. GT also added an unrelated phrase (“lock up”).</p> <p>MC noted that this example shows semantic convergence but divergent object classification. Gemini aligned with Japanese craft traditions, while Ernie generalized toward textual artifacts.</p>
 <p>Japanese Image 4</p>	<p>For Japanese Image 4, Gemini said the image features the Japanese poem Iroha (いろは), The script used is a combination of Kanji: Chinese characters &amp; Hiragana: A native Japanese phonetic script. The poem is deeply rooted in Buddhist philosophy regarding the transience of life. It translates to: “<i>Though their colors are fragrant, the flowers will eventually scatter. Who in our world can remain forever? Crossing today the deep mountains of existence, I shall no longer see shallow dreams, nor be intoxicated by them</i>”. Ernie-ViLG indicated that the text is written in Japanese calligraphy, using kanji characters. The phrase 我々は常に色が散りぬる世を誰ぞ知らん translates as “Who among us knows the world where colors always fade?” GT rendered a simplified and partially correct translation “The colors are beautiful but fade away, but who is the common man in this world?” The phrasing is poetic but structurally incoherent.</p> <p>MC noted semantic alignment with textual drift: Both Gemini and Ernie captured the emotional and philosophical core, but Gemini anchored its reading in the historical Iroha, while Ernie offered a plausible but divergent reconstruction.</p>
 <p>Japanese Image 5</p>	<p>For Japanese Image 5, Gemini interpreted the image as a martial arts maxim and translated the phrases as follows: 修行常待 – Training is always present; 越舟 (artist’s name); 初段 First-degree black belt; 梶次 Likely the practitioner’s surname. Gemini framed the image as a Dōjō inscription emphasizing lifelong discipline. Ernie-ViLG identified the text as a Buddhist phrase with uncertain structure and offered multiple rearrangements and interpretations, signalling low confidence and defaulting to Buddhist vocabulary. The standard order for a meaningful phrase would likely be: “肉身菩提 常行” is less common; a more typical arrangement might be 菩提常行 肉身 – but this still feels fragmented. A More Likely Interpretation: The calligraphy appears to be “常行 肉身菩提” or possibly “菩提常行. 肉身” as separate lines, but the full meaning isn’t immediately clear. And went on and on. It gave what it considered a plausible translation based on the characters “<i>The physical body’s enlightenment comes through constant practice.</i>” GT gave a partial translation: “regular practice .. newbie to taojue”. The translation is incoherent, vague and fragmented</p> <p>MC added that both Gemini and Ernie-ViLG misread the characters, but each did so through a distinct cultural lens: Gemini projected a Zen/Dōjō maxim and Ernie projected a Buddhist philosophical phrase. This is a case of cultural projection under stylization, where models infer meaning from visual and thematic cues rather than accurate character recognition.</p>

	<p>For Japanese Image 6, Gemini correctly translated the large black vertical text to Kumadori (隈取); the small vertical text to Tenugui Luncheon Mat (手ぬぐいランチオンマット, and identified the bottom text ©SHOCHIKU as a production company in Tokyo. On the contrary Ernie-ViLG mistranslated the large, bold characters 妖艶化 as "bewitching transformation" or "enchanted makeover." It misread and partially mistranslated the Left side text to "Kinukai Lunch Mat" and fabricated a brand name "Kinukai" from the word てぬぐい. GT translated it as Tenugui placemat. SHOCHIKU. GT's translation is minimal and literal, often limited to one or two surface-level words without context.</p> <p>MC noted that this example illustrates how stylized packaging can trigger semantic hallucination, with AI models projecting culturally plausible but incorrect readings. Gemini's interpretation was grounded in textual fidelity, while Ernie-ViLG relied on aesthetic inference, reinforcing the need for robust character recognition in multimodal analysis. MC's analysis highlights how stylized packaging can trigger semantic hallucination in some models, especially when character recognition is weak.</p>
	<p>In Japanese Image 7, Gemini identified the yellow cloth as a traditional Japanese imagery and text related to Kabuki theater. It translated the Main Text in White Characters to Kumadori (隈取); the Small Vertical Text (Bottom Left) to Shochiku Co., Ltd. It explained that the item is likely a Tenugui (traditional hand towel) or a Furoshiki (wrapping cloth) sold at a Kabuki theater. Ernie-ViLG indicated that the text 鳳凰牌 is repeated multiple times across the fabric, identified it as Chinese and translated it to "Phoenix Brand." GT did not recognize it and gave nothing.</p> <p>MC indicated that Gemini correctly identified the cloth as Japanese Kabuki merchandise, translating 隈取" as Kumadori and noting Shochiku Co., Ltd. Ernie misclassified it as Chinese, interpreting phoenix motifs and hallucinating 鳳凰牌 Phoenix Brand, despite the clear Japanese branding. GT gave no response. MC noted that stylized packaging can trigger cultural misclassification when models rely on visual patterns over textual anchors.</p>

The above data analysis shows that Gemini is the most contextually reliable model. Even when it misreads characters, it reconstructs a meaning that fits the cultural and linguistic domain of the image, whereas Ernie is linguistically creative but unstable. It often "hedges" with several interpretations, revealing low confidence in character recognition. Its translations are sometimes meaningful but rarely faithful to the actual text. On the contrary, GT provides partial translations at best - correct lexical fragments without syntactic or semantic integration. It is useful only for confirming isolated word meanings, not for interpreting full inscriptions.

Across the Japanese calligraphy and label images, Gemini and Ernie converge in meaning, but diverge in textual fidelity. They often converge on broad themes, such as impermanence, discipline, or craftsmanship, but diverge sharply in character-level accuracy and cultural framing. This divergence illustrates how multimodal models may rely on visual pattern recognition over textual anchors, leading to culturally biased misclassification. The case highlights the need for accurate script recognition and contextual grounding when interpreting stylized East Asian calligraphy and labels.

GT consistently produced literal, word-for-word translations with no cultural explanation, no genre awareness, and no interpretive reasoning, behaving essentially like a raw dictionary tool rather than a multimodal translator. It performed reasonably well only when the text is standard, high-contrast, and non-stylized—such as modern printed characters, short phrases, or common Chinese idioms like 福, 壽, and 祿. However, GT struggled severely with stylized or artistic calligraphy, frequently producing OCR errors, nonsense strings, or comically incorrect responses. It showed no ability to distinguish between poems, idioms, blessings, or decorative inscriptions, and it lacked object recognition or contextual inference. GT performed worst with cursive scripts (草書, 行書), classical poetry, artistic brushwork, vertical layouts, and Japanese texts mixing kana with kanji, as well as culturally loaded idioms that require interpretive knowledge. Its limitations make it unreliable for complex or stylized calligraphy, but precisely because of its literal, surface-level behavior, GT serves as an ideal baseline model in this study—providing a clear contrast to the more advanced multimodal reasoning demonstrated by Gemini.

