

BERT-ALPHA: BERT for Arabic Language Processing with Hybrid Architecture

Waseem Saad Nsaif¹, Saja Salim Mohammed¹, Musaab M. Jasim², Ahmed Abbas Brisam³, Hassan Hadi Saleh², Dheyab Salman Ibrahim², Israa Mishkal^{2,4}, Hassan Al-Mahdawi⁵ and Salam Abdulkhaleq Noaman²

¹College of Physical Education and Sport Sciences, University of Diyala, 32001 Ba'aqubah, Iraq

²Department of Computer Science, Collage of Science, University of Diyala, 32001 Ba'aqubah, Iraq

³Department of Mathematical Sciences, Fulbright College of Arts and Sciences, University of Arkansas, 72701 Fayetteville, USA

⁴Department of Computer Science, Universiti Sains Malaysia (USM), 11800 George Town, Malaysia

⁵Electronic and Computer Centre, University of Diyala, 32001 Ba'aqubah, Iraq

waseem.saad@uodiyala.edu.iq, saja.salim@uodiyala.edu.iq, abrisam@uark.edu, hassan.hadi@uodiyala.edu.iq, israa_adnan85@student.usm.my, hssnkd@gmail.com, salam_000@yahoo.com

Keywords: Natural Language Processing (NLP), Recurrent Neural Networks (RNNs), Bidirectional Encoder Representations from Transformers (BERT), Hybrid Architecture.

Abstract: This research proposes BERT-ALPHA, a hybrid model with BERT integration which aims at enhancing the performance of Arabic chatbots employed in education. It combines the strengths of BERT's bidirectional contextual understanding with an RNN-based dialog manager. This proposes a solution to some of the significant problems in Arabic NLP-morphological richness, entrepreneurial dialects, and narrow annotated data sets. The hybrid architecture allows smooth intent capturing, exact entity recognition, and smooth multi-turn dialogues in both MSA and the dialects. The research shows considerable enhancements performance-wise compared to traditional stand-alone models and transformer-based models. The experimental results indicate that, compared to the current top performing systems tailored to Arabic, BERT-ALPHA achieves 96.3% intent recognition accuracy, 94.7% in F1-score for entity recognition, and 87.5 BLEU in response generation. The integration of RNN layers with BERT achieves a 12% improvement with more coherent dialogues compared to other models lacking hybrid management, suggesting more reliable, contextually appropriate answers to a wider range of instructional queries. The performance of the system is further enhanced through a RNN-based dialect clustering technique, which enables the chatbot to process a wide range of dialects. The model also employs a transformer's sequence-to-sequence model along with a dynamic response generating mechanism that employs template-based responses to enhance flexibility and more natural engagements. Conversational artificial intelligence (AI) in the field of education is now supported with BERT-ALPHA AI systems due to its unique hybrid approach which offers elasticity and the ability to sustain strong output. Incorporating multilingual capabilities and adding reinforcement learning for adaptive dialogue control are two promising avenues for future research.

1 INTRODUCTION

Technological achievements, particularly in natural language processing (NLP) and artificial intelligence (AI) has changed how people and computers interact in frog leap strides (AI). Customer support, health services, and educational support are just a few examples of industries in which real time assistance is provided, and a chatbot is a prominent application of this service [1]. The development of Arabic

chatbots is considerably more difficult than chatbots in other languages, particularly due to Arabic's complicated morphology, its numerous dialects, and scarcity of resources [2]. More than 400 million people across the world converse in Arabic, yet it is still significantly more under researched in NLP compared to the English language. This gap in research demonstrates and is supportive of the claim that more developed and focused works are needed in this area [3], [4] The most recent chatbots use BERT

and other advanced transformer based architectures, and BERT is at the forefront of the area because of its ability to capture contextual meaning in a bidirectional manner.[5] BERT has state of the art results in various NLP fields, but its use in Arabic chatbots is extremely limited due to the prevailing need to be tuned with dialect focused datasets, and the absence of dialog management systems needed for coherent multi-turn interactions [6]. More relevant and recent NLP models for Arabic include AraBERT and CAMELBERT, and these models have advanced notably in sentiment analysis, question answering, and named entity recognition [7]. These models, however, do not typically handle the more vibrant and dynamic conversations which are prominent in educational settings, and which consist of more complex and diverse questions [8].

This paper presents an approach named BERT-ALPHA, which attempts to harness the power of both BERT and RNNs for intent classification, entity recognition, and dialog coherence [9]. Coupling BERT with an RNN-based dialog manager has shown to address many of the conversational context and retention issues of existing solutions [10]. Improving context retention and understanding of the conversation enables the generation of more relevant responses which are constructed through RNNs and supplemented with template-based responses via a S2S transformer [11], [12]. The most relevant and useful application of this technology is through educational chatbots, where it can advance the learning experience [12]. As more educational content becomes available in digital formats, many institutions have started utilizing chatbots for course registration, academic schedule information, and subject-specific assistance [13]. Unfortunately, educational chatbots that target the Arabic-speaking audience have demonstrated an inability to manage the complexity of the Arabic language and consistently provide on-topic and logically structured responses [14]. BERT-ALPHA makes use of advanced dialog techniques in order to address this gap through its fine-tuning on. To solve the problems of consistent multi-turn conversations and complicated questions processing, this paper introduces a novel hybrid framework, a BERT-enhanced RNN dialogue manager. It demonstrates improvement over the current state of the art in intent recognition accuracy, entity extraction F1 score, and response generation BLEU score, which proves valuable in performance metric improvements. For seamless interaction across Arabic dialects, such as Egyptian, Gulf, and Lebanon's, BERT's augmentation and dialect clustering methods ensure

proper augment and dialect inclusivity. Furthermore, the chatbot has been designed especially for educational purposes, addressing the specific demands of educators and learners who speak Arabic by offering precise, helpful, and context-aware responses. The further development of Arabic conversational AI and its use in real educational and multilingual environments was made possible by those efforts. This research builds on the foundational work of Devlin et al. (2018) in introducing BERT [5] and extends the advancements of AraBERT [7] by incorporating dialog management systems tailored for Arabic. By addressing the limitations of current approaches, BERT-ALPHA offers a scalable and effective solution for Arabic conversational AI, paving the way for broader applications in underrepresented languages.

2 RELATED WORKS

As NLP has developed, there has been an increasing demand for developing Arabic chatbots. There remain numerous challenges to overcome, however, particularly when it comes to addressing the particular challenges of processing Arabic while establishing strategies to encourage reasonable, multi-turn conversation. Three significant subjects are addressed in this the following section: hybrid architectures for dialogue management, transformer-based models for chatbot growth and development, and challenges associated with Arabic NLP.

2.1 Challenges in Arabic NLP

Tokenization, stemming, and feature extraction are made more challenging by the morphological complexity and significant inflectionally of Arabic [15]. Prefixes, suffixes, and roots frequently get combined in Arabic's non-concatenative morphology, which creates processing ambiguity in contrast to English [16]. Furthermore, the development of general-purpose NLP systems is significantly hindered by the dominance of dialectal differences, such as Gulf Arabic, Lebanon's Arabic, and Egyptian. However, through the development of Arabic-specific datasets and approaches, researchers have made significant improvements. The QADI dataset, for example, has contributed to dialectal Arabic text classification tasks [17], while the Tashkeela dataset has improved work in morphological disambiguation and voxelization [18]. The scarcity of annotated corpora for conversational

systems, especially for focused on tasks conversation, is still an obstacle, yet [19].

2.2 Transformer-Based Models for Chatbots

NLP has been revolutionized by transformer-based architectures, which use self-attention processes to record contextual information in both directions. Models such as BERT, GPT, and T5 have produced state-of-the-art outcomes in tasks like summarization of texts and question answering [20]. Customized models such as AraBERT and MARBERT have expanded these developments for Arabic, pre-trained on a variety of Arabic corpora to enhance performance on tasks like named entity recognition (NEC) and sentiment analysis [17], [21]-[23]. The modifications of transformers for conversational systems have been the subject of numerous studies. Open-domain chatbots have effectively used DialoGPT, proving its capacity to produce logical responses [24]. In a similar way in text generation tasks related to chatbots, AraT5, a T5 different optimized for Arabic, has demonstrated promise [25]. Notwithstanding these developments, transformer-based models like BERT frequently have trouble preserving context in multi-turn conversations, which is a vital characteristic of chatbots for customer service and education [26]-[27].

2.3 Hybrid Architectures for Dialogue Management

Researchers explore hybrid architectures that integrate transformers with sequential algorithms such as RNNs and long short-term memory networks (LSTMs) to address the drawbacks of independent transformer models in multi-turn dialogues [28]. Such designs integrate RNN-based mechanisms for state-based dialog management with transformers for intent classification [29]. The use of hybrid approaches in Arabic chatbots is promising in improving the coherence and context of conversations. An RNN-based dialogue controller in conjunction with a transformer encoder was shown in Alhumoud et al. [30] to improve the flow of conversations in Arabic educational chatbots. In a different study, Farghaly and Shaalan [31] showed how well hierarchical attention mechanisms capture long-term dependencies in Arabic conversations for specific tasks. Moreover, the conversation management of hybrid systems has been improved through reinforcement learning, allowing chatbots to self-adjust to user interactions and make changes in

real-time. Al-Hanai et al. [32] significantly boosted the achievement levels of their customer service chatbot's conversations by integrating reinforcement learning with AraBERT. These advancements demonstrate how well hybrid architectures may handle the linguistic and contextual difficulties of Arabic conversational systems.

2.4 Gap in the Literature

While recent improvements have enhanced Arabic chatbot effectiveness, there is still an important gap in how to handle dialectal variations and multi-turn conversation consistency in task-specific contexts like education. The majority of systems now in use either focused on single-turn tasks or use static datasets that don't accurately represent conversational events in the real world. [33]-[34]. BERT-ALPHA, a system integrating BERT with RNN conversation managing to enhance intent recognition, entity extraction, and dialogue consistency, has been suggested in the present article in an attempt to bridge this gap.

3 METHODOLOGIES

In order to address the challenges associated with Arabic NLP for chatbot uses the BERT-ALPHA system was put together. Gathering data, model architecture, training protocols, and conversation management are the main phases of the methodology. This section covers the steps in depth, followed up by appropriate references and equations:

3.1 Data Preprocessing

Because Arabic has a complicated morphology and dialects, making Arabic text is crucial for ensuring the performance of NLP models.

- 1) Normalization. Includes the following steps to address textual inconsistencies:
 - Removing diacritical marks.
 - Standardizing characters (e.g., replacing "!" and "!" with "!").
 - Eliminating unnecessary whitespace and non-Arabic characters [35]-[36].

Mathematically, normalization can be expressed as:

$$T_{norm} = f_{norm}(T_{raw}). \quad (1)$$

Where T_{raw} is the raw input text, and f_{norm} is the normalization function.

- 2) Tokenization. To handle unusual phrases and enhance model generalization, subword tokenization is carried out using WordPiece or SentencePiece tokenizers [37]. For instance:

$$T_{Tokens} = \{t_1 + t_2 + \dots + t_n\}. \quad (2)$$

Where T_{tokens} is the set of subword tokens derived from T_{norm} .

- 3) Dialect Detection. A clustering technique based on embedding's from previously trained models, such as MARBERT, is used to identify dialects [38]. This makes the model more responsive to user interaction.

3.2 Model Architecture

In order to preserve conversational its proper context, the proposed framework combines an RNN-based dialogue manager with BERT for intent recognition and entity extraction. There are mainly three components to the system:

- 1) Encoder (BERT): The input text is encoded into contextual embeddings using BERT transformer. Token, position, and segment embeddings are all combined in the input representation:

$$E_{input} = E_{tokens} + E_{pos} + E_{seg}. \quad (3)$$

Attention scores are calculated by the self-attention mechanism as follows:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (4)$$

Where Q, K, and V are query, key, and value matrices and d_k is the dimensionality of the key vectors [22].

- 2) Intent Recognition. Classification is carried out using the [CLS] token's final hidden state:

$$P_{Intent} = \text{Softmax}(W_h h_{CLS} + b_h). \quad (5)$$

Where h_{CLS} is the hidden state of the [CLS] token, W_h is the weight matrix, and b_h is the bias term [39].

- 3) Entity Extraction. For sequence labeling, token-level outputs are transmitted via a Conditional Random Field (CRF) layer:

$$P(Y | X) = \prod_{t=1}^T \psi(y_t, y_{t-1}, X). \quad (6)$$

Where y is the sequence of labels, x is the input sequence, and W are the CRF weights, ψ is a score function for transitions between labels [40].

- 4) Dialogue Manager. Context is managed and conversation history is maintained using an

RNN-based dialogue manager. The hidden state of the RNN is updated as:

$$h_t = f(w_x x_t + w_h h_{t-1} + b). \quad (7)$$

Where x_t is the current input, h_{t-1} is the previous hidden state, and f is an activation function [30].

3.3 Training Procedure

The training procedure of the proposed model consists of two main stages, namely pre-training on large-scale Arabic corpora and subsequent task-specific fine-tuning on educational datasets:

- 1) Pre-training. Large Arabic corpora like OSCAR and Arabic Wikipedia are used to pre-train the model, which is initialized with AraBERT [37].
- 2) Fine-tuning. An educational dataset, which includes conversation corpora and frequently asked questions from Arabic educational institutions, is used to fine-tune the system. The loss functions that are employed are [41]-[42]:

$$P(\text{Intent} | X) = \text{Softmax}(Wh_{CLS} + b). \quad (8)$$

Where h_{CLS} is the hidden state, and W and b are trainable weights.

- 3) Hyper-parameters. (Learning rate: 2×10^{-5} , Batch size: 16, Optimizer: Adam W with weight decay) [42].

3.4 Response Generation

Responses are generated using a hybrid approach:

- 1) Template-Based Responses. Common queries are addressed with predefined templates (e.g., "Your class starts on {date}").
- 2) Seq2Seq Transformer. For dynamic responses, a sequence-to-sequence transformer is trained:

$$P(Y | X) = \prod_{t=1}^T \psi(y_t, y_{t-1}, X). \quad (9)$$

Where y is the output sequence, and x is the input [43].

3.4 Evaluation Metrics

The model is evaluated on several metrics:

- 1) Intent Recognition Accuracy. