

Human vs. AI: A Comparative Analysis of Translating Near-Synonymy in the Quran

Bilal M.J. Ibrahim Zakarneh

Department of Languages and Culture, College of Humanities and Sciences, Ajman University,
Ajman, UAE

Email: b.ibrahim@ajman.ac.ae

Yasser Mohammed Hamid Ahmed Alrefaee

Al-Baydha University, Yemen

yasser.alrefaee@gmail.com

How to cite:

Zakarneh, B., & Alrefaee, Y. (2026). Human vs. AI: A Comparative Analysis of Translating Near-Synonymy in the Quran. *International Journal of Linguistics and Translation Studies*, 7(4).35-49. <https://doi.org/10.36892/ijlts.v7i4.786>

ARTICLE HISTORY

Received:
25/08/2026

Accepted:
29/09/2026

Keywords:

Qur'anic translation, near-synonymy, GPT-4, lexical semantics, tafsir, translation evaluation

Abstract

Near-synonymy in the Qur'an challenges translators because related lexical items can differ in intensity, scope, expressive force, and contextual function. This exploratory qualitative study compares selected English expressions by Arberry and Irving with excerpts reported as GPT-4 outputs. Eleven lexical items in ten verse contexts are examined through a translation-oriented adaptation of Murphy's Relation by Contrast approach. The analysis draws on published, tafsir-informed semantic profiles and distinguishes local semantic adequacy, preservation of lexical contrast, contextual fit, and English idiomaticity. The results vary by item. All three products maintain lexical differentiation between bathth and huzn in Qur'an 12:86. Irving's 'famine' and the AI expression 'severe hunger' express different aspects of the masghabah/jū' contrast. Arberry's 'grievous sin' makes religious condemnation explicit, while 'fault' remains defensible for khaṭī'ah. The AI excerpts make regret and private speech explicit, but 'despair impatiently' is an unusual collocation. These findings support evaluation by specific semantic features rather than an overall human-AI ranking. The contribution is an account of how lexical contrast, interpretive choice, and English usage interact in selected translation examples. Incomplete AI provenance, excerpt-level evidence, and the absence of independent ratings limit the conclusions. A supervised revision workflow is proposed for subsequent testing.

1. INTRODUCTION

Qur'anic translation requires decisions about lexical items whose meanings overlap but are not interchangeable in every context. Research on Qur'anic near-synonymy shows that apparently synonymous expressions may differ in denotation, connotation, intensity, participant structure, temporal orientation, or contextual function (Abdul-Raof, 2001; Abdul-Ghafour et al., 2017, 2020, 2022; Alrefaee et al., 2023). Near-synonymy is therefore best treated as a relational problem: two items share a semantic domain, yet the choice of one rather than the other is motivated by a contextually relevant difference.

For a translator, two questions follow. First, is an English expression acceptable in its immediate verse? Second, does the set of English choices convey the relevant differences among related Arabic items? These questions can yield different answers. ‘Hunger’ may be appropriate for *jū*, yet its unmodified repetition for *masghabah* can leave severity unexpressed. Assessment must therefore consider both individual expressions and their relationships. It must also consider whether syntax or surrounding wording conveys a distinction that the selected English noun does not express by itself.

Large language models make this problem newly important. They can generate fluent translations and explanations, but their outputs depend on the model label, prompt, conversation context, sampling behavior, and possible exposure to published translations during training. The GPT-4 technical report states that the model was trained on a mixture of publicly available and licensed data but does not disclose the complete training corpus (OpenAI, 2023). Consequently, an acceptable model output demonstrates performance on a particular prompt; it does not by itself prove independent recovery of an exegetical distinction.

The comparison concerns three differently produced translation products. Arberry and Irving produced complete translations, whereas the AI material comprises selected expressions reported under a GPT-4 label. The study examines the semantic effects of these choices within identified occurrences; it does not treat their production conditions as a matched experiment.

The study addresses three research questions:

1. How adequately do the selected Arberry, Irving, and reported GPT-4 expressions convey the contextually supported meanings of the Arabic items?
2. Which relevant lexical contrasts are expressed, weakened, or left uncertain in each output set?
3. What do these observations imply for a supervised human-AI workflow in Qur’anic translation?

The study distinguishes the adequacy of an individual expression from its ability to preserve a contrast with a related item. Cases identified in earlier near-synonym research are reassessed alongside the AI excerpts, using verse-specific lexical and exegetical evidence (Abdul-Ghafour et al., 2022; Alrefaee et al., 2023). The contribution is the explicit reasoning behind these comparisons across grief, hunger, wrongdoing, and concealment. The literature-guided sample supports close analysis of the selected cases rather than generalization to the Qur’anic lexicon.

2. LITERATURE REVIEW

2.1. Qur’anic Near-Synonymy

Near-synonyms share substantial meaning without necessarily having identical distributions or contextual uses. A translation can convey their shared semantic domain while leaving a relevant difference implicit. In Qur’anic discourse, such differences must be established through contextual and exegetical evidence rather than assumed from lexical variation alone (Abdul-Ghafour et al., 2022; Alrefaee et al., 2023).

A broad English item such as ‘hunger,’ ‘secret,’ or ‘fault’ can convey a shared semantic domain while leaving a more specific feature implicit. Assessment therefore considers the expression

within its verse and its relationship to other selected expressions. Repeating an English word is evidence of reduced lexical differentiation only when the Arabic distinction is supported and is not conveyed elsewhere in the translation.