Together, these examples demonstrate that accurate Japanese calligraphy interpretation requires visual recognition, deep linguistic grounding, cultural knowledge, and genre awareness capabilities that only Gemini consistently exhibited.


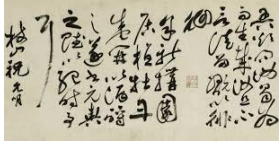
### 3.3 Chinese Calligraphic Text Recognition, Decoding and Translation by Gemini, Ernie-ViLG And GT

Data analysis has shown that out of the 10 Chinese calligraphic text images, Gemini translated 9/10 correctly, Ernie-ViLG gave 3/10 correct translations and GT gave 0/10 fully correct translations, no responses to 4/10 and partial, fragmented, incoherent translation to 6/10. Gemini’s responses consist of character-level identification, full translations, literal and cultural explanations, poetic or idiomatic reconstructions, author or source recognition, script-style notes, and breakdowns of individual characters or phrases. It often includes contextual information such as historical background, genre, or cultural significance. Ernie-ViLG’s responses consist of partial character readings, multiple possible interpretations, alternative phrase reconstructions, stylistic observations, and contextual guesses based on visual cues. It often provides several candidate readings, brief translations, and comments on script difficulty or ambiguity. GT’s responses consist of isolated words, literal lexical fragments, partial phrase extractions, and OCR-based text snippets. When text is detected, it provides short, direct word-level translations without additional explanation, context, or structural integration.

**Table 3: Gemini, Ernie-ViLG and GT’s Translation of Chinese Calligraphic Text Images**

 <p><b>Chinese Image 1</b></p>	<p>For Chinese Image 1, Gemini translated 知足常樂 to "Contentment brings constant happiness." It gave a character breakdown: 知 To know /To be aware of; 足 Enough / Sufficient /Satisfied; 常 Always/Constantly/Often; 樂 Joy/Happiness/ Pleasure. Ernie-ViLG described the calligraphy as abstract and highly cursive, making it difficult to read with certainty. It offered uncertain interpretations depending on whether the phrase is a standalone artistic piece or a known phrase such as 春盡誰憐 or 春盡誰憐 (When spring ends, who will pity it?); translated phrases as a corrupted or stylized version of something like: "春盡誰憐" (Chūn jìn shuí lián) – "When spring ends, who will pity [it]" and treated the image as expressive art rather than a literal message, signalling low confidence in textual recognition. For a precise translation, Ernie asked for a clearer or more standard calligraphy sample would be needed and If the author has additional context, that would help. GT could not detect text and gave nothing.</p> <p>MC noted semantic divergence caused by stylization. Gemini prioritized semantic recognition. It identified the phrase 知足常樂 and offered a precise character breakdown and philosophical context rooted in Taoist thought. But Ernie-ViLG interpreted the calligraphy as abstract and difficult to read, suggesting uncertainty, but ultimately framing the piece as expressive rather than literal.</p>
 <p><b>Chinese Image 2</b></p>	<p>For Chinese Image 2, Gemini identified the five large characters as the "Five Blessings" in Chinese culture and translated them as: 福 (Fortune/Good Luck; 祿 Prosperity/Success; 壽 (Shòu): Longevity/ Long Life; 禧 Happiness/Joy; 財 (Wealth/Riches. It also decoded the smaller inscription 時在戊戌年 created in the Year of the Earth Dog (2018); 冬 [Name]; 書 - written in the winter by [Artist's Name]. Ernie-ViLG also correctly translated the phrase 福祿壽喜財 to "Fortune, Prosperity, Longevity, Happiness, Wealth". It identified the smaller text at the bottom as a signature or seal, and possibly the date or a poetic phrase, but it considered it not clearly legible in this image. GT rendered a partial translation (Wealth, longevity, fortune and fortune) which are semantically aligned but structurally incoherent.</p> <p>MC noted partial convergence with divergence in depth: Gemini and Ernie-ViLG succeeded on culturally standard content but differed in precision. Gemini decoded the inscription linguistically; Ernie-ViLG deferred to visual caution. The case illustrates high agreement on surface meaning with divergence in interpretive depth.</p>
 <p><b>Chinese Image 3</b></p>	<p>For Chinese Image 3, Gemini identified the poem as Du Fu’s “Quatrain” (絕句) and translated it as follows: “Two golden orioles sing in the green willows, A row of white egrets flies into the blue sky. My window frames the thousand-year snow of the Min Mountains, By my door are moored boats from distant Eastern Wu.” Gemini offered a fluent, poetic translation with smooth phrasing and cultural grounding. Ernie-ViLG also recognized the text as a famous couplet from classical Chinese poetry. It reads 兩個黃鸝鳴翠柳;行白鷺上青天。;窗含西嶺千秋雪;門泊東吳萬里船 which translate to "Two golden orioles sing amid the green willows, A row of white egrets ascends into the blue sky, The window frames the eternal snows of</p>

