Journal of Interdisciplinary Qur'anic Studies

Contemporary Thoughts Press

Shahid Beheshti University

Journal of Interdisciplinary Qur'anic Studies Vol.4, Issue 2, December 2025, 11-23

Generating Inferences from Qur'anic Verses: A Computational Text Mining Approach

Samira Abdi

Master's degree in Computational Qur'an Mining, Interdisciplinary Qur'anic Studies Research Institute, Shahid Beheshti University, Tehran, Iran.

Alireza Talebpour [1]

Associate Professor, Computer Science and Engineering Department, Shahid Beheshti University, Tehran, Iran.

Ahmad Mahmoudi-Aznaveh

Assistant Professor, Cyberspace Research Institute, Shahid Beheshti University, Tehran, Iran.

Mostafa Moradi

Assistant professor, Interdisciplinary Qur'anic Studies Research Institute, Shahid Beheshti University, Tehran, Iran.

Article History: Received 16 February 2025; Accepted 6 April 2025

ABSTRACT:

Original Paper

The Recent advancements in deep learning have yielded novel and significant capabilities in natural language processing (NLP) and automatic inference generation. These capabilities are particularly critical due to their resemblance to human reasoning. At the same time, interdisciplinary initiatives have led to substantial advancements in the realms of knowledge and technology. In this study, the Qur'an is examined as a rich source of multiple concepts and teachings. The objective of this research is to employ natural language processing algorithms to derive meaningful and accurate inferences from the English translation of the Qur'an. Inference is defined as the process of deriving a new and logical sentence from two basic and related sentences. The research methodology introduces a model that utilizes transformers and pre-trained language models. Consequently, we construct the set of all unique unordered verse pairs (i<j) from 6,236 verses, totaling 19,440,730 pair evaluations. A fine-tuned BERT-based classifier labels each pair as either exhibiting or not exhibiting a syllogistic relation. Pairs predicted as "Yes" and exceeding a confidence threshold of 0.80 proceed to the subsequent stage, which is inference generation, while all other pairs are discarded. In the following phase, large language models are employed to generate inferences from the selected pairs of verses.

1. Corresponding Author. Email Address: talebpour@sbu.ac.ir

http://dx.doi.org/10.37264/JIQS.V4I2.1

KEYWORDS: The Qur'an, Artificial Intelligence, Inference, Computational Qur'an Mining, Deep Learning, Natural Language Processing (NLP), Interdisciplinary Qur'anic Studies.

1. Introduction

Inference, as one of the critical processes in human thought, plays a fundamental role in knowledge analysis. The Holy Qur'an, which serves as the foundation for the present research and is rich in knowledge and diverse teachings, underscores the significance of the process of inference and the potency of human reasoning in attaining profound truths. It invites humankind to engage in contemplation of the signs of creation and to draw logical conclusions from them. In this context, the role of inference in comprehending the phenomena of the universe becomes more pronounced. Interdisciplinary research facilitates the integration of diverse fields of knowledge.

Inference from English text represents a sophisticated challenge in NLP, moving beyond superficial pattern matching to derive implicit information and structured knowledge that is not directly stated within a text (Dagdelen et al. 2024). Unlike traditional extraction methods that focus on explicit, token-level matches, inference seeks to reconstruct deep semantic structures, identifying compositional relations, nested attributes, and context-dependent meanings often distributed across sentences and even paragraphs (Dagdelen et al. 2024). The ultimate goal of inference generation is to transform loosely structured, free-form language into analyzable variables, which supports critical applications such as clinical research, materials discovery, and legal analysis (L. Wang et al. 2024). This reconstruction of underlying knowledge is vital for building comprehensive, structured databases that enhance both scientific discovery and real-world decision support systems (Gu et al. 2025).

Inference generation bridges three foundational fields: natural language inference (NLI), information extraction (IE), and domain-specific reasoning (Dagdelen et al. 2024). Like NLI, it requires models to assess logical consequences from a passage; like IE, it demands mapping extracted or inferred knowledge into structured representations. Importantly, domain-specific inference often requires deep conceptual schemas, which general-purpose language models struggle to represent adequately (Dagdelen et al. 2024). For instance, inferring social determinants of health (SDoH) from clinical notes demands familiarity with clinical semantics, social contexts, and healthcare documentation conventions (Gu et al. 2025). Similarly, comprehending research papers in materials chemistry requires

understanding domain-specific nomenclature and implicit conventions (Dagdelen et al. 2024). The ability to fuse linguistic understanding with specialized knowledge enables scalable mining of systems like electronic health records (EHRs) that previously resisted quantitative analysis (Gu et al. 2025). Such advancements herald a major transformation in how unstructured professional texts can be converted into actionable knowledge.

