information

MDPI

Article

Fine-Tuning QurSim on Monolingual and Multilingual Models for Semantic Search

Tania Afzal [1,2,†], Sadaf Abdul Rauf [1,2,†], Muhammad Ghulam Abbas Malik [3,†] and Muhammad Imran [4,5,*,†]

[1] Speech and Language Processing Group, Fatima Jinnah Women University, Rawalpindi 46000, Pakistan; 20-20931-013@cs.fjwu.edu.pk (T.A.); sadaf.arauf@fjwu.edu.pk (S.A.R.)
[2] Department of Computer Science, Fatima Jinnah Women University, Rawalpindi 46000, Pakistan
[3] Interdisciplinary Sustainable Systems (IS2) Group, College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia; amaalik@psu.edu.sa
[4] Education Research Lab, Prince Sultan University, Riyadh 11586, Saudi Arabia
[5] Department of English Language and Literature, Khazar University, Baku AZ1096, Azerbaijan
* Correspondence: mimran@psu.edu.sa
† These authors contributed equally to this work.

Abstract: Transformers have made a significant breakthrough in natural language processing. These models are trained on large datasets and can handle multiple tasks. We compare monolingual and multilingual transformer models for semantic relatedness and verse retrieval. We leveraged data from the original QurSim dataset (Arabic) and used authentic multi-author translations in 22 languages to create a multilingual QurSim dataset, which we released for the research community. We evaluated the performance of monolingual and multilingual LLMs for Arabic and our results show that monolingual LLMs give better results for verse classification and matching verse retrieval. We incrementally built monolingual models with Arabic, English, and Urdu and multilingual models with all 22 languages supported by the multilingual paraphrase-MiniLM-L12-v2 model. Our results show improvement in classification accuracy with the incorporation of multilingual QurSim.

Keywords: semantic similarity; Quranic verse classification; verse retrieval; semantic search

check for updates

Academic Editor: Paulo Quaresma

Received: 28 October 2024
Revised: 15 January 2025
Accepted: 18 January 2025
Published: 23 January 2025

Citation: Afzal, T.; Abdul Rauf, S.; Malik, M.G.A.; Imran, M. Fine-Tuning QurSim on Monolingual and Multilingual Models for Semantic Search. *Information* **2025**, *16*, 84. https://doi.org/10.3390/info16020084

1. Introduction

Research in natural language processing (NLP) has seen a spotlight on large language models (LLMs) such as BERT [1], GPT [2], and XLM [3], which have enabled machines to capture context and relationships between text like never before. BERT emerged as a breakthrough that significantly advanced English-based tasks. The initial LLMs were specific to the English language; however, their multilingual counterparts were later introduced, e.g., mBERT [1] was expanded to more than 100+ languages.

Modern Arabic natural processing, like most fields, has succumbed to deep learning to address the complexities of the Arabic language [4]. Existing work has focused mainly on monolingual models such as the BERT base and AraBERT [5] to handle rich morphological and complex syntactic structures [6–8]. Hybrid approaches such as Roberta [9] with LSTM and CNN improve classification by improving semantic understanding and addressing class imbalances [10].

Monolingual models generally perform better, but multilingual models benefit from the shared representation from multiple languages [11,12]. However, as the size of multilingual models increases, their explanation becomes less reliable compared to that of monolingual models [13]. We explore these effects by evaluating the performance of monolingual and multilingual models across various languages. An interesting feature of our work is our multilingual corpus, which is based on human translations of the Holy Quran in multiple languages.

The Holy Quran is the holy book of Muslims revealed over fourteen centuries ago in Arabic. It has 114 Surahs (Chapters) and 6348 verses (sentences) with 78,000 words [14]. Arabic in the Holy Quran is unique and distinct, known for its eloquence, precision, and clarity. The Arabic language has a complex and rich grammar system, with intricate rules for pronunciation, morphology (word structure), and syntax (sentence structure). According to [15]'s Holy Quran corpus, God's words differ from other corpora such as newspapers, speeches, etc. Although the Holy Quran is a sacred scripture in Islam, it has been analyzed and studied from time to time from the perspective of NLP to gain insight into its linguistic and textual features.

QurSim [16] is an Arabic corpus that measures the relatedness of sentences for short texts (verses) based on the Holy Quran. The QurSim dataset is intended to incorporate pairs of Quranic verses along with their corresponding semantic similarity scores. These verse pairs were chosen based on their connection and contextual relevance, as obtained from the commentary on the Holy Quran by the famous Islamic scholar Ibn-e-Kathir. The commentary of [17] provides distinctive insights into the interpretation and context of Quranic verses, making the QurSim dataset a useful and credible resource for analyzing semantic alignment in the Holy Quran.

We extend the Qursim corpus to include all language translations available on tanzil.net. We conducted experiments using various monolingual and multilingual models, including AraBERTv0.2 and CAMelBERT-CA, which were trained on many Arabic datasets along with the multilingual paraphrase-MiniLM-L12-v2 model. Our testing involved three distinct languages for monolingual evaluation: Arabic, Urdu, and English. For multilingual evaluation, we combined Arabic, Urdu, and English to test how the model handles these languages and to assess the semantic connections among Quranic verses.

In this study, we addressed two main tasks: classification and verse retrieval by measuring semantic similarity between multiple languages. Our main contributions include the following:

1. A detailed analysis using the original and multilingual version of QurSim aimed at verse classification and retrieval using monolingual and multilingual models.
2. The release of a multilingual version of QurSim for the research community: https://github.com/sabdul111/QurSimMultilingual (accessed on 27 October 2024).

We start with an overview of the corpus and our approach in Section 2. Section 3 details the LLM and methodology. Section 4 presents results, While Section 5 focuses on benchmarking and Section 6 briefly reviews past work in the light of the Arabic language. Finally, Section 7 concludes this paper.