	<p>Western Ridge, The gate parks boats from Eastern Wu, a thousand miles away." GT gave a hallucinated and incoherent translation "Two flutes, yellow rooster crows in the dark, A line of egrets flies up to the blue sky, meeting at the window, cannons crackling in the autumn, Tangmen guards against intentional attacks. Meiwan Sable., which are unrelated and structurally broken.</p> <p>MC noted full convergence between Gemini and Ernie-ViLG: Both models accurately identified the poem and conveyed its literary context and emotional resonance, differing only in stylistic choices. This case illustrates high cultural awareness and semantic precision under ideal conditions.</p>
 <p>Chinese Image 4</p>	<p>For Chinese Image 4, Gemini identified the idiom 萍水相逢 and translates as "To meet like drifting duckweed on the water." It preserved the poetic imagery and offered a literal rendering. Ernie-ViLG also recognized the idiom 萍水相逢 and gave multiple interpretations "Meet by chance like driftwood and water" or more naturally in English: "Meet by chance" or "a chance encounter." Ernie's translations reflected both semantic recognition and interpretive flexibility. GT failed to translate it and rendered nothing.</p> <p>MC noted full convergence with stylistic variation and interpretive layering. Gemini and Ernie correctly identified the idiom and conveyed its meaning, with Gemini emphasizing metaphor and Ernie-ViLG offering layered interpretations. Both translations are correct because this phrase is common in Chinese idioms, easy to recognize visually, frequently used in literature and calligraphy and well-represented in training data for both models.</p>
 <p>Chinese Image 5</p>	<p>For Chinese Image 5, Gemini identified the text as the idiom 萍水相逢 and translated it to: "To meet like drifting duckweed on the water". It provided a precise character breakdown: 萍 Duckweed; 水 Water; 相 Each other / Mutually; 逢 To meet / To encounter. It emphasized the poetic metaphor. Ernie-ViLG indicated that the text is written in a stylized, abstract form of Chinese calligraphy, which makes it somewhat challenging to decipher. However, it appears to be a creative or artistic rendition of the phrase 大富貴 每日金運 which can be translated as: "Great wealth and prosperity; daily golden fortune." Ernie broke it down into: 大富貴 "Great wealth and prosperity" or "immense riches and honor; 每日 "Every day" or "daily"; 金運 (jīn yùn) – "Golden fortune" or "good luck with wealth." The calligraphy is bold and expressive, emphasizing the themes of abundance and luck. The abstract style adds an artistic flair, making it more of a decorative piece than traditional calligraphy. GT could not detect text and yielded nothing.</p> <p>MC noted, Gemini and Ernie-ViLG interpreted the same image in completely different ways. Gemini identified the idiom 萍水相逢 ("to meet like drifting duckweed on water"), offering a precise breakdown of each character and its poetic meaning, while Ernie-ViLG read the image as 大富貴 每日金運 ("great wealth and daily fortune"), emphasizing decorative themes of prosperity. This divergence highlights how stylization and cultural priors influence model interpretation. Gemini leaned toward literary idioms, whereas Ernie-ViLG favored auspicious phrases common in visual training data. The contrast reveals how multimodal models may activate distinct semantic pathways when faced with ambiguous or expressive calligraphy.</p>
	<p>For Chinese Image 6, Gemini identified the four large characters as representing a well-known philosophical idiom. The phrase is 知足常樂 translates to "Contentment brings constant happiness." Or more literally: "To know [when one has] enough is to be always joyful." Gemini broke the phrase down into: 知 To know/to be aware of, 足 enough/ sufficient/satisfied, 常 always/constantly/often, 樂 joy/happiness/pleasure. Ernie-ViLG identified the characters are 常樂知足 which translates to "Always content, satisfied with what one has." Breaking it down: 常樂 "Always happy or constant joy." 知足 "Content with what one has or satisfied." The phrase reflects a mindset of gratitude and avoiding excessive desire. GT produced a vague and partial response: "Always happy to know" - semantically close but structurally incoherent.</p>

<p><b>Chinese Image 6</b></p>	<p>MC noted semantic convergence with structural variation: Both models correctly conveyed the message of gratitude and sufficiency, yet diverged in character order and framing. The case highlights how multimodal models may agree on meaning while differing in textual parsing — a valuable nuance in stylized script interpretation.</p>
 <p><b>Chinese Image 7</b></p>	<p>For Chinese Image 7, Gemini identified the text a poem by Du Fu, titled "Quatrain" (绝). The poem is read from top to bottom, starting from the far-right column and translated it fluently: Two golden orioles sing in the green willows, A row of white egrets flies into the blue sky, My window frames the thousand-year snow of the Western Peaks, By my door are moored boats from the distant Eastern Wu. Ernie-ViLG said the text in the image is written in Chinese cursive script (草书), which is highly abstract and fluid, making precise reading challenging. However, based on the brushstrokes and common phrases, it appears to be a poetic or artistic expression. It is assuming the characters are: 每作勢，鸞蛟出，筆飛是也。and gave uncertain translations: 每作勢- "Every time [one] assumes a posture" or "Whenever [the brush] takes form." 鸞蛟出- "A phoenix and a dragon emerge". 筆飛是也 - "The brush flies - that is it!". It gave an assumably full Poetic translation: "Whenever the brush takes form, a phoenix and dragon emerge - the brush flies, and that is it!" GT failed to detect or translate the text.</p> <p>MC noted semantic divergence due to stylization and training bias: Gemini demonstrated literary recognition and textual grounding, whereas Ernie-ViLG defaulted to decorative interpretation. This case illustrates that AI model identity (e.g., being "Chinese") does not guarantee calligraphic literacy, especially under expressive brushwork.</p>
 <p><b>Chinese Image 8</b></p>	<p>For Chinese Image 8, Gemini identified the text as Du Fu's poem 闻官军收河南河北 "Hearing of the Recovery of the Imperial Capital by Government Forces" and translated it fluently to: <i>Suddenly it is heard that our armies have retaken the north, When the news reached me, tears drenched my clothes. I look at my wife and children, seeing their grief vanish, I pack my books at random, wild with joy. Singing loud in the light of day, we shall drink our fill, With the green spring as our companion, it's time to go home. It recognized the layout and matched the stylized script to a known literary source.</i> Ernie-ViLG said that the text in the image is written in Chinese cursive script (草书), which is highly stylized and abstract, making it challenging to read. It identified it as a Taoist reflection. It added that after careful examination, the characters appear to be: 無為而治，不言之教，牡丹花之為花也，則其華乎？and gave a possible translation: 無為而治 "Govern by doing nothing that goes against nature"; 不言之教 "Teaching without words"; &amp; 牡丹花之為花也，則其華乎？ "As for the peony being a flower, is it not magnificent?" GT could not detect text and could not translate it.</p> <p>MC noted this as one of the most striking divergence cases in the study: Gemini leaned toward historical-literary recognition, while Ernie-ViLG activated philosophical-poetic inference. The case illustrates how cursive calligraphy (草书) can trigger fundamentally different semantic pathways depending on model training and interpretive bias, revealing the limits of multimodal consistency under expressive brushwork.</p>
	<p>For Chinese Image 9, Gemini gave two conflicting interpretations: First, it claimed the text was Du Fu's poem 闻官军收河南河北 ("Hearing of the Recovery of the Imperial Capital"), It translated it as: <i>Suddenly it is heard that our armies have retaken the north, When the news reached me, tears drenched my clothes. I look at my wife and children, seeing their grief vanish, I pack my books at random, wild with joy. Singing loud in the light of day, we shall drink our fill, With the green spring as our companion, it's time to go home. Then shifted to Wang Xizhi's Preface to the Orchid Pavilion Gathering. It reads: "In the ninth year of Yonghe, in the year of Guichou, in the early part of late spring, we gathered at the Orchid Pavilion in Shanyin, Kuaiji. Neither matches the visible characters.</i> Ernie-ViLG also said that the text is written in Chinese cursive script (草书), which is highly stylized and abstract, making precise reading challenging. However, based on the brushstrokes and common phrases, it attempted an interpretation of</p>