Research on Qur'anic near-synonymy documents differences in contextual and expressive meaning and their implications for translation (Abdul-Ghafour et al., 2017, 2020, 2022; Alrefaee et al., 2023). Such research supports combining lexical analysis with tafsir, but exegetical interpretation is not a uniform answer key. A distinction proposed for one verse should not be assumed to define every occurrence of the same Arabic item.

2.2.Sacred-Text Translation and Exegetical Accountability

Sacred-text translation intensifies lexical-equivalence problems because the target text represents a religious source. Qur'anic translation necessarily interprets syntax, implicit relations, polysemy, and near-synonymy. Abdul-Raof (2001) therefore treats discourse, texture, and exegesis as integral: formal closeness alone may still obscure rhetorical organization or contextually activated meaning.

This responsibility creates tension between lexical transparency and natural English. A literal equivalent may preserve a distinction but sound obscure; a communicative equivalent may read smoothly while narrowing the Arabic. Neither strategy is automatically superior. The judgment depends on purpose, audience, the feature at stake, and whether compensation is possible through modifiers, syntax, or notes. Lexical choice, contextual fit, and idiomaticity should therefore be assessed separately.

Exegetical descriptions such as 'more intense,' 'collective,' or 'unintentional' require a stated source and a specified context. In this study, a profile is an interpretive criterion for the selected occurrence, not an invariant dictionary definition. Alternative English wording remains acceptable where it conveys the relevant meaning, even if it does not match a reference gloss exactly.

2.3.Large Language Models in Multilingual and Arabic Translation

Instruction-tuned LLMs widen the range of translation tools because they can follow contextual instructions, generate alternatives, explain choices, and revise output. These affordances are useful for difficult texts, but they complicate evaluation: wording may change with the prompt, model snapshot, conversation history, and decoding conditions.

Large-scale evidence shows uneven performance. Zhu et al. (2024) found that GPT-4 surpassed a strong supervised baseline in a substantial minority of translation directions but still trailed commercial systems, especially in lower-resource settings. General benchmark success therefore cannot establish sensitivity to rare, domain-specific lexical distinctions.

Arabic adds challenges of historical variation, morphology, and diglossia. Kadaoui et al. (2023) found different results across ten Arabic varieties: LLMs were competitive in some dialectal directions but trailed commercial systems on Classical Arabic and MSA. Competence in contemporary Arabic therefore cannot be assumed to entail reliable handling of Qur'anic constructions or exegetically specialized senses.

2.4. AI-Assisted Translation of Religious and Qur'anic Texts

Religious-text research reports useful output alongside domain-specific difficulties. Wei (2024) evaluates three LLMs translating classical Chinese Buddhist material and identifies problems with specialized language and interpretation. This is evidence from a different source tradition and language pair; it provides background for domain-sensitive evaluation rather than direct proof of a Qur'anic translation error.

Khair and Sawalha (2025) report an LLM-assisted translation workflow for Islamic literature from the Al-Shamela library. Their use of back-translation and embedding similarity offers scalable validation. However, broad sentence similarity does not, by itself, show whether a translation preserves a particular lexical distinction. This limitation motivates a complementary qualitative assessment.

Qur'an-focused studies use different objects of assessment. Gaanoun and Alsuhaibani (2025) apply AI-assisted sentiment analysis to seven existing English translations. Metwally and Bin-Hady (2025) compare Saheeh International with the Monica application for metonymic expressions. These studies concern particular texts, products, and evaluative features, which limits their relevance to a general comparison of human and AI translation.

Shormani and Alfahad (2025) analyze ChatGPT translations of three religion-oriented academic abstracts and report semantic and register problems. More directly, Shehab (2026) examines AI renderings of Qur'anic near-synonyms, and Shehab and Badah (2026) compare three LLMs with Abdel Haleem's translation across sixteen cognitive-synonym pairs. Their findings motivate attention to context and interpretive criteria while establishing that AI translation of Qur'anic near-synonyms is already an active research area.

2.5. Evaluation Prompting and Reproducibility

Qian et al. (2024) show that the information supplied to an LLM and the prompting method affect machine-translation evaluation. For qualitative comparison, this dependence reinforces the need to document the verse, textual source, prompt, and evaluative criterion. Automated evaluation research provides methodological background; the present judgments rest on the lexical and contextual evidence discussed for each item.

Changes to commercial models complicate reproducibility, and a model-family label does not identify a particular snapshot or interface. Repeated generations can reveal variability that an isolated excerpt cannot show. Interpretation of the present AI material is therefore restricted to the recorded expressions and their reported provenance.

Fluency and fidelity must also be separated. Polished English can conceal generalization, explicitation, or inconsistency, while a marked phrase may preserve an important distinction. Multidimensional assessment should explain each judgment and reduce the risk of choosing a preferred rendering first and constructing its rationale afterward.

2.6. Research Gap and Contribution

Previous research has examined Qur'anic near-synonymy in human translation and extended the inquiry to AI output (Abdul-Ghafour et al., 2022; Alrefaee et al., 2023; Shehab, 2026; Shehab & Badah, 2026). The present study brings selected grief-related items and contrasts involving hunger, wrongdoing, and concealment into one excerpt-level comparison. It separates four evaluative dimensions and examines judgments that can become overly

dependent on a preferred English gloss. The contribution lies in identifying what a rendering conveys, which contrast remains implicit, and where alternative interpretations require a qualified judgment.

3. THEORETICAL AND EVALUATIVE FRAMEWORK

3.1.Relation by Contrast and Synonymy

Murphy's (2003, Chapter 4) Relation by Contrast approach provides a metalexical account of synonymy based on similarity among word-concepts whose forms differ. Following its translation-oriented application in Abdul-Ghafour et al. (2022), RC-S is used here to organize comparisons among related lexical choices. The four dimensions in Table 1 are an operational adaptation developed for this analysis. They are not a standardized scoring scale supplied by Murphy; Arabic contextual evidence establishes the contrasts to be evaluated.

3.2.Local Meaning and Relational Contrast

Local semantic adequacy concerns the meaning conveyed in the selected verse. Relational contrast concerns whether the wording differentiates an item from its selected neighbor along a source-supported dimension. For example, 'hunger' and 'severe hunger' distinguish intensity, while 'hunger' and 'famine' additionally suggest scarcity affecting a wider community. These alternatives may preserve different aspects of a contrast without being equally specific.

3.3.Contextual Fit and Idiomaticity

Contextual fit concerns the expression's compatibility with the surrounding discourse, including its participants, evaluative force, and rhetorical function. Idiomaticity concerns natural English phrasing, syntax, and register. An expression can satisfy one criterion while raising difficulties under another, so semantic coverage and target-language naturalness are assessed separately.

3.4.Evidence and Interpretive Confidence

The contextual profiles combine the tafsir-informed analysis in Abdul-Ghafour et al. (2022) with directly consulted commentaries. Al-Tabari (n.d.) is consulted for Qur'an 12:86, and the online abridged Ibn Kathir commentary for 6:64, 69:50, 4:112, and 14:21 (Ibn Kathir, n.d.-a, n.d.-b, n.d.-c, n.d.-d). The online sources are cited as reproduced commentaries, without assigning them an unverified print-edition date. Each profile guides assessment of a particular occurrence; alternative interpretations remain relevant when the evidence does not uniquely establish a meaning.

Table 1. *Qualitative evaluation rubric*

Dimension	Primary question	Evidence considered	Descriptive terms
Local adequacy	Does the expression convey the supported local meaning?	Denotation and expressive force in the selected context	Aligned / partially aligned / not aligned
Relational contrast	Is the relevant difference conveyed?	Lexical choice, modifiers, and available context	Preserved / weakened / uncertain
Contextual fit	Does the expression suit the discourse?	Participants, evaluation, and interpretive narrowing	Strong / defensible / weak
Idiomaticity	Is the expression natural English?	Collocation, syntax, and register	Natural / marked / awkward

The rubric supports descriptive judgments, with their reasons stated in the item-level discussion. 'Aligned,' 'partially aligned,' and 'not aligned' refer to semantic correspondence

rather than exact agreement with a preferred gloss. ‘Preserved,’ ‘weakened,’ and ‘uncertain’ describe the relevant lexical contrast. Tables 2 and 3 present the expressions and contextual profiles; no aggregate score or statistical ranking is calculated. A dimension is not assessable when the excerpt does not provide sufficient evidence.

4. METHODOLOGY

4.1. Research Design

The study uses an exploratory qualitative comparison of selected textual cases (Creswell & Poth, 2018). The evidence consists of translation excerpts, published lexical analyses, and directly consulted verse commentaries. This design supports explanation of particular lexical choices and contrasts, with conclusions bounded by the available texts and their different production conditions.

4.2. Corpus and Selection

The corpus comprises 11 lexical items in 10 identified verse contexts, listed in Appendix B. Selection follows the contrasts discussed in Abdul-Ghafour et al. (2022) and Alrefaee et al. (2023). Only excerpts associated with an identified Qur’anic occurrence are included; an unlocated ghamm expression is excluded. Lexical labels identify an item or word family rather than always reproducing its inflected surface form. For example, *jazaʿ* labels the verbal form *a-jaziʿ nā* in Qur’an 14:21.

4.3. Translation Sources

The human comparison uses selected expressions attributed to Arberry and Irving in the supplied data record. Abdul-Ghafour et al. (2022) independently reproduce the six additional human examples in Table 3. All selected Arberry expressions were corroborated against the Quranic Arabic Corpus, and accessible or indexed IslamAwakened reproductions corroborate several Irving expressions (Dukes, n.d.; IslamAwakened, n.d.). These sources establish the reproduced wording rather than a particular printed edition. Accessible human clauses are considered where surrounding wording qualifies a lexical judgment; the comparison does not characterize either translator’s complete work.

The AI data record identifies the excerpts as GPT-4 outputs collected on October 21, 2025, and supplies the prompt reproduced in Appendix A. The prompt requests a modern, natural-English rendering with neighboring Arabic verses for context. Complete responses and original conversation records are not part of the analytic dataset. The precise model snapshot, interface, input strings, decoding settings, retries, and selection procedure are therefore unspecified. GPT-4 is retained as a reported attribution; the analysis does not independently authenticate the collection metadata.