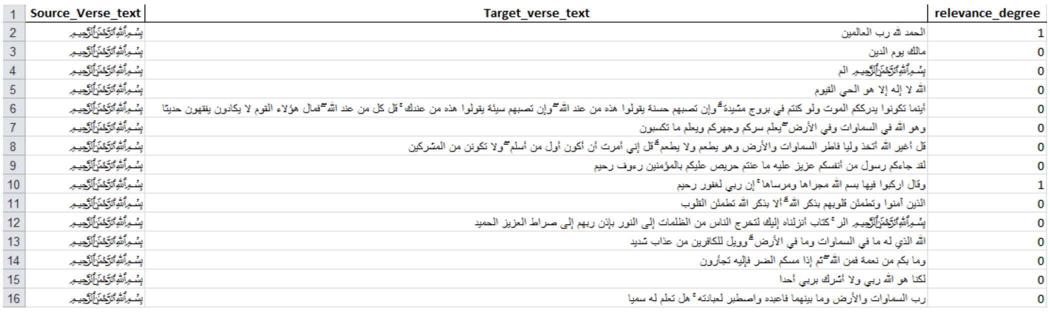
2. Corpus Collection and Preprocessing

The data in this study were taken from the QurSim [16] dataset based on Tafseer Ibn-e-Kathir [17]. This dataset includes verse pairs from all chapters of the Holy Quran categorized as similar and non-similar. Each pair includes a source verse and target verse, along with a relevance degree rating of (2, 1, 0), where a degree of 2 shows strong similarity, a degree of 1 indicates weak similarity, and a degree of 0 represents no similarity. The original dataset also includes chapter and verse numbers for both the source and target verses. We fetched the actual verse text and created the dataset with verses. The total original QurSim corpus has a collection of 7679 verse pairs.

To obtain similar and non-similar verses from the Holy Quran to align with the QurSim dataset. We extracted raw text from tanzil.net and applied the following preprocessing steps. Initially, we removed punctuation and stop words using Python 3 regex libraries without removing verse numbers. Next, we extracted similar verses from the original Quranic dataset based on their verse numbers. After extracting similar verses, we removed the verse numbers to eliminate the impact of numbers in sentence embeddings. For this process, we extracted the source and the target verse separately.

The dataset consists of a CSV file containing three columns: source surah verse, target surah verse, and relevance degree. For the multilingual setting, translations were mapped in accordance with the verse numbers in the QurSim dataset. The languages included Arabic (ar), English (en), Urdu (ur), Bulgarian (bg), Czech (cs), German (de), Persian (fa), Hindi (hi), Indonesian (id), Italian (it), Japanese (ja), Korean (ko), Kurdish (ku), Malay (ms), Dutch (nl), Polish (pl), Romanian (ro), Russian (ru), Albanian (sq), Swedish (sv), Thai (th), and Turkish (tr).

In our experimental setting, we selected pairs with a strong degree of similarity (2) and pairs with no similarity (0), representing these pairs as 1 and 0, respectively, while excluding pairs with weak similarity (1), as shown in Figure 1. This resulted in a total of 3961 pairs for our training set. To balance our training set, we selected a set of non-similar verse pairs from the Holy Quran, comprising an additional 2197 pairs. The total number of verse pairs used in this experiment are summarized in Table 1. We expanded our dataset by including translations from 16 authors in English and 8 authors in Urdu, and for multilingual, we incorporated translations in 22 different languages, resulting in a total of 84 translations, as illustrated in Table 2.



1	Source_Verse_text	Target_verse_text	relevance_degree
2	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	الحمد لله رب العالمين	1
3	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	مالك يوم الدين	0
4	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ الم	0
5	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	الله لا إله إلا هو الحي القيوم	0
6	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	أينما تكونوا يدرككم الموت ولو كنتم في بروج مشيدة ۗ وإن تصبهم حسنة يقولوا هذه من عند الله ۖ وإن تصبهم سيئة يقولوا هذه من عندك ۚ قل كل من عند الله ۖ فمال هؤلاء القوم لا يكادون يفقهون حديثا	0
7	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	وهو الله في السماوات وفي الأرض ۖ يعلم سركم وجهركم ويعلم ما تكسبون	0
8	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	قل أغير الله أتخذ وليا فاطر السماوات والأرض وهو يطعم ولا يطعم ۗ قل إني أمرت أن أكون أول من أسلم ۖ ولا تكونن من المشركين	0
9	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	لقد جاءكم رسول من أنفسكم عزيز عليه ما عنتم حريص عليكم بالمؤمنين رءوف رحيم	0
10	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	وقال اركبوا فيها بسم الله مجراها ومرساها ۚ إن ربي لغفور رحيم	1
11	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	الذين آمنوا وتطمئن قلوبهم بذكر الله ۗ ألا بذكر الله تطمئن القلوب	0
12	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ الر ۚ كتاب أنزلناه إليك لتخرج الناس من الظلمات إلى النور بإذن ربهم إلى صراط العزيز الحميد	0
13	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	الله الذي له ما في السماوات وما في الأرض ۗ وويل للكافرين من عذاب شديد	0
14	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	وما بكم من نعمة فمن الله ۖ ثم إذا مسكم الضر فإليه تجأرون	0
15	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	لكنا هو الله ربي ولا أشرك بربي أحدا	0
16	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	رب السماوات والأرض وما بينهما فاعبده واصطبر لعبادته ۚ هل تعلم له سميا	0

Figure 1. QurSim dataset with labeled pairs of source and target verses with relevance degree (1 for relevant, t and 0 for non-relevant).

Table 1. Relevance degrees and values of QurSim.

Relevance Degree	Verse Pairs
2	3079
0	882
0 (ours)	2197
Relevance Total	6158

Table 2. Relevance degrees and values of ar, en, ur, bg, cs, de, fa, hi, id, it, ja, ko, ku, ms, nl, pl, ro, ru, sq, sv, th, and tr languages. Multiple author translations contribute to bigger sets for some languages.