 <p><b>Chinese Image 9</b></p>	<p>并送请·兄查可即复·送请查复·并此致候·兄案概是号与" that is more natural contextual translation: This seems to be part of a formal or semi-formal letter, possibly from historical correspondence. A plausible interpretation could be: "I am sending this for your review; please check and reply promptly. I am also sending this for your examination and response. With this, I send my regards. The case in question is likely related to [something]." Ernie broke it down to: 并送请 – "I am also sending this for your request/review."; 兄查可即复 "Brother, please check and reply immediately."; 送请查复 "Sending for your examination and reply."; 并此致候 "With this, I send my regards"; 兄案概是号与 "The case in question is likely related to [this number/reference]." The last part is unclear due to cursive abstraction. GT produced a fragmented and incoherent response: "Send for investigation. Eastern Beauty Wu Sochi".</p> <p>MC noted this as a case of dual hallucination: Gemini hallucinated based on poetic priors; Ernie-ViLG hallucinated based on bureaucratic ones. The actual text contains literary references such as 谢郢集, 宋玉, and 琴书萧然, indicating a poetic dedication or preface. Neither model performed accurate character recognition. The case highlights how stylized cursive script (草书) triggers divergent semantic pathways and remains a major challenge for multimodal AI.</p>
 <p><b>Chinese Image 10</b></p>	<p>For Chinese Image 10, Gemini identified the object as a Chinese Zisha (purple clay) tea mug and interpreted the inscriptions as classic poetic phrases and a famous landmark reference. The text reads from right to left and translated 东方明珠 as Oriental Pearl; 德泽万物生光辉 "Bountiful grace spreads to all things, bringing radiance to the world"; 庚辰 (Gēng Chén), a specific year in the Chinese sexagenary cycle (year 2000 or 1940); 福 (Fú), "Good Fortune."; 长歌行 <i>Chang Ge Xing</i> "Han Dynasty poem title". Gemini framed the mug as a celebration of prosperity, nature's beauty, and good fortune. Ernie-ViLG gave a very brief response and translated the characters as "The orient is as bright as pearls. The sun nourishes all things and shines with vitality." The visible characters on the mug are: 东方咏洪, 应是瀑布, 物美宜. These lines suggest poetic phrasing, possibly describing the beauty or abundance of the East, with references to "grace" (泽), "distribution" (布), and "suitability" or "quality" (宜). The phrase 咏洪 could imply "singing of grandeur" or "praising vastness." Ernie's gave the following full translation: "The orient is as bright as pearls. The sun nourishes all things and shines with vitality." GT produced a fragmented response: "Food Tower" - semantically unrelated and structurally incoherent.</p> <p>MC noted thematic convergence between Gemini and Ernie-ViLG. Both recognized the poetic and auspicious tone, but in different phrasing. This example shows high agreement on meaning and object context, with divergence in stylistic rendering. Ernie is semantically elegant, but does not match the actual characters. It seems to have paraphrased the characters based on visual style and thematic expectations rather than accurate character recognition.</p>

Across the Chinese calligraphic dataset, Gemini was the most accurate and contextually grounded model. It consistently recognized characters even under moderate stylization, identified idioms and classical poems, and matched cursive or semi-cursive forms to known literary sources. Gemini's translations were fluent, culturally informed, and supported by detailed character breakdowns, showing strong multimodal integration between visual recognition and linguistic knowledge. Its few errors occurred only under extreme cursive abstraction, but even then, Gemini attempted to reconstruct meaning using literary priors. Overall, Gemini demonstrated the highest fidelity to the actual text and the strongest cultural and historical grounding.

Ernie-ViLG showed partial success, performing well on common idioms, standard phrases, and culturally familiar expressions, but struggling with stylized or expressive calligraphy. Its responses often included multiple possible readings, hedged interpretations, and thematic guesses based on visual cues rather than accurate character recognition. In several cases, Ernie reconstructed plausible but incorrect poetic or philosophical phrases, revealing reliance on cultural priors rather than textual fidelity. While Ernie occasionally converged with Gemini on meaning, it diverged sharply in accuracy, especially under cursive script (草书). Ernie is therefore a moderately capable model with unstable performance under stylization.

Although Ernie-ViLG is a Chinese model, it performed poorly on Chinese calligraphy because its architecture is not designed for character-level recognition but for image generation and cultural interpretation. Calligraphy—especially cursive forms like 草书—is a visual art with merged strokes, omitted radicals, and stylistic distortions that even trained human readers struggle to decode, and Ernie lacks the OCR-style mechanisms needed to parse these shapes as text. Instead, it relies heavily on semantic priors, inferring meaning from the overall visual impression rather than the literal strokes, which leads it to default to familiar themes from its training data—Taoist sayings, auspicious blessings, poetic clichés, or bureaucratic formulas—whenever the script becomes ambiguous. Stylization further breaks its recognition pipeline, causing it to describe mood or brush energy rather than identify characters, a pattern reinforced by training bias toward decorative inscriptions rather than labelled cursive calligraphy. In short, Ernie is culturally expressive but not textually precise, and its identity as a “Chinese model” does not grant calligraphic literacy; its errors reflect architectural limits, training gaps, and the inherent difficulty of reading stylized calligraphy.

GT performed the weakest on Chinese calligraphy, failing to detect text in many images and producing fragmented, incoherent, or unrelated responses when it did. Its translations were limited to isolated lexical fragments with no syntactic structure, cultural context, or genre awareness. GT could only handle clean, printed, high-contrast characters and collapsed entirely under cursive, artistic, or expressive brushwork. Because it lacks multimodal reasoning and relies solely on OCR-based extraction, GT is unsuitable for interpreting calligraphy. However, its literal, surface-level behavior makes it a useful baseline for contrasting the more advanced multimodal capabilities of Gemini and Ernie-ViLG.

## 4. Discussion

### 4.1 Summary of Results

Results of the current study revealed that the three AI models’ ability to recognize, transcribe and translate calligraphic text images in Arabic, Japanese and Chinese varied. Across all three languages, Gemini is the most consistently reliable model. It recognizes characters with high accuracy in non-extreme stylization, identifies idioms, poems, Qur’anic phrases, proverbs, and classical texts, and provides layered explanations—linguistic, cultural, historical, and stylistic. In Arabic, it distinguishes between Qur’anic verses, proverbs, and decorative scripts; in Japanese, it identifies objects (torii, kokeshi, tenugui) and matches stylized text to known poems; in Chinese, it recognizes classical poetry, idioms, and literary sources. Its main weakness is high-prestige hallucination: when scripts become extremely cursive or ambiguous, Gemini sometimes defaults to famous texts with great confidence. But overall, it demonstrates the strongest multimodal grounding across all three writing systems.

Ernie-ViLG show cultural fluency but unstable textual fidelity across all three languages. In Arabic, it often misreads letters but produces. In Japanese, it frequently projects meaning based on genre expectation - Buddhist phrases, auspicious packaging, or martial arts maxims - rather than reading the actual characters. In Chinese, it performs best on common idioms and standard scripts but collapses under cursive calligraphy, generating elegant but unrelated poetic or philosophical reconstructions. Ernie’s pattern is consistent: it recognizes *themes* (poetry, blessings, philosophy) but struggles with *literal character recognition*, relying heavily on cultural priors.