Historically, early extraction methods targeted surface-level patterns, often relying on predetermined rules or keyword matching (Dagdelen et al. 2024). Such techniques failed to capture the complex relational and semantic structures that characterize true inference. The emergence of large language models (LLMs) has enabled a paradigmatic shift, allowing models to perform deep semantic reasoning, infer implicit attributes, and yield structured, machine-readable outputs such as JSON objects. This progression has made it possible to convert raw, unstructured text such as clinical admission records or scientific literature into analyzable datasets (L. Wang et al. 2024). Their ability to model contextual meaning and reconstruct relational structures makes LLMs indispensable for building domain-specific knowledge bases at scale (Dagdelen et al. 2024). Large language models such as GPT-3, GPT-4, PaLM, and Llama have demonstrated remarkable emergent abilities in natural language understanding, long-form reasoning, and open-ended question answering (F. Wang et al. 2024).

Their vast pretrained knowledge allows them to perform complex tasks such as linking dopants to host materials in chemistry papers, extracting biomedical attributes, and interpreting legal concepts across documents (Dagdelen et al. 2024; L. Wang et al. 2024). Their proficiencies extend to multi-hop reasoning, synthesizing information across multiple passages, and generating coherent outputs across diverse domains (F. Wang et al. 2024). These abilities make LLMs promising cognitive tools for automating intellectual tasks traditionally reserved for domain experts.

The necessity of this research can be summarized in terms of time and cost efficiency. The substitution of human reasoning with intelligent inference has been demonstrated to enhance the efficiency of analysis while concomitantly reducing expenses. This transformation plays a pivotal role across various fields and will have substantial impacts. The application of large language models in analyzing pertinent Qur'anic verses has the potential to yield novel insights into Qur'anic exegesis and to enhance comprehension.

The primary objective of this study is to automate the process of inference from Qur'anic verses using advanced NLP algorithms. Interdisciplinary Qur'anic studies, which are aimed at establishing a

connection between religion and science, demonstrate that Qur'anic teachings and scientific findings are not only compatible but can also facilitate the achievement of desired outcomes. They encourage researchers to examine this foundational text from multiple angles and incorporate it into their studies. This research is of particular significance as it is the first to apply intelligent inference to the Holy Qur'an for analysis. The subsequent section will expound upon the research methodology.

However, LLMs come with substantial limitations. Models like PaLM 540B and Llama-3.1 405B require enormous computational resources, and their high inference latency makes them impractical for real-time applications or on-device deployment (F. Wang et al. 2024). Their reliance on cloud APIs introduces ethical and privacy concerns, particularly when handling sensitive data like medical or legal documents. Moreover, general-purpose LLMs often lack the specialized knowledge necessary for expert-level reasoning in medicine, law, or materials science, leading to hallucinations or domain inaccuracies (F. Wang et al. 2024). In code generation, LLMs can inadvertently produce insecure or vulnerable scripts and even be exploited to create polymorphic malware, raising serious concerns in cybersecurity. Attribution studies show that although LLM-generated code mimics human style, its origin can be detected using inference-based authorship classification techniques with high accuracy Choi & Mohaisen 2025).

In response to these challenges, Small Language Models (SLMs) have emerged as efficient alternatives that offer a balance between performance, accessibility, and cost-effectiveness. SLMs provide low latency, lightweight fine-tuning, and enhanced adaptability for domain-specific applications, especially in settings with privacy constraints or limited hardware resources (F. Wang et al. 2024). Their ability to run on consumer-grade GPUs makes them suitable for localized tasks such as clinical feature extraction, legal attribute classification, and educational analytics (L. Wang et al. 2024). Quantized versions of SLMs, such as the INT4 version of Qwen-14B-Chat, have demonstrated high accuracy (97.28%) and even a 0% null ratio, eliminating unsupported predictions while preserving inference quality (L. Wang et al. 2024).