Language	Verse Pairs
Arabic	6.1 K
English	98.5 K
Urdu	49 K
Ar + En + Ur	154 K
Multilingual	517.2 K

3. Methodology

In this section, we explore the pair similarity of the Holy Quran using the original and multilingual QurSim datasets. We analyzed how the monolingual and multilingual models work effectively in capturing the semantic similarity of verses. We undertook two tasks for finding verse similarity: Quranic verse classification and verse retrieval.

3.1. Quranic Verse Classification

Verse classification relates to computing verse relatedness using LLMs. To calculate how closely one verse is similar to another, we converted the dataset into an embedding vector space to find its relatedness.

3.2. Language Model Selection

We carefully selected a state-of-the-art language model, which is based on Arabic CA and MSA, including well-known monolingual models AraBertv0.2 base [18] with a size of 77 GB with 136 M parameters based on the BERT architecture, designed and pre-trained for processing Arabic text. Trained on a large Arabic corpus, including Arabic Wikipedia, news articles, and other publicly available data sources. It was trained for tasks such as named entity recognition, sentiment analysis, and question answering.

CAMelBERT-CA [19], which was trained in classical Arabic with 6 GB data, is also based on the BERT architecture, part of the CAMeL NLP toolkit, optimized for various Arabic dialects and Modern Standard Arabic (MSA). It includes multiple variants tailored to different Arabic datasets. For example, CAMelBERT-CA is trained on classical Arabic texts, while other variants may focus on different dialects. It was designed to perform tasks such as Named Entity Recognition (NER), sentiment analysis, and classification for Arabic dialects, MSA, and classical Arabic.

For the multilingual experiments, we selected the transformer-based paraphrase-multilingual-MiniLM-L12-v2 model [20] based on the transformer architecture, but a distilled version of it. It is a smaller, lighter-weight model compared to BERT and similar models, providing efficient performance for large-scale applications, and is optimized for high-quality sentence embeddings and used for multilingual tasks such as semantic similarity, clustering, and paraphrase detection. Pre-trained on a multilingual corpus covering many languages (not limited to Arabic) and fine-tuned using the Sentence-BERT (SBERT) framework, its primary strength lies in efficient sentence similarity tasks, making it highly effective for multilingual semantic search and sentence-level embeddings across various languages. These models were chosen based on their ability to perform NLP tasks and their adaptability to multilingual settings. We fine-tuned our models on the QurSim dataset, which includes translations in multiple languages, using Google Colab. The dataset was split into training, evaluation, and test sets using a 90:5:5 ratio.

Training Dataset (90%): Used for model training and parameter optimization.

Validation Dataset (5%): Used for hyperparameter tuning and monitoring the model during train to prevent overfitting.

Test Dataset (5%): Reserved for the final evaluation to check the model's ability to generalize to unseen data. To avoid redundancy, division was performed at the verse-pair level to ensure that no verse appeared in more than one subset. Translations of the same verse across multiple languages were kept within the same subset, preventing overlap between training, evaluation, and test sets.

AraBERTv0.2 and CAMelBERT-CA were used for monolingual fine-tuning and, sentence transformers/paraphrase-MiniLM-L12-v2 model was used for multilingual fine-tuning. For each of the three models, we utilize the AutoTokenizer for tokenization and pass the source verse text and target verse text as input columns. We employ the Auto-

ModelForSequenceClassification model for all three, which is part of the Hugging Face Transformers library and provides an easy way to load any pre-trained transformer model. During the fine-tuning process, the model was trained to 50 epochs, with a batch size of 16 and a learning rate of 2 $\times 10^{-5}$. The Adam optimizer was employed to update model parameters during training.

3.3. Verse Retrieval

Verse retrieval in this process is based on cosine similarity, a method widely used to measure the similarity between two vectors by calculating the cosine of the angle between them. It is often used to measure the semantic similarity between two documents or vectors [21]. The formula for cosine similarity is as follows:

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2 \cdot \sum_{i=1}^{n} B_i^2}} \tag{1}$$

where A and B are two vectors representing the source and target verses. It calculates the cosine angle between two vectors. When the angle is small, the cosine similarity approaches 1, indicating the highest similarity. When an angle is large, the cosine similarity approaches 0, indicating the lowest similarity. In verse retrieval, this process involves converting both the source and target verses into a vector representation. The verses are tokenized using the AutoTokenizer and converted into embeddings using BERT-based models, including https://huggingface.co/aubmindlab/bert-base-arabertv02 (accessed on 27 October 2024), https://huggingface.co/CAMeL-Lab/bert-base-arabic-CAMelBERT-ca (accessed on 27 October 2024), and https://huggingface.co/sentence-transformers/paraphrase-MiniLM-L12-v2 (accessed on 27 October 2024). Cosine similarity is calculated for each input source verse against all target verses. The target verse with the highest cosine similarity score is considered the most relevant to the source verse.

3.4. Model Training

The dataset comprised verse pairs with binary labels that indicate relevance (1 for relevant, 0 for not relevant). QurSim dataset was used with 6158 verse pairs, split into 80% training and 20% validation subsets. The model was fine-tuned using the AdamW optimizer with a learning rate of 2 $\times 10^{-5}$. The training was conducted in batches of 16, over 5 epochs. Validation metrics such as precision, recall, F1-score, and accuracy were evaluated to monitor the models' performance.

4. Similarity Classification Results

Table 3 shows the results for the monolingual and multilingual models. Monolingual LLMs achieved the best results for Arabic, with both CAMelBERT-CA and AraBERTv0.2 having F1-scores of 0.90, while multilingual models had F1-scores of 0.78.

However, English and Urdu for multilingual models exhibited good scores of 0.989 and 0.980 due to the model being fine-tuned on large language translation datasets. The multilingual model with 22 languages also achieved a good score.

The confusion matrix provides a detailed evaluation of the model's prediction on the test dataset, highlighting the classification performance.

True Positives (TP) represent the value of 141 in the matrix, which shows the cases where the model correctly predicted similar verses.