GT is the weakest across Arabic, Japanese, and Chinese calligraphy. It handles only clean, printed, high-contrast text and fails almost entirely on stylized, artistic, or cursive scripts. In Arabic, it produces isolated word fragments with no syntactic or semantic coherence. In Japanese, it extracts random nouns or partial phrases without context. In Chinese, it rendered incoherent or unrelated strings, often failing to detect text at all. GT has no genre awareness, no cultural reasoning, and no interpretive capacity. It functions as a dictionary-level OCR tool, making it unsuitable for calligraphy but useful as a baseline for comparison.

### 4.2 How and Why Gemini, Ernie-ViLG and GT Process Calligraphy Differently

To understand why the three AI models behaved so differently, it is necessary to consider both how modern AI systems process calligraphy in general and how each model applies its own recognition strategy.

Modern AI models process calligraphic text through a combination of visual feature extraction and linguistic prediction. Contemporary neural models use deep convolutional and transformer-based encoders to interpret strokes, curves, and spatial patterns holistically. This allows them to infer characters even when the calligraphy is stylized, elongated, or artistically distorted. However, calligraphic scripts as Arabic thuluth, Chinese cursive, and Japanese semi-cursive pose unique challenges because they break the standard rules of spacing, baseline alignment, and letter shape consistency. Successful decoding therefore depends on visual recognition and the model’s ability to match ambiguous shapes to known verses, idioms, or poetic structures.

In this study Gemini, Ernie and GT’s accuracy in recognizing Arabic, Japanese and Chinese calligraphic text images varied because they approach calligraphic text in different ways. Each model relies on a different recognition strategy. Gemini does not truly “read”



calligraphy stroke-by-stroke; instead, it uses broad multilingual training, pattern matching, and semantic retrieval to infer which classical poem, proverb, or idiom the visible strokes most closely resemble. This allows it to succeed when the calligraphy corresponds to well-known texts, even if the characters are partially stylized or incomplete. Ernie-ViLG, by contrast, struggles with literal recognition and instead interprets calligraphy through cultural priors: when it cannot decode the strokes, it defaults to common themes in Chinese art—Taoist sayings, auspicious blessings, or bureaucratic formulas - producing elegant but often inaccurate paraphrases. GT relies almost entirely on OCR designed for printed or clean handwritten text, so when confronted with artistic calligraphy—where strokes merge, distort, or follow historical styles- it fails to segment characters and renders random fragments or nonsense. In essence, Gemini recognizes calligraphy through *literary pattern inference*, Ernie through *cultural projection*, and GT through *rigid OCR*, which explains their sharply different performance across the current dataset.

### **4.3 Why Gemini Outperformed Ernie and GT in Arabic, Japanese and Chinese Calligraphy Recognition**

To explain Gemini's superior performance, it is necessary to examine how its multimodal architecture differs fundamentally from Ernie ViLG and GT. Gemini succeeded across Arabic, Japanese, and Chinese calligraphic images, despite the scripts being visually complex and linguistically diverse because it combines broad multilingual training with strong pattern-matching abilities and a large internal library of classical texts, idioms, and poetic structures. Gemini uses contextual reasoning to match partially legible strokes to known literary sources, allowing it to identify poems, proverbs, and set phrases even when the calligraphy is stylized or incomplete. Its training includes a wide range of global scripts, so it can recognize the visual logic of Arabic curves, Japanese kanji and kana combinations, and Chinese brush forms, giving it a cross-script advantage that Ernie and GT lack. Gemini also excels at reconstructing meaning from fragments. When it sees a few identifiable characters, it searches for the most plausible canonical text that fits the pattern, which often leads to accurate identifications in Arabic and Japanese, and reasonably strong performance in Chinese. Although this strategy sometimes produces high-prestige hallucinations, it generally allows Gemini to outperform the other models by integrating visual cues, cultural knowledge, and literary priors into a single interpretive process. In short, Gemini succeeds because it is not just reading the calligraphy—it is reasoning through it.

Gemini outperformed Ernie-ViLG on Japanese and Chinese calligraphy because Ernie-ViLG is built on modern printed Chinese, typed characters, web text, captioned images, and standard handwriting, but traditional calligraphy as cursive, semi-cursive and expressive brushwork belongs to a completely different visual system that requires years of human exposure to stroke habits, historical styles, brush pressure, and ink flow. Without extensive labelled calligraphy data, a model perceives these forms as abstract art rather than readable text, which is why Ernie, despite being “Chinese,” lacks true calligraphy literacy and falls back on generic philosophical or Taoist phrases when it cannot decode the strokes. Gemini, by contrast, appears to have broader training on classical literature, stronger pattern-matching for famous poems, and better semantic retrieval when the calligraphy resembles known works. When it encounters a poem like Du Fu's, it recognizes the structure, matches it to a canonical source, and retrieves the correct text. The divergence between the two models illustrates the core insight of this study: multimodal systems do not actually *read* calligraphy- they infer meaning based on their training priors.

GT failed to identify Arabic calligraphy and to translate Arabic, Japanese and Chinese calligraphic images because its image-translation system is built on OCR (optical character recognition) designed for *printed, typed, or very clean handwriting*, not for artistic scripts. Arabic calligraphy uses elongated strokes, decorative ligatures, variable letter shapes, and complex flourishes that break the assumptions of OCR, causing GT to mis-segment letters or fail to detect them entirely. Japanese and Chinese calligraphy pose the same challenge: brush strokes merge, radicals distort, proportions shift, and characters lose the rigid geometry that OCR depends on. Although GT offers an “image translation” option in its interface, this feature is optimized for standard text in photos, not for stylized or historical scripts. Without training on thousands of labeled calligraphy samples, GT cannot map the visual forms to actual characters, so it produces random fragments, unrelated words, or complete nonsense. In short, GT's failure is not a menu-design issue—it is a limitation of OCR-based systems that cannot interpret calligraphy as language, only as unreadable shapes.