These results underscore the promise of SLMs in delivering reliable inference in privacy-sensitive or resource-limited environments. Several architectural innovations further optimize inference systems. Transformer-based language models form the backbone of modern inference, learning contextual semantics, bridging inferences, and elaborative reasoning through self-supervised objectives like masked language modeling (Kumar et al. 2024). Meta-learning and few-shot learning allow models to adapt

rapidly to new tasks from limited examples, enabling incremental learning without full retraining (Kumar et al. 2024). Further, non-autoregressive (NAT) approaches using CTCPMLM and connectionist temporal classification (CTC) loss dramatically improve inference speed, achieving up to 16.35× acceleration while preserving translation quality (Syu et al. 2024). Hybrid inference architectures combine lightweight named entity recognition (NER) models with LLMs, enabling scalable knowledge extraction from massive corpora of scientific literature, such as extracting over one million property records from polymer research articles (Gupta et al. 2024).

Inference-based systems have demonstrated compelling utility in real-world domains. In healthcare, modular LLM pipelines have been deployed to extract clinical features from 25,709 pregnancy cases using Qwen-14B-Chat and Baichuan2-13B-Chat, achieving high precision and low hallucination rates (L. Wang et al. 2024). In legal analytics, few-shot learning approaches enable the extraction of structured attributes from criminal case documents, significantly improving legal judgment and statute prediction tasks (Adhikary et al. 2024). In educational research, cognitive studies show that human inference depends on background knowledge, vocabulary depth, and strategic reading behaviors, highlighting parallels between human and machine inference. These findings suggest that robust inference, whether in humans or machines, relies not only on language processing but also on knowledge integration and reasoning strategies (Cain et al. 2024).

Concrete examples of Qur'anic inference. In this work, we use the term inference to denote the task of producing a concise conclusion sentence that is logically and semantically supported by two input verses (premises). For example, combining "*And sing His praises morning and evening*" (Q. 33:42) with "*And glorify Him in the night*" (Q. 52:49) yields the inferred statement "Praise God at all times," which summarizes the shared temporal instruction across both verses. Likewise, pairing "*And gave him abundant wealth*" (Q. 74:12) with "*The Thamud denied (the truth)*" (Q. 91:11) leads to the inference that denial can be driven by attachment to worldly wealth. These examples illustrate why inference matters: it can surface implicit, cross-verse conclusions that are not explicitly stated in either verse alone, supporting scalable exploration of thematic relationships in Qur'anic exegesis.

In conclusion, the landscape of inference from English text has evolved from surface-level pattern matching to deep semantic reasoning powered by advanced neural architectures. Large language models have transformed knowledge extraction, but limitations regarding cost, accuracy, domain

alignment, and privacy have spurred the rise of Small Language Models and hybrid architectures. These developments signal a future in which inference systems combining semantic reasoning, domain knowledge, and computational efficiency serve as foundational tools in science, medicine, law, and beyond.

2. Methodology

We trained a binary classifier using the Avicenna syllogistic reasoning dataset. Each instance contains two premises and a binary label indicating whether a syllogistic relation exists between the premises ("yes") or not ("no"). We used the dataset-provided "Syllogistic relation" field as ground truth; labels were not newly assigned by the authors. The labels in the Avicenna dataset were annotated by the dataset authors following explicit annotation guidelines; further details regarding the annotation protocol and quality are provided in Aghahadi and Talebpour (2022).

Dataset split: The Avicenna dataset contains 6,000 instances. We used the provided 4,800 instances as the training pool and 1,200 instances as the held-out test set (20%). From the 4,800 training instances, we set aside 10% (480 instances) as a validation set and used the remaining 4,320 instances for training. The validation set was used for model selection, while the test set was used only once for final reporting.

Model details: We fine-tuned a pretrained BERT model (bert-base-cased) for binary sequence classification (yes/no). Input pairs were formatted as: premise1 [SEP] premise2, tokenized using AutoTokenizer with a maximum sequence length of 512 and truncation enabled. We trained for 10 epochs with a learning rate of 2e-5, batch size 32, and weight decay 0.01, selecting the best checkpoint based on validation performance.

Verse pair construction: We extracted 6,236 verses from the English translation file (en.ahmedali). We enumerated all unique unordered verse pairs by selecting indices $i<j$, resulting in 19,440,730 candidate pairs. Each pair was then passed to the trained classifier to obtain a predicted label and confidence score.

Filtering weak pairs: We first discarded all pairs predicted as "no" by the classifier, retaining 205,013 candidate "yes" pairs. We then applied an automatic confidence threshold (score ≥ 0.80) to remove weak or low-confidence pairs, yielding 9,820 high-confidence pairs.

Human evaluation feasibility: A human expert did not evaluate all 205,013 candidate pairs. Expert review was applied only after automatic

filtering on the high-confidence subset (and/or on a small stratified sample for quality control), making the manual step practically feasible.