True Negative (TN) represents the value of 138, which is predicted as a negative class where the verses are not similar.

False Positive (FP) represents the value of 9, where the model predicted the verses as similar, even though they are not similar.

False Negative (FN) represents the value of 20; the model predicted that the verses are not similar, but they are similar in reality.

High TP (141) and TN (138) values show that the model accurately predicts the majority of verses. The low values of FP (9) and FN (20) show a small number of values that show non-similar verses as similar and that the model makes few errors in misclassifying verses. Monolingual models show high performance in predicting the verse semantic similarity for the Arabic language as they were specifically trained on the Arabic dataset, allowing them to capture Arabic syntax and morphology. In contrast, multilingual models show limitations in this context due to training on datasets spanning multiple languages, resulting in less focus on Arabic data. As a result, they can struggle to capture complex syntactic features, which can cause low performance. However, multilingual models show the highest scores for English and Urdu due to the availability of extensive training data for these languages. The large fine-tuning dataset increases performance. The dataset used for fine-tuning the model in English and Urdu was significantly larger compared to that for Arabic, which limits the depth of learning for the latter. Detailed results are provided in the Appendix A (Tables A1 and A2).

Table 3. Test scores for monolingual and multilingual LLMs.

Model	Language	Test Loss	Test Accuracy	Precision	Recall	F1-Score	Confusion Matrix
Monolingual LLMs							
AraBertv0.2	Ar	0.0974	0.905	0.908	0.905	0.905	$\begin{vmatrix} 141 & 9 \\ 20 & 138 \end{vmatrix}$
CAMelBERT-CA	Ar	0.092	0.902	0.902	0.902	0.902	$\begin{vmatrix} 137 & 13 \\ 17 & 141 \end{vmatrix}$
Multilingual LLMs							
paraphrase-MiniLM-L12-v2	Ar	0.183	0.782	0.784	0.782	0.781	$\begin{vmatrix} 109 & 41 \\ 26 & 32 \end{vmatrix}$
paraphrase-MiniLM-L12-v2	En	0.008	0.989	0.989	0.989	0.989	$\begin{vmatrix} 2500 & 12 \\ 42 & 2373 \end{vmatrix}$
paraphrase-MiniLM-L12-v2	Ur	0.0164	0.980	0.981	0.980	0.980	$\begin{vmatrix} 1233 & 33 \\ 14 & 1184 \end{vmatrix}$
paraphrase-MiniLM-L12-v2	Ar + En + Ur	0.0151	0.981	0.982	0.981	0.981	$\begin{vmatrix} 3788 & 34 \\ 105 & 3771 \end{vmatrix}$
paraphrase-MiniLM-L12-v2	Ar + MultiLang	0.0145	0.982	0.982	0.982	0.982	$\begin{vmatrix} 10{,}581 & 120 \\ 261 & 10{,}591 \end{vmatrix}$

Verse Retrieval Results

Table 4 presents a detailed analysis of the performance metrics for the models, including the precision, recall, F1-score, accuracy, and confusion matrix. CAMeLBERT demonstrated the highest performance across all metrics, achieving the highest precision, recall, and F1-score. The CAMeLBERT confusion matrix has the lowest false positive and false negative rates, contributing to its high accuracy. The results of verse retrieval using an Arabic-based monolingual model are shown in Table 5, and the results of the multilingual model are shown in Table 6.

Table 4. Semantic search metrics result.

Model	Precision	Recall	F1-Score	Accuracy	Confusion Matrix
AraBERT	0.88	0.87	0.87	0.88	$\begin{vmatrix} 2890 & 345 \\ 384 & 2541 \end{vmatrix}$
CAMelBERT	0.91	0.91	0.91	0.91	$\begin{vmatrix} 2964 & 271 \\ 272 & 2653 \end{vmatrix}$
MiniLM	0.78	0.83	0.80	0.81	$\begin{vmatrix} 2543 & 692 \\ 499 & 2426 \end{vmatrix}$

The results in Table 6 indicate that [6:100] (Chapter 6, Verse 100) and [17:67] exhibit the highest cosine similarity of 0.996. But verse [6:100] describes the false attribution of the partner and offspring to God. On the other hand, verse [17:67] focuses on the tendency of humans to turn away from God when they are protected from harm, emphasising ingratitude. This shows that while the cosine similarity score shows a high degree of similarity, the verses do not convey the same meaning.

Verses [6:100, 6:145, 25:2], presented in Table 5, provide meaningful results, effectively highlighting the theme of monotheism and God's dominance over heaven and the earth. These verses declare that God has no son or partner in dominion and that He is the solitary Creator and Determiner of all things. This emphasizes the fundamental monotheistic idea that God is distinct and unsurpassed in His characteristics and authority. In this context, it is notable that the monolingual models AraBERTv0.2 and CAMelBERT-CA perform well in identifying the similarity between these verses. Although the multilingual model shows the highest similarity, the verses are not identical. This distinction emphasizes the need to consider semantic content and cosine similarity scores for more accurate interpretation.

Table 5. Verse retrieval using AraBERTv0.2 and CAMelBERT-CA.

	AraBERTv0.2			CAMelBERT-CA	
Chapter:Verse	**Input Source Verse**	**Relevance Target Verse**	**Similarity Score**	**Relevance Target Verse**	**Similarity Score**
(7:100, 6:6, 30:9)	أولم يهد للذين يرثون الأرض من بعد أهلها أن لو نشاء أصبناهم بذنوبهم ونطبع على قلوبهم فهم لا يسمعون Has it not become clear to those who inherited the earth after its [previous] people that if We willed, We could afflict them for their sins? But We seal over their hearts so they do not hear.	ألم يروا كم أهلكنا من قبلهم من قرن مكناهم في الأرض ما لم نمكن لكم وأرسلنا السماء عليهم مدرارا وجعلنا الأنهار تجري من تحتهم فأهلكناهم بذنوبهم وأنشأنا من بعدهم قرنا آخرين Have they not seen how many a generation before them We have destroyed whom We had established on the earth such as We have not established you? And We poured out on them rain from the sky in abundance and made the rivers flow under them. Yet We destroyed them for their sins, and created after them other generations.	0.902	ولم يسيروا في الأرض فينظروا كيف كان عاقبة الذين من قبلهم كانوا أشد منهم قوة وأثاروا الأرض وعمروها أكثر مما عمروها وجاءتهم رسلهم بالبينات فما كان الله ليظلمهم ولكن كانوا أنفسهم يظلمون Do they not travel in the land, and see what was the end of those before them? They were superior to them in strength, and they tilled the earth and populated it in greater numbers than these (pagans) have done, and there came to them their Messengers with clear proofs. Surely, Allah wronged them not, but they used to wrong themselves.	0.871