4.4. Analytical Procedure

The analytical unit is a selected English expression associated with an identified Arabic occurrence. Analysis proceeds by establishing the contextual profile from the cited source, examining each expression against that profile, and comparing the products for the same occurrence. Related items are then considered together along a supported semantic dimension. The discussion records alternative interpretations and separates local adequacy, contrast preservation, contextual fit, and idiomaticity. Accessible human verse wording is checked for contextual qualification; AI judgments remain restricted to the supplied excerpts. English

glosses accompanying Arabic fragments are explanatory aids. Missing AI responses are not reconstructed, and no newly generated output is substituted for the recorded expressions.

4.5. Reliability and Limitations

The analysis has no blinded independent ratings or agreement statistics. Its principal limitations are the literature-guided sample, excerpt-level evidence, and incomplete AI provenance. The AI excerpts cannot establish whole-verse completeness, consistent model behavior, or compensation through unavailable surrounding wording. The three products were created under different conditions, so any apparent advantage concerns an item and an evaluative dimension. Semantic profiles are contextual interpretations rather than compulsory target-language strings. Possible exposure to published translations during model training remains an alternative explanation, not evidence that copying occurred.

5. FINDINGS AND DISCUSSION

The findings distinguish semantic explicitness from accuracy. A longer expression can make a feature visible while also adding an interpretation that the Arabic context does not require. The discussion therefore considers defensible alternatives alongside the strengths and limitations of each lexical choice.

5.1. Grief-Related Lexemes

5.1.1. *Huẓn and Bathth*

إِنَّمَا أَشْكُو بَثِّي وَحُزْنِي إِلَى اللَّهِ

“I only complain of my bathth and my huẓn to Allah” (Qur’an, Yusuf 12:86).

Al-Tabari’s commentary on 12:86 reports several interpretations of bathth, including worry, need, and intense sorrow, and connects the word with expressing the condition that troubles the speaker (Al-Tabari, n.d.). The passage therefore supports attention to emotional force and expression without fixing a single intensity scale. Irving’s ‘sorrow/sadness,’ Arberry’s ‘anguish/sorrow,’ and the AI pair ‘anguish/grief’ all maintain lexical differentiation. ‘Anguish’ conveys strong distress, so departure from a preferred ‘sorrow/sadness’ pair does not establish an error. The complaint to God also supplies an expressive setting that the selected English nouns need not encode separately.

5.1.2. *Karb and Distress*

وَمِنْ كُلِّ كَرْبٍ

“and from every karb” (Qur’an, Al-An’am 6:64).

The Ibn Kathir commentary places 6:64 within deliverance from dangers on land and at sea, including storms, and from distress more generally (Ibn Kathir, n.d.-a). Arberry and the AI excerpt use ‘distress’; Irving uses ‘anxiety.’ ‘Distress’ covers a broader range of hardship, whereas ‘anxiety’ foregrounds its emotional aspect. Irving’s choice is consequently narrower at the lexical level. Neither the selected phrase nor the consulted commentary establishes a fixed duration, so prolonged suffering is not imposed as a necessary meaning of karb here.

5.1.3. *Ḥasrah and Regret*

وَإِنَّهُ لَحَسْرَةٌ عَلَى الْكَافِرِينَ

“And indeed, it will be a ḥasrah upon the disbelievers” (Qur’an, Al-Haqqah 69:50).

The abridged Ibn Kathir commentary explains ḥasrah in 69:50 through anguish associated with rejection on the Day of Resurrection (Ibn Kathir, n.d.-b). Arberry's 'sorrow' conveys the emotional burden; Irving's 'despair' emphasizes hopelessness. The AI expression 'a source of regret' makes regret explicit, but this explicitness does not establish unique superiority over 'sorrow' or 'anguish.' The commentary also discusses different referents for the pronoun, so the excerpt alone cannot settle the full causal relation. The eschatological setting further shows why regret should not be equated with exclusively past grammatical reference.

5.1.4. Jaza' and Patient Endurance

سَوَاءٌ عَلَيْنَا أَجَزْنَا أَمْ صَبَرْنَا

"It is all the same for us whether we show jaza' or are patient" (Qur'an, Ibrahim 14:21).

In 14:21, a-jazi' nā contrasts with ṣabarnā, or patient endurance. The Ibn Kathir commentary discusses agitated distress in opposition to bearing torment patiently (Ibn Kathir, n.d.-d). Irving's 'whether we panic or act patient' expresses the opposition directly, although 'act patient' is marked. Arberry's 'we cannot endure' presents a failure of endurance and is less explicit about agitation. The AI phrase 'we despair impatiently' combines emotional collapse with impatience but forms an unusual English collocation. These alternatives express different aspects of the opposition; 'panic' is not required as its only possible English realization.