### **4.4 Comparison with Findings of Prior Studies**

Current results both confirm and challenge patterns reported by prior studies in the literature. Studies on East Asian calligraphy recognition, such as Bao et al. (2025), Peng et al. (2022), Si (2024), and the Kuzushiji research by Ueki & Kojima (2021), generally report high accuracy when models are trained on large, labelled datasets of calligraphic characters, especially when the scripts follow semi-standardized forms. This is partially consistent with the current finding that Gemini performs well when the calligraphy resembles known poems or canonical structures, effectively mirroring and matching the pattern successfully. Similarly, research on style classification and writer identification by (Hu & Wu, 2024; Zhu & Zhu, 2024; Wang & Zong, 2023) echoes current findings in that AI models can excel in genre or style recognition even when they struggle with literal reading. Current findings showed that Ernie recognized cultural themes but failed at character-level accuracy. However, current results diverge sharply from studies that reported strong OCR-based performance on Chinese or Arabic calligraphy (e.g., Alsalamah, 2020; Hussein, 2021; Kaoudja et al., 2021), because these AI systems were trained specifically on calligraphic datasets, whereas Gemini, Ernie, and GT were not. Current

findings also contrast with deep learning work on Arabic scripts by (Zerdoumi et al., 2023; Zafar & Iqbal, 2022), which demonstrates that specialized AI models can decode even historical or stylized forms—a capability absent in the general-purpose models tested in the current study. Overall, the current study reinforces the literature showing that AI succeeds when trained on calligraphy, but it also exposes a gap in that general multimodal models, despite their linguistic breadth, lack true calligraphic literacy, especially in Arabic and highly cursive Japanese and Chinese scripts.

## 5. Recommendations

### 5.1 Implications for Education, Linguistic Landscapes, and Public Communication

Findings of this study highlight several important implications for calligraphy education, linguistic landscapes, and public communication. First, because general-purpose AI systems such as Ernie ViLG and Google Translate cannot reliably read stylized Arabic, Japanese, or Chinese calligraphy, art and calligraphy training programs should integrate AI literacy into their curricula. Building on earlier work in art education by (Al Jarf, 2009, Al Jarf, 2013, Al Jarf, 2021d), art students should be trained to critically evaluate AI responses and recognition of calligraphic texts, understand the limitations of current recognition technologies, and use AI tools as supportive - not authoritative - resources when analyzing calligraphic texts.

Second, current results have direct relevance for bilingual shop signs, the linguistic landscape and product labels. Prior research showed that stylized calligraphy, unconventional transliterations, and creative branding already challenge human readers. This study demonstrates that AI systems struggle even more with bilingual calligraphy in real-world signages. Misrecognition affects translation accuracy, digital mapping, accessibility, and automated linguistic landscape analysis. Future AI development should therefore incorporate calligraphy-specific datasets and training protocols that the artistic and cultural dimensions of Saudi public signage.

Third, current findings carry implications for tourism and public communication. Tourists increasingly rely on AI-based camera translation to navigate unfamiliar environments, landmarks, and museum collections, yet stylized bilingual signage remains largely unreadable to current AI systems. This can lead to mistranslations, misunderstandings, and reduced cultural accessibility. Improving AI recognition of calligraphic scripts - and designing public signage with digital readability in mind - would enhance navigation, cross-cultural communication, and visitor experience. As personal travel experiences show, AI has the potential to transform linguistic accessibility, but only if it can accurately interpret the calligraphic forms that shape real-world communication.

### 5.2 Recommendations for AI Development and Future Research

Based on findings of this study, several recommendations emerge for improving AI performance on Arabic, Japanese, and Chinese calligraphy, as well as for future research on text recognition by A. General multimodal AI models struggle because they are not trained on stylized scripts. Creating large, labelled datasets of Arabic Thuluth/Diwani, Japanese semi-cursive, and Chinese cursive would significantly improve recognition accuracy and reduce hallucinations. Gemini's success shows the value of combining global visual reasoning with linguistic priors. Future AI systems should integrate OCR for clean segments, multimodal reasoning for stylized strokes & cultural/genre priors for canonical texts.

AI models need modules that can reconstruct brush motion, stroke order, and pressure - especially for cursive Chinese and artistic Arabic scripts. This would help distinguish ornamentation from meaningful strokes. Future models should include uncertainty calibration, hallucination-detection layers and cross-checking mechanisms against visual evidence to prevent overconfident misreadings. Gemini's cross-script identification and recognition advantage suggests that training on multiple writing systems strengthens pattern recognition. Future models should incorporate more Arabic calligraphy styles, more historical Japanese scripts and more Chinese cursive exemplars to improve generalization.

AI should not replace expert readers but should support them. Practical applications include preliminary transcription for scholars, style classification for museums, and metadata generation for digital archives. Human oversight remains essential for ensuring accuracy in calligraphic script recognition. Finally, future research may explore the ability of Gemini, ERNIE-ViLG, and GT to recognize English and Arabic handwritten texts, as well as compare human (especially student) and AI performance in recognizing Arabic calligraphic verses—an area that remains open for further investigation.

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

**ORCID ID:** <https://orcid.org/0000000262551305>

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

## References

- [1] Al-Jarf, R. (2025a). AI translation of full-text Arabic research articles: The case of educational polysemes. *Journal of Computer Science and Technology Studies*, 7(1), 311-325. [Google Scholar](#)
- [2] Al-Jarf, R. (2025b). AI translation of the Gaza-Israel war terminology. *International Journal of Linguistics, Literature and Translation*, 8(2), 139-152. [Google Scholar](#)
- [3] Al-Jarf, R. (2025c). Arabic transliteration of borrowed English nouns with /g/ by Artificial Intelligence (AI). *Journal of Computer Science and Technology Studies*, 7(9), 245-252. [Google Scholar](#)
- [4] Al-Jarf, R. (2025e). Can AI decode and interpret encrypted Arabic on Facebook and YouTube to evade algorithmic moderation. *Journal of Computer Science and Technology Studies*, 7(12), 307-321. <https://doi.org/10.32996/jcsts.2025.7.12.40>. [Google Scholar](#)
- [5] Al-Jarf, R. (2025f). Can Artificial Intelligence (AI) translate Arabic abu-brand names with different prompts. *Journal of Computer Science and Technology Studies*, 7(9), 768-779. [Google Scholar](#)
- [6] Al-Jarf, R. (2025h). *Can students learning Arabic as a foreign language use Arabic YouTube videos narrated by Artificial Intelligence (AI) for listening practice*. 2nd International Forum on Teaching Arabic in the Modern World: Traditions and Innovations. Sheikha Fatima bint Mubarak Center for Education. Primakov International School Moscow, Russia. November 15–16, 2025. <https://www.researchgate.net/publication/398106697>. [Google Scholar](#)
- [7] Al-Jarf, R. (2025i). Copilot's English translation of contrastive emphatic negation in Arabic discourse: An analytical study. *International Journal of Linguistics, Literature and Translation*, 8(12), 214-230. DOI: 10.32996/ijllt.2025.8.12.24. [Google Scholar](#)
- [8] Al-Jarf, R. (2025j). Copilot vs DeepSeek's translation of denotative and metonymic abu- and umm- animal and plant folk names in Arabic. *Journal of Computer Science and Technology Studies*, 7(10), 367-385. [Google Scholar](#)
- [9] Al-Jarf, R. (2025k). DeepSeek, Google translate and Copilot's translation of Arabic grammatical terms used metaphorically. *Journal of Computer Science and Technology Studies*, 7(3), 46-57. [Google Scholar](#)
- [10] Al-Jarf, R. (2025l). Google Translate then and now: Translations from five languages into English and Arabic (2012–2025). *Journal of Computer Science and Technology Studies*, 7(12), 413-427. DOI: 10.32996/jcsts.2025.7.12.50. [Google Scholar](#)
- [11] Al-Jarf, R. (2025m). Human vs AI translation of common names of chemical compounds: A comparative study. *Frontiers in Computer Science and Artificial Intelligence*, 4(4), 11-24. <https://doi.org/10.32996/fcsai.2025.4.4.2>. [Google Scholar](#)
- [12] Al-Jarf, R. (2025n). Pan Arab linguistic and translation errors and strategies in bilingual linguistic landscapes. *International Journal of Translation and Interpretation Studies*, 5(3), 17-38. <https://doi.org/10.32996/ijtis.2025.5.3.4> [Google Scholar](#)
- [13] Al-Jarf, R. (2025o). Pronunciation errors in Arabic YouTube videos narrated by AI. *Frontiers in Computer Science and Artificial Intelligence*, 4(2), 01-12. <https://doi.org/10.32996/fcsai.2025.2.2.1>. [Google Scholar](#)
- [14] Al-Jarf, R. (2025p). *Pronunciation errors in AI-narrated Arabic YouTube videos*. LICCS Online Conference on Teaching and Research in Language and Culture: Past, Present and AI. Babeş-Bolyai University, Cluj-Napoca, Romania. September 11-12, 2025. [Google Scholar](#)
- [15] Al-Jarf, R. (2025q). Specific linguistic questions that Artificial Intelligence (AI) cannot answer accurately: Implications for Digital Didactics. *Frontiers in Computer Science and Artificial Intelligence*, 4(4), 43-61. <https://doi.org/10.32996/fcsai.2025.4.4.4>. [Google Scholar](#)
- [16] Al-Jarf, R. (2025r). *To publish or not to publish AI-generated research articles in scholarly journals: A perspective from editors and publishers*. I2COMSAPP International Conference on Artificial Intelligence and its Practical Applications in the Age of Digital Transformation. 2nd Edition. Faculty of Sciences and Techniques. Nouakchott University, Nouakchott, Mauritania. October 22-24, 2025. [Google Scholar](#)
- [17] Al-Jarf, R. (2025s). Translation of Arabic expressions of impossibility by AI and student-translators: A comparative study. *Journal of Computer Science and Technology Studies*, 7(8), 288-299. [Google Scholar](#)
- [18] Al-Jarf, R. (2025t). Translation of Arabic folk medical terms with om and abu by AI: A comparison of Microsoft Copilot and DeepSeek. *Journal of Medical and Health Studies*, 6(4), 45-58. [Google Scholar](#)
- [19] Al-Jarf, R. (2025u). Translation of English and Arabic "sleep" terms and formulaic expressions by Artificial Intelligence: A comparison of Copilot and DeepSeek. *International Journal of Linguistics, Literature and Translation*, 8(11), 95-108. [Google Scholar](#)
- [20] Al-Jarf, R. (2025v). Translation of zero-expressions by Microsoft Copilot and Google Translate. *Journal of Computer Science and Technology Studies*, 7(2), 203-216. [Google Scholar](#)
- [21] Al-Jarf, R. (2024a). Definite or Indefinite? The case of Arabic product names as judged by student translators. *International Journal of Linguistics, Literature and Translation*, 7(3), 83-92. DOI: 10.32996/ijllt.2024.7.3.10. [Google Scholar](#)
- [22] Al-Jarf, R. (2024b). English language representation in Korean linguistic landscapes. *International Journal of Asian and African Studies*, 3(2), 01-10. DOI: <https://doi.org/10.32996/ijaas.2024.3.2.1>. [Google Scholar](#)
- [23] Al-Jarf, R. (2024c). Students' assignments and research papers generated by AI: Arab instructors' views. *Journal of Computer Science and Technology Studies*, 6(2), 92-98. [Google Scholar](#)

- [24] Al-Jarf, R. (2024d). To translate or not to translate: The case of Arabic and foreign shop names in Saudi Arabia. *International Journal of Translation and Interpretation Studies*, 4(1), 33-40. DOI: 10.32996/ijtis.2024.4.1.5. [Google Scholar](#)
- [25] Al-Jarf, R. (2024e). *Translation of medical terms by AI: A comparative linguistic study of Microsoft Copilot and Google Translate*. I2COMSAPP'2024 Conference, Nouakchott, Mauritania. [Google Scholar](#)
- [26] Al-Jarf, R. (2024f). *Translation of medical terms by AI: A comparative linguistic study of Microsoft Copilot and Google Translate*. In Y. M. Elhadj et al. (Eds.), I2COMSAPP 2024, LNNS 862, pp. 1–16. [https://doi.org/10.1007/978-3-031-71429-0\\_17](https://doi.org/10.1007/978-3-031-71429-0_17). Springer Nature Switzerland AG 2024. [Google Scholar](#)
- [27] Al-Jarf, R. (2023). Semantic and syntactic anomalies of Arabic-transliterated compound shop names in Saudi Arabia. *International Journal of Arts and Humanities Studies (IJAHs)*, 3(1), 1-8. DOI: 10.32996/ijahs.2023.3.1.1. [Google Scholar](#)
- [28] Al-Jarf, R. (2022a). Deviant Arabic transliterations of foreign shop names in Saudi Arabia and decoding problems among shoppers. *International Journal of Asian and African Studies (IJAAAS)*, 1(1), 17-30. DOI: 10.32996/ijaas.2022.1.1.3. [Google Scholar](#)
- [29] Al-Jarf, R. (2022b). Dominance of foreign shop names over Arabic names in Saudi Arabia: Promotional, sociocultural and globalization issues. *International Journal of Middle Eastern Research (IJMER)*, 1(1), 23-33. DOI: 10.32996/ijmer.2022.1.1.5. [Google Scholar](#)
- [30] Al-Jarf, R. (2021a). An investigation of Google's English-Arabic translation of technical terms. *Eurasian Arabic Studies*, 14, 16-37. [Google Scholar](#)
- [31] Al-Jarf, R. (2021b). Linguistic-cultural characteristics of hotel names in Saudi Arabia: The case of Makkah, Madinah and Riyadh Hotels. *International Journal of Linguistics, Literature and Translation (IJLLT)*, 4(8), 160-170. DOI: 10.32996/ijllt.2021.4.8.23. [Google Scholar](#)
- [32] Al-Jarf, R. (2021c). Teaching English with linguistic landscapes to Saudi students studying abroad. *Asian Journal of Language, literature and Culture Studies (AJL2CS)*, 4(3), 1-12. ERIC ED619894. [Google Translate](#)
- [33] Al-Jarf, R. (2021d). Testing reading for specific purposes in an art education Course for graduate students in Saudi Arabia. *International Journal of Advance and Innovative Research*, 8 (1), 32-42. ERIC ED617119. [Google Scholar](#)
- [34] Al-Jarf, R. (2016). Issues in translating English technical terms to Arabic by Google Translate. *TICET 2016 Conference*, Khartoum, Sudan. [Google Scholar](#)
- [35] Al-Jarf, R. (2013). Teaching and assessing graduate students' research skills in english for art education Purposes. 1st International Conference on Teaching English for Specific Purposes: "Connect and Share". University of Niš, Faculty of Electronic Engineering, Serbia. pp. 771-780. ERIC ED610674 [Google Scholar](#)
- [36] Al-Jarf, R. (2012). *Electronic translation between Arabic and European languages: Current status and future Perspectives*. 6th Annual Conference of Ibn Sina Institute for Human Sciences titled: The Future of Arabic Language Teaching in Europe. LILLE, France. June 22-24. [Google Scholar](#)
- [37] Al-Jarf, R. (2009). Using online instruction in English for art education. *Asian EFL Journal Teaching Articles*, 34, February, 50-60. ERIC ED634168. [Google Scholar](#)
- [38] Akbar, M. & Hates, K. (2025). AI analysis of Arabic calligraphy characteristics in digital advertisements. *IEEE 4th International Conference on Computing and Machine Intelligence (ICMI)* (pp. 1-5). IEEE.
- [39] Al Hamad, M. (2025). *The rise and development of the Arabic script: From silent stones to the Qur'an's voice*. In National Museum of World Writing Systems (Ed.), *The spread of phonetic scripts: Beyond the Tower of Babel* (pp. 275–314). National Museum of World Writing Systems.
- [40] Al-Hmouz, R. (2020). Deep learning autoencoder approach: Automatic recognition of artistic Arabic calligraphy types. *Kuwait Journal of Science*, 47(3).
- [41] Alsalamah, S. (2020). *Combining image and text processing for the computational reading of Arabic Calligraphy*. The University of Manchester. United Kingdom.
- [42] Bao, X., Wang, Z., Gu, J. & Huang, C. (2025). Calligraphic OCR for Chinese calligraphy recognition. *2025 Conference on Empirical Methods in Natural Language Processing* (pp. 4865-4877).
- [43] Batjargal, B. (2022). Recognizing Kuzushiji in Japanese historical documents. *International ARC Seminar Review. Art Research*, 22, 2.
- [44] El-Desouky, A., et al. (1991). A handwritten Arabic character recognition technique for machine reader. *3rd International Conference on Software Engineering for Real Time Systems*. pp. 212-216. IET.
- [45] González, Á. & Cabarrocas, S. (2025). Artificial intelligence for calligraphic writer identification: The case of Lope de Vega's Autographs. *Hipogrifo: Revista de Literatura y Cultura del Siglo de Oro*, 13(1), 517-532.
- [46] Gürer, D. & Gökbay, İ. (2023). Arabic calligraphy image analysis with using traditional machine learning algorithms and transfer learning. *2023 Innovations in Intelligent Systems and Applications Conference (ASYU)* (pp. 1-7). IEEE.
- [47] Hu, X., & Wu, F. (2024). Applications of machine learning in recognizing Chinese calligrapher's handwriting styles. *5th International Conference on Computer Information and Big Data Applications* (pp. 548-552).
- [48] Hussein, A. (2021). Fast learning neural network based on texture for Arabic calligraphy identification. *Indonesian Journal Of Electrical Engineering And Computer Science*, 21(3), 1794-1799.

- [49] Jiang, Z., et al. (2023). When artificial intelligence comes to the Chinese calligraphic landscape: The coming transformation. *Geography Compass*, 17(1), e12670.
- [50] Kaoudja, Z., Kherfi, M. L., & Khaldi, B. (2021). A new computational method for Arabic calligraphy style representation and classification. *Applied Sciences*, 11(11), 4852.
- [51] Lin, T. et al. (2025). Future ink: The collision of AI and Chinese calligraphy. *ACM Journal on Computing and Cultural Heritage*, 18(1), 1-17.
- [52] Liu, M., & Zhang, Y. (2024). Intelligent recognition algorithm for calligraphy fonts based on texture mapping. *Computer-Aided Design & Applications*, 21(S3), 2024, 197-210.
- [53] Peng, X. et al. (2022). Calligraphy character detection based on deep convolutional neural network. *Applied Sciences*, 12(19), 9488.
- [54] Salamah, S. & King, R. (2018). Towards the machine reading of Arabic calligraphy: A letters dataset and corresponding corpus of text. In *2018 IEEE 2nd international workshop on Arabic and derived script analysis and recognition (ASAR)* (pp. 19-23). IEEE.
- [55] Si, H. (2024). Analysis of calligraphy Chinese character recognition technology based on deep learning and computer-aided technology. *Soft Computing*, 28(1), 721-736.
- [56] Sumayli, A., & Alkaoud, M. (2025). Handwritten Arabic calligraphy generation: A systematic literature review. *International Journal of Advanced Computer Science and Applications*
- [57] Ueki, K., & Kojima, T. (2021). Survey on deep learning-based Kuzushiji recognition. In *International Conference on Pattern Recognition* (pp. 97-111). Cham: Springer International Publishing.
- [58] Wang, Y., & Zong, Y. (2023, December). Calligraphy font recognition algorithm based on improved DenseNet network. *2023 Global Conference on Information Technologies and Communications (GCITC)* (pp. 1-5). IEEE.
- [59] Xing, B., et al. (2025). AI-driven exploration system of emotional and stylistic expressions in calligraphy. *Journal of Engineering Design*, 1-31.
- [60] Xu, C., Liang, J., & Ling, X. (2025). A machine learning-based calligraphy font recognition system using hog features and support Vector machine. *11th International Conference on Computing and Artificial Intelligence (ICCAI)* (pp. 175-180). IEEE.
- [61] Yang, C., Osman, M. & Sarvghadi, M. (2025). A web-based platform for Korean Calligraphy learners using deep learning. *IEEE*.
- [62] Zafar, A., & Iqbal, A. (2022). Application of soft computing techniques in machine reading of Quranic Kufic manuscripts. *Journal of King Saud University-Computer and Information Sciences*, 34(6), 3062-3069.
- [63] Zerdoumi, S. et al. (2023). A deep learning based approach for extracting Arabic handwriting: applied calligraphy and old cursive. *PeerJ Computer Science*, 9, e1465.
- [64] Zhu, Y., & Zhu, Y. (2024). PagodaNet: Light weight AI Model for Chinese calligraphy styles recognition. *4th International Conference on Software Engineering and Artificial Intelligence (SEAI)* (pp. 22-27). IEEE.