Table 5. *Cont.*

	AraBERTv0.2		CAMelBERT-CA		
Chapter:Verse	**Input Source Verse**	**Relevance Target Verse**	**Similarity Score**	**Relevance Target Verse**	**Similarity Score**
(7:159, 16:36, 43:60)	ومن قوم موسى أمة يهدون بالحق وبه يعدلون And of the people of Musa (Moses) there is a community who lead (the men) with truth and establish justice therewith	ولقد بعثنا في كل أمة رسولا أن اعبدوا الله واجتنبوا الطاغوت فمنهم من هدى الله ومنهم من حقت عليه الضلالة فسيروا في الأرض فانظروا كيف كان عاقبة المكذبين And We certainly sent into every nation a messenger, [saying], "Worship Allah and avoid Taghut". And among them were those whom Allah guided, and among them were those upon whom error was [deservedly] decreed. So proceed through the earth and observe how was the end of the deniers.	0.821	ولو نشاء لجعلنا منكم ملائكة في الأرض يخلفون And if We willed, We could have made [instead] of you angels succeeding [one another] on the earth.	0.847
(6:100, 6:145, 25:2)	وجعلوا لله شركاء الجن وخلقهم وخرقوا له بنين وبنات بغير علم سبحانه وتعالى عما يصفون But they have attributed to Allah partners—the jinn, while He has created them—and have fabricated for Him sons and daughters. Exalted is He and high above what they describe.	وجعلوا لله مما ذرأ من الحرث والأنعام نصيبا فقالوا هذا لله بزعمهم وهذا لشركائنا فما كان لشركائهم فلا يصل إلى الله وما كان لله فهو يصل إلى شركائهم ساء ما يحكمون And the polytheists assign to Allah from that which He created of crops and livestock a share and say, "This is for Allah", by their claim, "and this is for our partners [associated with Him]". But what is for their "partners" does not reach Allah, while what is for Allah—this reaches their "partners". Evil is that which they rule.	0.848	الذي له ملك السماوات والأرض ولم يتخذ ولدا ولم يكن له شريك في الملك وخلق كل شيء فقدره تقديرا He to whom belongs the dominion of the heavens and the earth and who has not taken a son and has not had a partner in dominion and has created each thing and determined it with [precise] determination.	0.788

Table 6. Verse retrieval using paraphrase-MiniLM-L12-v2.

Chapter:Verse	**Input Source Verse**	**Relevance Target Verse**	**Similarity Score**
(7:100, 6:59)	أولم يهد للذين يرثون الأرض من بعد أهلها أن لو نشاء أصبناهم بذنوبهم ونطبع على قلوبهم فهم لا يسمعون Has it not become clear to those who inherited the earth after its [previous] people that if We willed, We could afflict them for their sins? But We seal over their hearts so they do not hear.	وعنده مفاتح الغيب لا يعلمها إلا هو ويعلم ما في البر والبحر وما تسقط من ورقة إلا يعلمها ولا حبة في ظلمات الأرض ولا رطب ولا يابس إلا في كتاب مبين And with Him are the keys of the Ghaib (all that is hidden), none knows them but He. And He knows whatever there is in (or on) the earth and in the sea; not a leaf falls, but he knows it. There is not a grain in the darkness of the earth nor anything fresh or dry, but is written in a Clear Record.	0.997

Table 6. *Cont.*

Chapter:Verse	Input Source Verse	Relevance Target Verse	Similarity Score
(7:159, 5:20)	ومن قوم موسى أمة يهدون بالحق وبه يعدلون And of the people of Musa (Moses) there is a community who lead (the men) with truth and establish justice therewith	وإذ قال موسى لقومه يا قوم اذكروا نعمة الله عليكم إذ جعل فيكم أنبياء وجعلكم ملوكا وآتاكم ما لم يؤت أحدا من العالمين And (remember) when Musa (Moses) said to his people: "O my people! Remember the Favour of Allah to you, when He made Prophets among you, made you kings, and gave you what He had not given to any other among the 'Alamin (mankind and jinns, in the past)".	0.747
(6:100, 17:67)	وجعلوا لله شركاء الجن وخلقهم وخرقوا له بنين وبنات بغير علم سبحانه وتعالى عما يصفون But they have attributed to Allah partners—the jinn, while He has created them—and have fabricated for Him sons and daughters. Exalted is He and high above what they describe	وإذا مسكم الضر في البحر ضل من تدعون إلا إياه فلما نجاكم إلى البر أعرضتم وكان الإنسان كفور And when harm touches you upon the sea, those that you call upon besides Him vanish from you except Him (Allah Alone). But when He brings you safely to land, you turn away (from Him). And man is ever ungrateful.	0.996

5. Benchmarking QurSim: A Comparative Evaluation of Classification and Semantic Search

The Doc2vec-based approach proposed by [22] uses cosine similarity to measure the semantic relationship between pairs of Quranic verses. In contrast, our methodology employs transformer-based models such as AraBERT and CAMeLBERT to reformulate the task as a classification and semantic search problem. On the QurSim dataset, our models achieved an F1-score of 91.3% and an accuracy of 92.5%, demonstrating superior performance in capturing semantic similarity relationships compared to the scores reported in prior work (see Table 7).