Table 2. Grief-related expressions in identified contexts

Item and verse	Contextual profile	Arberry	Irving	Reported GPT-4
ḥuzn 12:86	Sadness in a paired complaint to God; no unique English gloss required.	sorrow	sadness	grief
bathth 12:86	Reported senses include worry, need, and intense or expressed sorrow; no fixed scale is imposed.	anguish	sorrow	anguish
karb 6:64	Distress in a context of dangers and deliverance; duration is not fixed.	distress	anxiety	distress
ḥasrah 69:50	Anguish or regret in an eschatological setting; the causal referent is interpretive.	sorrow	despair	a source of regret
jaza' 14:21	Agitated distress contrasted with patient endurance; a verbal form occurs here.	we cannot endure	we panic	we despair impatiently

Note. The profiles draw on Al-Tabari (n.d.) for 12:86 and Ibn Kathir (n.d.-a, n.d.-b, n.d.-d) for 6:64, 69:50, and 14:21, respectively. Arberry wording is corroborated in Dukes (n.d.); selected Irving wording is reproduced in accessible or indexed IslamAwakened records (n.d.). GPT-4 wording is reported excerpt data, with provenance described in Appendix A. Lexical differentiation does not by itself establish complete preservation of a semantic contrast.

5.2. Additional Near-Synonym Comparisons

5.2.1. Jū' and Masghabah

لَا يُسْمِنُ وَلَا يُغْنِي مِنْ جُوعٍ

"It neither nourishes nor avails against hunger [jū']" (Qur'an, Al-Ghashiyah 88:7).

أَوْ إِطْعَمٌ فِي يَوْمٍ مَسْغَبَةٍ

"Or feeding, on a day of masghabah" (Qur'an, Al-Balad 90:14).

Abdul-Ghafour et al. (2022) discuss masghabah as more severe or widespread hunger than jū'. All three products use 'hunger' for jū'. For masghabah, Irving's 'famine' highlights wider scarcity, whereas the AI expression 'severe hunger' highlights intensity. Both distinguish the

items, but on different dimensions. Arberry repeats ‘hunger,’ yet the complete phrase ‘a day of hunger’ situates feeding within a time of hardship (Dukes, n.d., Qur’an 90:14). The intensity difference remains lexically implicit; the shared noun alone does not demonstrate complete contextual loss. Neither thirst nor community-wide famine is treated as a necessary feature of every occurrence.

5.2.2. *Khiṭ’ and Khaṭī’ah*

إِنَّ قَتْلَهُمْ كَانَ خِطْئًا كَبِيرًا

“Indeed, their killing was a great khiṭ’” (Qur’an, Al-Isra’ 17:31).

In 17:31, the prohibition of killing children for fear of poverty and the adjective kabīran establish grave wrongdoing. Abdul-Ghafour et al. (2022) favor a rendering with ‘sin.’ Arberry’s ‘grievous sin’ conveys religious condemnation explicitly; the reported AI ‘grave wrong’ conveys moral gravity without the same religious label. Irving’s ‘serious blunder’ emphasizes a serious error and can weaken the explicit moral category. No expression is marked deficient solely because it does not separately state deliberateness, which is supported by the prohibited act and need not be added to the English noun.

وَمَنْ يَكْسِبْ خَطِيئَةً أَوْ إِثْمًا

“And whoever earns a khaṭī’ah or a sin [ithm]” (Qur’an, Al-Nisa’ 4:112).

In 4:112, khaṭī’ah occurs alongside ithm, not khiṭ’. The comparison must therefore account for this local opposition as well as the wrongdoing in 17:31. Abdul-Ghafour et al. (2022) report interpretations contrasting minor or unintentional fault with graver or intentional sin. The abridged Ibn Kathir commentary uses a fault/sin distinction without explicitly requiring that binary interpretation (Ibn Kathir, n.d.-c). Arberry’s and the AI’s ‘fault’ are consequently defensible. Arberry’s full clause also retains the subsequent condemnation of false accusation and manifest sin (Dukes, n.d., Qur’an 4:112). Irving’s ‘blunder’ conveys error but makes moral accountability less explicit in the selected noun. ‘Mistake’ remains one interpretive option rather than a compulsory correction.

5.2.3. *Sirr and Najwa*

فَأِنَّهُ يَعْلَمُ السِّرَّ وَأَخْفَى

“He knows the secret [sirr] and what is even more hidden” (Qur’an, Taha 20:7).

لَا خَيْرَ فِي كَثِيرٍ مِّنْ نَّجْوَاهُمْ

“There is no good in much of their private speech [al-najwā]” (Qur’an, Al-Nisa’ 4:114).

Abdul-Ghafour et al. (2022) distinguish concealed information, al-sirr, from private speech, al-najwā. ‘Secret’ in Arberry and the AI excerpt, and Irving’s ‘secrets,’ convey concealment; the change in number requires clause-level assessment. In 4:114, the Ibn Kathir commentary explains the speech category and its exceptions for charity, goodness, and reconciliation (Ibn Kathir, n.d.-c). Arberry’s ‘conspiring’ and Irving’s ‘intrigue’ emphasize the negative opening and may narrow the lexical category. Arberry nevertheless retains the positive exceptions in his full clause (Dukes, n.d., Qur’an 4:114). The AI’s ‘secret talk’ states private speech directly and is idiomatic. The difference concerns lexical emphasis, not evidence that every human rendering treats all private conversation as malicious.