Figure 1 illustrates the research framework designed to analyze and generate inferential statements from the Qur'anic text in English. This architecture comprises several interconnected components.

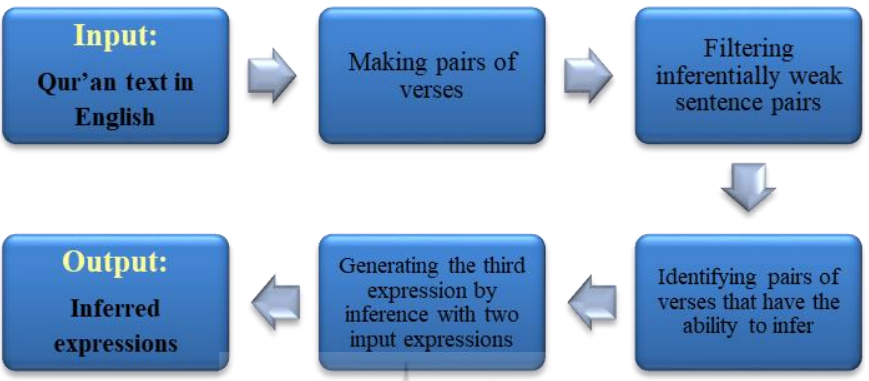

Figure 1. The overall architecture of the proposed framework

2.1. English Qur'an Text as Input and Model Construction

The technical process of this study begins with the use of the English-translated Qur'anic text as the primary input, which assumes a pivotal role in subsequent analysis and processing. In view of the paucity of labeled datasets in the domain of Qur'anic textual inference, the Avicenna dataset, a benchmark in natural language generation and inference, was utilized for model construction (Aghahadi & Talebpour 2022).

2.2. Construction of Qur'anic Ordered Pairs

In this phase, transformers were employed for model building and text classification. The architecture is predicated on the attention mechanism, thereby enabling the model to accurately capture relationships between words and sentences (Bostrom et al. 2022). Subsequently, the Bidirectional Encoder Representations from Transformers (BERT) model was implemented, capitalizing on its pretraining on extensive unlabeled text corpora to attain robust performance (Devlin et al. 2019). Following the deployment of the model, all Qur'anic verses were processed in pairs, with the model predicting potential relationships for each verse pair.

2.3. Filtering Weak Inference Pairs

In order to enhance the quality and relevance of the data, model-based filtering criteria were applied to remove verse pairs labeled as "no" due to weak inferential connections. This step ensures that only meaningful pairs (labeled "yes") proceed to subsequent stages. As not all constructed pairs demonstrate robust logical or semantic coherence, this phase necessitates model-driven evaluation to eliminate weaker pairs. In essence, the model learns to accurately classify strong (related) and weak (unrelated) pairs.

2.4. Identifying Strong Inference Pairs

After applying the trained BERT-based classifier to all unique unordered verse pairs, a total of 205,013 verse pairs were predicted as exhibiting a syllogistic relation (label "yes"). These pairs represent candidate inference relationships between Qur'anic verses generated automatically by the model. To remove weak or unreliable predictions, an automatic confidence-based filtering step was applied. Only verse pairs with a predicted probability score greater than or equal to 0.80 were retained. This threshold reduced the number of candidate pairs from 205,013 to 9,820 high-confidence inference pairs. Human expert review was not conducted on the entire set of 205,013 candidate pairs due to feasibility constraints. Instead, expert evaluation was applied only after the automatic filtering stage, either on the resulting high-confidence subset or on a small stratified sample of it, with the goal of quality control and validation rather than exhaustive annotation. This process ensures both scalability and reliability in identifying strong inference pairs.

2.5. Generating Inferential Statements

In the final step, a generative inference model was employed, utilizing LLMs to produce inferences. These models analyze the inferential relationship between input verse pairs and generate logically coherent conclusions. This capability is derived from the proficiency of LLMs in semantic understanding and context-aware content generation. The Llama3 model family, which was introduced by Meta Research, signifies a novel generation of foundational AI models. These models are available in three parameter sizes (8B, 70B, and 405B) and have been shown to optimize computational resource utilization (Dubey et al. 2024). The system's capabilities are characterized by its ability to demonstrate competitive performance in comparison to models that represent the pinnacle of technological advancement, as well as its capacity to exhibit behavior that

is aligned with human tendencies. The present study implemented the Meta Llama 3.1 8B variant. The ensuing sections are dedicated to the presentation and discussion of the results and analyses.