Table 7. Benchmark results for similarity search and classification tasks on the QurSim dataset. MiniLm = paraphrase-multilingual-MiniLM-L12-v2.

Model	Method	Dataset	Fine-Tuning	F1-Score	Accuracy
Similarity Search Task					
Baseline Doc2vec	Cosine Similarity	Qursim	Ar	0.67	0.76
AraBERTv02(Ours)	Cosine Similarity	Qursim	Ar	0.87	0.88
CAMelBERT	Cosine Similarity	Qursim	Ar	0.91	0.91
MiniLM	Cosine Similarity	Qursim	Ar	0.80	0.81
Classification Task					
AraBERTv02	Transformers	Qursim	Ar	0.90	0.90
CAMelBERT	Transformers	Qursim	Ar	0.90	0.90
MiniLM	Transformers	Qursim	Ar	0.78	0.78
MiniLM	Transformers	Qursim	En	0.98	0.98
MiniLM	Transformers	Qursim	Ur	0.98	0.98
MiniLM	Transformers	Qursim	A + En + Ur	0.98	0.98
MiniLM	Transformers	Qursim	Ar + MultiLang	0.98	0.98

6. Related Works

The advancement of pre-trained word embeddings [23,24] in conjunction with transformer-based models [1,25] has enabled the seamless integration of cross-lingual learning [3,26]. This integration enables the incorporation of semantic knowledge [27] within a single language and across multiple languages. Ref. [28] utilized multilingual large language models (MLLMs) to assess semantic similarity across Quranic translations, providing insights into the semantic connections between various languages. Ref. [29] studied cross-lingual semantic similarity tailored to Al-Quran verses in different languages. Their approach integrates both word alignment and semantic vector representations, with the experimental results highlighting its efficacy in discerning verses that exhibit semantic similarity.

Ref. [30] conducted a study on the semantic relatedness of Quranic verses using the AraBERT model. They used the QurSim dataset and compared two versions of AraBERT (v2 and v0.2). Their work addressed challenges such as imbalanced datasets and lexical synonyms, achieving an accuracy of 92 percent with AraBERT v0.2. Ref. [22] employed NLP techniques to gauge semantic similarity between Quranic verses. They employed the Doc2Vec model from [31], trained with Gensim, and assessed performance using cosine similarity. Their model achieved an accuracy of 76 percent, encountering limitations in capturing contextual distinctions between verses. Ref. [32] assessed ten English translations of Quranic verses, employing both cosine similarity and semantic similarity measures. The findings indicated that semantic similarity surpassed cosine similarity, particularly following preprocessing, underscoring its effectiveness in capturing similarities within translations.

7. Conclusions

In this study, we analyzed large language models (LLMs), including monolingual and multilingual models, to assess their effectiveness with a specific focus on the Arabic language. We employed AraBERTv0.2 and CAMelBERT for monolingual experiments and paraphrase-multilingual-MiniLM-L12-v2 for the multilingual approach. Our evaluation comprised two tasks, classification and verse retrieval, using the QurSim dataset to identify the most semantically relevant verses.

Our findings indicate that AraBERTv0.2 generally outperforms overall, but its performance is occasionally low due to the limitation of a small dataset. Although the multilingual model shows a high F1-score, a closer analysis reveals shortcomings in the results. Notably, when combining English, Urdu, and Arabic with other multiple-language translations, the multilingual model achieved the highest score due to being trained on a large dataset. In conclusion, for the Arabic language, monolingual models tend to perform well overall.

Author Contributions: Conceptualization, S.A.R. and M.G.A.M.; Software, T.A.; Validation, M.G.A.M.; Formal analysis, T.A. and Muhammad Imran; Investigation, T.A and S.A.R.; Resources, M.G.A.M.; Data curation, T.A. and S.A.R.; Writing—original draft, T.A. and S.A.R.; Writing—review & editing, M.G.A.M. and M.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Prince Sultan University for financial (APC) and technical support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: The authors would like to thank Prince Sultan University for providing financial and technical support.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

This section provides a detailed analysis of the similarity classification results presented in Tables A1 and A2. We conducted a comprehensive examination of the predictions generated by our models, drawing upon the extensive analysis we undertook. Notably, verses [6:152, 83:1], which are presented in Table A1, represent the pair verses. However, it is noteworthy that the AraBERTv0.2 model failed to predict and assign a label of 0 to these verses. [6:152] emphasizes fairness in the property of orphans and speaking the truth even when it may involve close relatives and fulfilling the commitments made by Allah. [83:1] suggests that those who are miserly or give less than what is due, particularly in the matter of charity, will face consequences. Both verses [20:132, 51:57] stress the importance of fairness and justice. The concepts of providing what is due and upholding fairness are essential elements in both giving to others and handling property. Both verses emphasize God's self-sufficiency and independence, highlighting that He is not dependent on people or any other being for material sustenance. God, on the other hand, is the ultimate Provider who cares for and supports His creation. However, both monolingual and multilingual models do not understand the contextual meaning of the verse. Table A2 demonstrates promising results for Urdu, English, and Arabic when utilizing paraphrase-multilingual-MiniLM-L12-v2. These results suggest that training on a large dataset shows favorable performance across all these languages.

Table A1. Classification of verse relevance degree using verses as input, where AraBERT denotes AraBERTv0.2 and MiniLM denotes MiniLM-L12-v2.