Table 3. Additional lexical contrasts in identified contexts

Item and verse	Contextual profile	Arberry	Irving	Reported GPT-4
jū‘ 88:7	General hunger.	hunger	hunger	hunger
masghabah 90:14	Severe hunger; a wider scarcity reading is discussed in the cited account.	hunger	famine	severe hunger
khiṭ’ 17:31	Wrongdoing in the prohibition of killing children; kabīran explicitly marks gravity.	grievous sin	serious blunder	grave wrong
khaṭī’ ah 4:112	Fault or wrongdoing alongside ithm; a minor/unintentional reading is not a compulsory gloss.	fault	blunder	fault
al-sirr 20:7	Concealed information.	secret	secrets	secret
al-najwā 4:114	Private speech with morally positive exceptions in the verse.	conspiring	intrigue	secret talk

Note. The profiles and human examples follow Abdul-Ghafour et al. (2022); Arberry expressions and surrounding wording are corroborated in Dukes (n.d.). The discussion of 4:112 and 4:114 also draws on Ibn Kathir (n.d.-c). AI expressions reproduce the recorded excerpts. Judgments distinguish the selected wording from contextual qualification in accessible human clauses.

5.3. Comparison across the Three Products

5.3.1. Arberry

Arberry’s selected expressions make distress and religious condemnation explicit and differentiate bathth from ḥuzn through ‘anguish/sorrow.’ His repetition of ‘hunger’ leaves the proposed jū’/masghabah difference lexically unstated. ‘Sorrow’ for ḥasrah and ‘we cannot endure’ for the jaza’ form express relevant aspects but make regret and agitation less explicit. ‘Fault’ for khaṭī’ ah is defensible; it should not be counted as an error simply because a source prefers ‘mistake.’

5.3.2. Irving

Irving differentiates bathth and ḥuzn, makes the scarcity reading of masghabah explicit with ‘famine,’ and directly conveys agitation with ‘panic.’ ‘Anxiety’ foregrounds one aspect of karb, while ‘despair’ foregrounds hopelessness rather than regret for ḥasrah. ‘Serious blunder’ can weaken the explicit religious condemnation in 17:31, whereas ‘blunder’ in 4:112 requires a more qualified judgment because the contextual interpretation of khaṭī’ ah is not uniquely settled.

5.3.3. AI Translation Excerpts

The AI excerpts make distress, regret, intensified hunger, and private speech explicit. ‘Grave wrong’ conveys moral gravity, and ‘fault’ remains defensible in Qur’an 4:112. The bathth/ḥuzn pair maintains lexical differentiation. ‘Despair impatiently’ is the clearest idiomatic weakness among the selected expressions. These are properties of the reported excerpts; their analysis cannot establish model-wide performance or an overall advantage over either human translator.

5.3.4. Implications for Human and AI Collaboration

The findings motivate a supervised workflow that remains to be tested. An LLM may propose alternatives, and a qualified reviewer may assess them against context, tafsir, lexical evidence, and English usage. The study has not measured the accuracy, speed, or reliability of a hybrid workflow. Its practical implication is to keep candidate generation, source validation, and editorial judgment distinguishable and documented.

5.4. Broader Implications

5.4.1. Local Adequacy and Contrast Preservation

The hunger comparison illustrates how local adequacy and relational differentiation can diverge. ‘Hunger’ conveys the shared domain, ‘severe hunger’ expresses intensity, and ‘famine’ suggests wider scarcity. Arberry’s ‘a day of hunger’ also places the condition within a period of hardship, although the noun leaves intensity implicit. Assessment must therefore identify the relevant feature and consider the clause in which it occurs. One repeated word does not by itself establish complete semantic loss.

Greater English lexical variety is not automatically more faithful. Alternating among ‘sorrow,’ ‘grief,’ ‘anguish,’ and ‘distress’ can create distinctions that are difficult to sustain consistently. The relevant goal is to convey a supported contextual difference, using lexical choice, modification, syntax, or explanatory notes where appropriate.

5.4.2. Differences in Semantic Emphasis

The excerpts show different choices of semantic emphasis. Arberry uses wording that directly states distress and sin, while Irving uses ‘famine’ and ‘panic’ to specify scarcity and agitation. These observations do not establish stable stylistic tendencies across either complete translation. Such an explanation would require a larger corpus and an analysis of the translators’ stated aims.

The reported AI material often uses phrases that make a feature visible, including regret, severity, and private speech. Explicitness is useful when justified by context, but it can overstate one interpretive dimension. The unusual combination ‘despair impatiently’ also shows that including several relevant ideas in a phrase does not guarantee natural English. These are observations about the available record, not explanations of the model’s internal reasoning.

5.4.3. Relationship to Recent AI Translation Findings

The item-specific pattern is compatible with benchmark research showing uneven LLM translation performance across languages and tasks (Kadaoui et al., 2023; Zhu et al., 2024). Studies of Buddhist texts, Islamic literature, and religion-oriented academic abstracts likewise motivate domain-sensitive review (Wei, 2024; Khair & Sawalha, 2025; Shormani & Alfahad, 2025). Their different datasets and validation methods preclude pooling their findings into a single human-AI ranking.

Gaanoun and Alsuhaibani (2025) examine sentiment preservation, and Metwally and Bin-Hady (2025) examine metonymy. Shehab (2026) and Shehab and Badah (2026) address Qur’anic near-synonyms more directly. The present analysis complements this work by distinguishing lexical generality, contextual narrowing, and idiomatic awkwardness, and by allowing that the same expression can be adequate on one dimension and limited on another.