3. Results

This study is among the first to apply advanced NLP algorithms and LLMs to generate intelligent inferences from Qur'anic verses. The findings indicate that these models can effectively identify semantic relationships between Qur'anic verses and produce logically coherent inferential statements. They further show that pre-trained language models function as effective instruments for the analysis of religious texts, particularly the Holy Qur'an. This approach has been demonstrated to reduce the time and costs associated with textual analysis while enabling the discovery of novel and deeper interconnections between verses.

In the contemporary digital research environment, interdisciplinary studies and diverse methodological strategies have played a pivotal role in propelling scientific inquiry forward. The integration of knowledge and methodologies from multiple domains has been demonstrated to facilitate the emergence of new insights, underscoring the transformative potential of cross-disciplinary collaboration in computational linguistics and theological studies.

3.1. Performance Evaluation

Table 1 presents the performance metrics of the proposed model.

Table 1. Performance metrics of the proposed model

Number of Verses	Number of Ordered Pairs	Verses Suitable for Inference	Accuracy
6,236	19,440,730	9,820	84%

As illustrated in Table 2, the model's outputs are presented in English.

Table 2. Sample outputs in English

No.	First Verse	Second Verse	Inferred
1	And glorify Him in the night, and when the stars begin to wane (Q. 52:49)	And sing His praises morning and evening (Q. 33:42)	Praise God at all times
2	The Thamud denied (the truth) in their perverseness (Q. 91:11)	And gave him abundant wealth (Q. 74:12)	Their denial was due to their love for and being enriched by worldly wealth.
3	Why then do they turn away from the admonition (Q. 74:49)	But those who try to subvert Our signs will be given over to punishment (Q. 34:38)	Those who deny or go against God's messages will face punishment.
4	When souls are reunited (with the bodies) (Q. 81:7)	Gather all the wicked together and their comrades, and those they had worshipped (Q. 37:22)	Judgment or a reckoning for the wicked will occur when souls are reunited with their bodies.
5	Whoever turns away from it will surely carry a burden on the Day of Judgement (Q. 20:100)	Surely the devotees will drink cups flavoured with palm blossoms (Q. 76:5)	Turning away from Allah will result in a burden on the Day of Judgement, while devotion to Him brings rewards and blessings.

3.2. Analysis of Inferential Sentences in the Output

In this section of the study, brief explanations are provided regarding the inferential outputs.

1. The first verse details the glorification of the Lord during nocturnal hours, while the subsequent verse underscores the glorification of God in the morning and evening. The proposed language model, by integrating these two verses with its temporal awareness, has deduced that the worship and praise of the Lord should be unceasing (in all circumstances).
2. Based on the Qur'anic verses, the people of Thamud turned to denial despite the abundant material blessings and vast wealth bestowed upon them. This rejection of the truth stemmed from their deep attachment to material possessions and worldly allurements. The present analysis indicates that accepting the truth and following divine commands requires a detachment from worldly attachments and an orientation toward the transcendent truth.
3. The first verse alludes to a collective renunciation of truth, while the subsequent verse delves into the repercussions of those who engage in the act of falsehood. In the interpretation of these verses, the model identifies "turning away" as a metaphor for rejection and explicitly describes the severe punishment awaiting such individuals.
4. The first verse alludes to the reanimation of the soul within the physical body, while the subsequent verse portrays the congregation of all transgressors and the deities they venerate. Despite the absence of

explicit mention of the Day of Judgment in these verses, the model, based on the concept of the soul's return to the body, has inferred a day of reckoning for the sinners.

5. The initial verse addresses those who disregard God and the Day of Judgment, burdening themselves with the weight of sins through persistent wrongdoing. In contrast, the second verse delineates the righteous, characterized as individuals who, through their pious deeds, find themselves immersed in divine and heavenly blessings. A thorough examination of these verses reveals a definitive conclusion: those who transgress will face dire consequences, while those who adhere to virtuous principles are poised to receive substantial rewards.

4. Conclusion

This study evaluated a computational model for Qur'anic analysis using a large-scale comparative framework. The model processed a corpus of 6,236 verses, from which 9,820 structured inferences were systematically derived. The evaluation was conducted through an automated process that generated 19,440,730 comparative data points. The results demonstrate that the model achieved a final inference accuracy of 84%. This outcome substantiates the efficacy of the proposed methodology, which integrates domain-specific knowledge from Qur'anic studies with advanced natural language processing techniques. The model's performance highlights the utility of LLMs in enhancing analytical capabilities for complex textual problems. By leveraging vast pre-trained knowledge, LLMs can significantly contribute to the precision and quality of scholarly inferences in the humanities. This research signifies a substantive advancement at the intersection of Islamic studies and artificial intelligence. It underscores the considerable potential of computational methods to facilitate rigorous, scalable analysis in religious and humanistic disciplines. The findings provide a robust foundation for future research, encouraging more extensive and applied explorations of digital tools in specialized scholarly domains.