			AraBERT	**MiniLM**	
Chapter:Verse	**Source Verse**	**Target Verse**	**Predicted Value**	**Predicted Value**	**QurSim Value**
(55:27, 76:9)	ويبقى وجه ربك ذو الجلال والإكرام And the Face of your Lord full of Majesty and Honour will abide forever	إنما نطعمكم لوجه الله لا نريد منكم جزاء ولا شكورا We feed you seeking Allah's Countenance only. We wish for no reward, nor thanks from you	1	1	1
(6:152, 83:1)	ولا تقربوا مال اليتيم إلا بالتي هي أحسن حتى يبلغ أشده وأوفوا الكيل والميزان بالقسط لا نكلف نفسا إلا وسعها وإذا قلتم فاعدلوا ولو كان ذا قربى وبعهد الله أوفوا ذلكم وصاكم به لعلكم تذكرون And come not near to the orphan's property, except to improve it, until he (or she) attains the age of full strength; and give full measure and full weight with justice. We burden not any person, but that which he can bear. And whenever you give your word (i.e., judge between men or give evidence, etc.), say the truth even if a near relative is concerned, and fulfill the Covenant of Allah, This He commands you, that you may remember.	ويل للمطففين Woe to those who give less [than due]	0	1	1

Table A1. *Cont.*

			AraBERT	MiniLM	
Chapter:Verse	**Source Verse**	**Target Verse**	**Predicted Value**	**Predicted Value**	**QurSim Value**
(28:47, 4:165)	ولولا أن تصيبهم مصيبة بما قدمت أيديهم فيقولوا ربنا لولا أرسلت إلينا رسولا فنتبع آياتك ونكون من المؤمنين And if (We had) not (sent you to the people of Makkah) in case a calamity should seize them for (the deeds) that their hands have sent forth, they should have said: “Our Lord! Why did You not send us a Messenger? We should then have followed Your Ayat (Verses of the Quran) and should have been among the believers.	رسلا مبشرين ومنذرين لئلا يكون للناس على الله حجة بعد الرسل وكان الله عزيزا حكيما Messengers as bearers of good news as well as of warning in order that mankind should have no plea against Allah after the Messengers. And Allah is Ever All-Powerful, All-Wise.	1	1	1
(20:132, 51:57)	وأمر أهلك بالصلاة واصطبر عليها لا نسألك رزقا نحن نرزقك والعاقبة للتقوى And enjoin As-Salat (the prayer) on your family, and be patient in offering them [i.e., the Salat (prayers)]. We ask not of you a provision (i.e., to give Us something: money, etc.); We provide for you. And the good end (i.e., Paradise) is for the Muttaqun.	ما أريد منهم من رزق وما أريد أن يطعمون I seek not any provision from them (i.e., provision for themselves or for My creatures) nor do I ask that they should feed me.	0	0	1

Table A2. Classification of verse relevance degree using verses as input using ar, ur, and en.

	Paraphrase-MiniLM-L12-v2		**Predicted Value**	**QurSim Value**
Chapter:Verse	**Source Verse**	**Target Verse**		
(21:8, 31:20)	Nor did We make them bodies that ate no food nor were they immortal	Some dispute about Allah though they have neither knowledge nor guidance nor an illuminating Book	1	1
(10:26, 55:60)	جن لوگوں ني نکی کی هي ان کي واسطي نکی بهی هي اور اضافه بهی اور ان کي چهروں پر نه ساہی ہوگی اور نه ذلت وه جنّت والي ہس اور وہس ہمدشه رہني والي ہس	کىا احسان کا بدلہ احسان کي علاوه کچھ اور بهی ہوسکتا هي	1	1

Table A2. *Cont.*

	Paraphrase-MiniLM-L12-v2		Predicted Value	QurSim Value
Chapter:Verse	**Source Verse**	**Target Verse**		
(6:100, 18:50)	وجعلوا لله شركاء الجن وخلقهم وخرقوا له بنين وبنات بغير علم سبحانه وتعالى عما يصفون But they have attributed to Allah partners—the jinn, while He has created them—and have fabricated for Him sons and daughters. Exalted is He and high above what they describe	وإذ قلنا للملائكة اسجدوا لآدم فسجدوا إلا إبليس كان من الجن ففسق عن أمر ربه أفتتخذونه وذريته أولياء من دوني وهم لكم عدو بئس للظالمين بدلا And (remember) when We said to the angels; “Prostrate to Adam”. So they prostrated except Iblis (Satan). He was one of the jinns; he disobeyed the Command of his Lord. Will you then take him (Iblis) and his offspring as protectors and helpers rather than Me while they are enemies to you? What an evil is the exchange for the Zalimun (polytheists, and wrong-doers, etc.).	1	1

References

1. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Burstein, J., Doran, C., Solorio, T., Eds.; Association for Computational Linguistics: Minneapolis, MN, USA, 2019; Volume 1 (Long and Short Papers); pp. 4171–4186. Available online: https://aclanthology.org/N19-1423 (accessed on 27 October 2024).
2. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving Language Understanding by Generative Pre-Training. OpenAI. 2018. Available online: https://openai.com/index/language-unsupervised/ (accessed on 27 October 2024).
3. Conneau, A.; Lample, G. Cross-lingual language model pretraining. In Proceedings of the Advances in Neural Information Processing Systems 22 (NIPS 2009), Vancouver, BC, Canada, 7–10 December 2009; Volume 32.
4. Alsaleh, D.; Larabi-Marie-Sainte, S. Arabic Text Classification Using Convolutional Neural Network and Genetic Algorithms. *IEEE Access* **2021**, *9*, 91670–91685. [CrossRef]
5. Antoun, W.; Baly, F.; Hajj, H. AraBERT: Transformer-based Model for Arabic Language Understanding. In Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, Marseille, France, 12 May 2020; Al-Khalifa, H., Magdy, W., Darwish, K., Elsayed, T., Mubarak, H., Eds.; European Language Resource Association: Marseille, France, 2020; pp. 9–15, ISBN 979-10-95546-51-1. Available online: https://aclanthology.org/2020.osact-1.2 (accessed on 27 October 2024).
6. Ayed, R.; Chouigui, A.; Elayeb, B. A new morphological annotation tool for Arabic texts. In Proceedings of the 2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA), Aqaba, Jordan, 28 October–1 November 2018; pp. 1–6.