5.4.4. Translation Practice and Future Research

A practical review begins with an identified occurrence and a documented semantic profile. Candidate expressions should then be compared within their full clauses and across the relevant lexical field. Reviewers should record the feature preserved, any interpretive addition, and alternatives that remain defensible. Editing should improve English usage without making a preferred gloss the sole test of fidelity.

A stronger follow-up design would use a larger occurrence-based corpus, matched instructions, several documented models, repeated generations, and independent expert ratings with source identity concealed where feasible. Complete Arabic inputs, outputs, prompts, model identifiers, dates, settings, and selection decisions should be archived. Agreement statistics could then complement, rather than replace, item-level explanation.

6. CONCLUSION

This exploratory comparison examined 11 lexical items in 10 identified Qur'anic verse contexts. A translation-oriented adaptation of RC-S organized four distinct questions: whether the English expression conveys local meaning, whether it differentiates related items, whether it fits the discourse, and whether it is idiomatic. The analysis shows why these judgments should not be collapsed into a single preferred-word test. A broad expression may be defensible while leaving a contrast implicit, and a more explicit phrase may introduce a narrower interpretation.

The results vary by item. All three products differentiate *batth* and *huzn* in 12:86. Irving's 'famine' and the reported AI 'severe hunger' express different aspects of the proposed *masghabah/jū'* contrast. Arberry's 'grievous sin' states religious condemnation directly, while 'grave wrong' conveys moral gravity. In 4:112, 'fault' remains defensible, and the evidence does not justify requiring 'mistake' as its replacement. The reported AI expressions make regret and private speech explicit, but 'despair impatiently' remains awkward. No overall superiority follows from these observations.

The practical implication is a supervised use of generative output as material for review. Candidate generation can be followed by source checking, comparison within full verse contexts, and editing for target-language usage. The present study proposes this process; it does not demonstrate that collaboration improves accuracy or efficiency. Qualified linguistic and exegetical judgment remains necessary to assess both implicit distinctions and interpretive additions.

The conclusions remain bounded by the selected occurrences, reliance on translation excerpts, incomplete AI provenance, and the absence of repeated generations or independent ratings. A larger study should preserve complete records and evaluate documented occurrences under matched conditions. The present comparison provides an explicit account of how lexical contrast, contextual interpretation, and English expression interact in selected Qur'anic translation examples.

DATA AVAILABILITY

The lexical expressions evaluated in this study are presented in Tables 2 and 3, with the verse register in Appendix B. Appendix A describes the reported AI collection metadata and prompt. Original AI conversation records and complete translation responses are not included in the analytic dataset.

ACKNOWLEDGMENT

This research received grant no. (506) from the Arab Observatory for Translation (an affiliate of ALECSO), which is supported by the Literature, Publishing & Translation Commission in Saudi Arabia.

REFERENCES

- Abdul-Ghafour, A. Q. K. M., Awal, N. M., Zainudin, I. S., & Aladdin, A. (2017). Meanings of near-synonyms and their translation issues in the Holy Qur'ān. *GEMA Online Journal of Language Studies*, 17(4), 258–273. <https://doi.org/10.17576/gema-2017-1704-17>
- Abdul-Ghafour, A. Q. K. M., Awal, N. M., Zainudin, I. S., & Aladdin, A. (2020). Meanings of the Qur'anic near-synonyms: Al-ṣirāt and al-sabīl (the path) and their English translations. *Akademika*, 90(3), 63–73. <https://doi.org/10.17576/akad-2020-9003-06>
- Abdul-Ghafour, A. Q. K., Aladdin, A., Zainudin, I. S., & Awal, N. M. (2022). Contextual meanings of Qur'anic near-synonyms and their translations. *Ijaz Arabi Journal of Arabic Learning*, 5(3), 796–811. <https://doi.org/10.18860/ijazarabi.v5i3.16519>
- Abdul-Raof, H. (2001). *Qur'an translation: Discourse, texture and exegesis*. Routledge. <https://doi.org/10.4324/9780203036990>
- Alrefaee, Y., Zakarneh, B., Kumar, T., & Abdul-Ghafour, A. Q. (2023). Contextual meanings of Qur'anic near-synonyms and their English translations with special reference to Huzn (grief) and its near-synonyms. *Journal of Intercultural Communication Research*, 52(5), 477–492. <https://doi.org/10.1080/17475759.2023.2250796>
- Al-Tabari. (n.d.). *Tafsir al-Tabari: Yusuf ayah 86* [Arabic commentary]. Quran.com. <https://quran.com/yusuf/86/tafsirs/ar-tafsir-al-tabari>
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE.
- Dukes, K. (n.d.). *The Quranic Arabic Corpus* [Linguistic and translation database]. Retrieved October 1, 2026, from <https://corpus.quran.com/>
- Gaanoun, K., & Alsuhaibani, M. (2025). Sentiment preservation in Quran translation with artificial intelligence approach: Study in reputable English translation of the Quran. *Humanities and Social Sciences Communications*, 12, Article 222. <https://doi.org/10.1057/s41599-024-04181-0>
- Ibn Kathir. (n.d.-a). *Ibn Kathir (abridged) tafsir of Al-An'am ayah 64*. Quran.com. <https://quran.com/al-anam/64/tafsirs/en-tafsir-ibn-kathir>
- Ibn Kathir. (n.d.-b). *Ibn Kathir (abridged) tafsir of Al-Haqqah ayah 50*. Quran.com. <https://quran.com/al-haqqah/50/tafsirs/en-tafsir-ibn-kathir>
- Ibn Kathir. (n.d.-c). *Ibn Kathir (abridged) tafsir of An-Nisa ayah 112*. Quran.com. <https://quran.com/an-nisa/112/tafsirs/en-tafsir-ibn-kathir>
- Ibn Kathir. (n.d.-d). *Ibn Kathir (abridged) tafsir of Ibrahim ayah 21*. Quran.com. <https://quran.com/ibrahim/21/tafsirs/en-tafsir-ibn-kathir>
- IslamAwakened. (n.d.). *The Qur'an* [Parallel translation database]. Retrieved October 1, 2026, from <https://islamawakened.org/quran/>
- Kadaoui, K., Magdy, S. M., Waheed, A., Khondaker, M. T. I., El-Shangiti, A. O., Nagoudi, E. M. B., & Abdul-Mageed, M. (2023). TARJAMAT: Evaluation of Bard and ChatGPT on machine translation of ten Arabic varieties. In *Proceedings of ArabicNLP 2023* (pp. 52–75). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.arabnlp-1.6>
- Khair, M. M., & Sawalha, M. (2025). Automated translation of Islamic literature using large language models: Al-Shamela Library application. In *Proceedings of the New Horizons in Computational Linguistics for Religious Texts* (pp. 53–58). Association for Computational Linguistics. <https://aclanthology.org/2025.clrel-1.5/>
- Metwally, A. A., & Bin-Hady, W. R. A. (2025). Exploring human vs. AI-powered translation to metonymic expressions: A case study of the Holy Quran. *Social Sciences & Humanities Open*, 12, Article 101615. <https://doi.org/10.1016/j.ssaho.2025.101615>
- Murphy, M. L. (2003). *Semantic relations and the lexicon: Antonymy, synonymy and other paradigms*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511486494>
- OpenAI. (2023). *GPT-4 technical report* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2303.08774>

- Qian, S., Sindhuja, A., Kabra, M., Kanojia, D., Orasan, C., Ranasinghe, T., & Blain, F. (2024). What do large language models need for machine translation evaluation? In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 3660–3674). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.214>
- Shehab, E. (2026). A shallow echo: Artificial intelligence and the semantic flattening of the Qur'an. *Learned Publishing*, 39(3), Article e2063. <https://doi.org/10.1002/leap.2063>
- Shehab, E., & Badah, A. M. (2026). Semantic precision challenges for large language models in Qur'anic cognitive synonyms translation. *Discover Artificial Intelligence*, 6, Article 637. <https://doi.org/10.1007/s44163-026-01207-1>
- Shormani, M. Q., & Alfahad, A. (2025). Artificial intelligence or human: The use of ChatGPT in the academic translation for religious texts. *SAGE Open*, 15(3). <https://doi.org/10.1177/21582440251343954>
- Wei, X. (2024). The use of large language models for translating Buddhist texts from classical Chinese to modern English: An analysis and evaluation with ChatGPT 4, ERNIE Bot 4, and Gemini Advanced. *Religions*, 15(12), Article 1559. <https://doi.org/10.3390/rel15121559>
- Zhu, W., Liu, H., Dong, Q., Xu, J., Huang, S., Kong, L., Chen, J., & Li, L. (2024). Multilingual machine translation with large language models: Empirical results and analysis. In *Findings of the Association for Computational Linguistics: NAACL 2024* (pp. 2765–2781). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.findings-naacl.176>

APPENDIX A. REPORTED AI RECORD

The retained data record identifies GPT-4 as the source and October 21, 2025, as the collection date. It supplies the following prompt template. These are reported metadata; an original transcript or model log is not included in the analytic dataset.

“Translate the following Qur’anic verse into modern, natural English. Provide the translation for the specified verse only, but include the full Arabic text of the verse immediately preceding it, the target verse, and the verse immediately following it for contextual reference.”

Tables 2 and 3 reproduce the selected AI expressions. Complete responses, exact Arabic input strings, verse-specific prompt instances, model snapshot, interface, decoding parameters, retries, and selection rules are unspecified in the record. The template requests neighboring Arabic verses, but their correct reproduction cannot be assessed without the original responses. No new generation is used to replace the historical excerpts.

APPENDIX B. CORPUS REGISTER

Table B1. *Included occurrences and the evaluated lexical items*

Verse	Lexical item or family	Comparison focus
12:86	ḥuzn and bathth	Differentiated emotional expression
6:64	karb	Breadth of distress
69:50	ḥasrah	Regret and anguish
14:21	jaza‘	Agitation and endurance
88:7	jū‘	General hunger
90:14	masghabah	Intensity and scarcity
17:31	khiṭ‘	Moral condemnation
4:112	khaṭī‘ah	Fault in relation to ithm
20:7	al-sirr	Concealed information

Verse	Lexical item or family	Comparison focus
4:114	al-najwā	Private speech

Note. There are 11 lexical items in 10 verse contexts because bathth and ḥuzn co-occur in 12:86. The grief profiles are supported by directly consulted commentaries cited in Table 2. English glosses accompanying Arabic fragments are explanatory aids. Human excerpt sources and accessible clause-level checks are described in the Translation Sources section; the Arberry wording is traceable by chapter and verse in Dukes (n.d.).