Acknowledgements

This paper has been extracted from the thesis prepared by Samira Abdi under the supervision of Dr. Alireza Talebpour, Dr. Ahmad Mahmoudi-Aznaveh and Dr. Mostafa Moradi in fulfilment of the requirements for the degree of Master of Computational Qur'an Mining. The authors extend their gratitude to the Interdisciplinary Qur'anic Studies Research Institute, Shahid Beheshti University, and the Research Deputy for supporting the research.

Declarations

Funding: No funding was received for conducting this study.
Conflict of Interest: The authors declare no competing interests.

References

Adhikary, S., Sen, P., Roy, D., & Ghosh, K. (2024). A case study for automated attribute extraction from legal documents using large language models. *Artificial Intelligence and Law*. https://doi.org/10.1007/s10506-024-09425-7

Aghahadi, Z., & Talebpour, A. (2022). Avicenna: A challenge dataset for natural language generation toward commonsense syllogistic reasoning. *Journal of Applied Non-Classical Logics*, 32(1), 55–71. https://doi.org/10.1080/11663081.2022.2041352

Ali, Ahmed (2001). *Al-Qurʾān: A contemporary translation*. Princeton University Press. https://tanzil.net/trans/

Bostrom, K., Sprague, Z., Chaudhuri, S., & Durrett, G. (2022). Natural language deduction through search over statement compositions. *arXiv:2201.06028*. https://doi.org/10.48550/arXiv.2201.06028

Cain, P. K., et al. (2024). The influence of reader and text characteristics on sixth graders' inference making. *Journal of Research in Reading*. 48(1), 24-45 https://doi.org/10.1111/1467-9817.12474

Choi, S., & Mohaisen, D. A. (2025). Attributing ChatGPT-generated source codes. *IEEE Transactions on Dependable and Secure Computing*, 22(4), 3602-3615. https://doi.org/10.1109/TDSC.2025.3535218

Dagdelen, J., Dunn, A., Lee, S., Walker, N., Rosen, A. S., Ceder, G., & Persson, K. A. (2024). Structured information extraction from scientific text with large language models. *Nature Communications*, 15, 1418. https://doi.org/10.1038/s41467-024-45563-x

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). 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* (4171–4186). Minneapolis, Minnesota. Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Dubey, A., et al. (2024). The Llama 3 herd of models. *arXiv:2407.21783*. https://doi.org/10.48550/arXiv.2407.21783

Gu, B., Shao, V., Liao, Z., Carducci, V., Brufau, S. R., Yang, J., & Desai, R. J. (2025). Scalable information extraction from free text electronic health records using large language models. *BMC Medical Research Methodology*, *25*. https://doi.org/10.1186/s12874-025-02470-z

Gupta, S., Mahmood, A.-U., Shetty, P., Adeboye, A., & Ramprasad, R. (2024). Data extraction from polymer literature using large language models. *Communications Materials*, *5*, 269. https://doi.org/10.1038/s43246-024-00708-9

Kumar, S., Sharma, A., Shokeen, V., Azar, A. T., Amin, S. U., & Khan, Z. I. (2024). Meta-learning for real-world class incremental learning: A transformer-based

approach. *Scientific Reports, 14*, 23092. https://doi.org/10.1038/s41598-024-71125-8

Syu, S., Xie, J., & Lee, H. (2024). Improving non-autoregressive translation quality with pretrained language model, embedding distillation and upsampling strategy for CTC. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32, 4121-4133. https://doi.org/10.1109/TASLP.2024.3451977

Wang, F., Zhang, Z., Zhang, X., Wu, Z., Mo, T., Lu, Q., Wang, S. et al. (2024). A comprehensive survey of small language models in the era of large language models: Techniques, enhancements, applications, collaboration with LLMs, and trustworthiness. *ACM Transactions on Intelligent Systems and Technology*. https://doi.org/10.48550/arXiv.2411.03350

Wang, L., Ma, Y., Bi, W., Lv, H., & Li, Y. (2024). An entity extraction pipeline for medical text records using large language models: Analytical study. *Journal of Medical Internet Research, 26*, e54580. https://doi.org/10.2196/54580