7. Bashir, M.H.; Azmi, A.M.; Nawaz, H.; Zaghouani, W.; Diab, M.; Al-Fuqaha, A.; Qadir, J. Arabic natural language processing for Qur'anic research: A systematic review. *Artif. Intell. Rev.* **2023**, *56*, 6801–6854. [CrossRef]
8. Salama; Aref, R.; Youssef, A.; Fahmy, A. Morphological word embedding for arabic. *Procedia Comput. Sci.* **2018**, *142*, 83–93. [CrossRef]
9. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv* **2019**, arXiv:1907.11692
10. Semary, N.A.; Ahmed, W.; Amin, K.; Pławiak, P.; Hammad, M. Improving sentiment classification using a RoBERTa-based hybrid model. *Front. Hum. Neurosci.* **2023**, *17*, 1292010. [CrossRef]
11. Feijó, D.d.; Moreira, V.P. Mono vs Multilingual Transformer-based Models: A Comparison across Several Language Tasks. *arXiv* **2007**, arXiv:2007.09757v1.
12. Přibáň, P.; Steinberger, J. Are the Multilingual Models Better? Improving Czech Sentiment with Transformers. In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021), Online, 1–3 September 2021; Mitkov, R., Angelova, G., Eds.; pp. 1138–1149. Available online: https://aclanthology.org/2021.ranlp-1.128 (accessed on 27 October 2024).
13. Zhao, Z.; Aletras, N. Comparing Explanation Faithfulness between Multilingual and Monolingual Fine-tuned Language Models. *arXiv* **2024**, arXiv:2403.12809.
14. Touati-Hamad, Z.; Laouar, M.R.; Bendib, I. Quran content representation in NLP. In Proceedings of the 10th International Conference on Information Systems and Technologies, Paris France, 12–14 August 2020; pp. 1–6.
15. Alhawarat, M. Extracting topics from the holy Quran using generative models. *Int. J. Adv. Comput. Sci. Appl.* **2015**, *6*, 288–294. [CrossRef]
16. Sharaf, A.M.; Atwell, E. QurSim: A corpus for evaluation of relatedness in short texts. In Proceedings of the LREC, Istanbul, Turkey, 23–25 May 2012; pp. 2295–2302.
17. Kathir, I. Tafsir. In *Islamic Bioethics: Problems and Perspectives*; Springer: Dordrecht, The Netherlands, 1966; Volume 3.
18. Antoun, W.; Baly, F.; Hajj, H. AraBERT: Transformer-based Model for Arabic Language Understanding. In Proceedings of the LREC 2020 Workshop Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; p. 9.
19. Inoue, G.; Alhafni, B.; Baimukan, N.; Bouamor, H.; Habash, N. The Interplay of Variant, Size, and Task Type in Arabic Pre-trained Language Models. In Proceedings of the Sixth Arabic Natural Language Processing Workshop, Kyiv, Ukraine, 19 April 2021; Habash, N., Bouamor, H., Hajj, H., Magdy, W., Zaghouani, W., Bougares, F., Tomeh, N., Farha, I.A., Touileb, S., Eds.; Association for Computational Linguistics: Kyiv, Ukraine, 2021; pp. 92–104. Available online: https://aclanthology.org/2021.wanlp-1.10 (accessed on 27 October 2024).
20. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; Inui, K., Jiang, J., Ng, V., Wan, X., Eds.; Association for Computational Linguistics: Hong Kong, China, 2019; pp. 3982–3992. Available online: https://aclanthology.org/D19-1410 (accessed on 27 October 2024).
21. Lahitani, A.R.; Permanasari, A.E.; Setiawan, N.A. Cosine Similarity to Determine Similarity Measure: Study Case in Online Essay Assessment. In Proceedings of the 2016 4th International Conference on Cyber and IT Service Management, Bandung, Indonesia, 26–27 April 2016; pp. 1–6.
22. Alshammeri, M.; Atwell, E.; Alsalka, M.a. Detecting semantic-based similarity between verses of the Quran with Doc2vec. *Procedia Comput. Sci.* **2021**, *189*, 351–358. [CrossRef]
23. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the Advances in Neural Information Processing Systems 26 (NIPS 2013), Lake Tahoe, NV, USA, 5–8 December 2013.
24. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
25. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
26. Artetxe, M.; Schwenk, H. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. In *Transactions of the Association for Computational Linguistics*; MIT Press: Cambridge, MA, USA, 2019; Volume 7, pp. 597–610.
27. Shah, T.Z.; Imran, M.; Ismail, S.M. A diachronic study determining syntactic and semantic features of Urdu-English neural machine translation. *Heliyon* **2024**, *10*, e22883. [CrossRef] [PubMed]
28. Afzal, T.; Rauf, S.A.; Majid, Q. Semantic Similarity of the Holy Quran Translations with Sentence-BERT. In Proceedings of the 2023 20th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, 22–25 August 2023; pp. 285–290. [CrossRef]

29. Amelia, R.; Bijaksana, M.A. Cross-Lingual Semantic Similarity in Pieces of Al-Quran Verses Translation Using Word Alignment and Semantic Vector Approach. *Int. Semin. Inf. Commun. Technol.* **2019**, *1*, 18–29.
30. Alsaleh, A.N.; Atwell, E.; Altahhan, A. Quranic verses semantic relatedness using arabert. In Proceedings of the Sixth Arabic Natural Language Processing Workshop, Kyiv, Ukraine, 19 April 2021; Habash, N., Bouamor, H., Hajj, H., Magdy, W., Zaghouani, W., Bougares, F., Tomeh, N., Farha, I.A., Touileb, S., Eds.; Association for Computational Linguistics: Kyiv, Ukraine, 2021; pp. 185–190.
31. Le, Q.; Mikolov, T. Distributed representations of sentences and documents. In *International Conference on Machine Learning*; PMLR: London, UK, 2014; pp. 1188–1196.
32. Al Ghamdi, N.M.; Khan, M.B. Assessment of performance of machine learning based similarities calculated for different English translations of Holy Quran. *Int. J. Comput. Sci. Netw. Secur.* **2022**, *22*, 111–118.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.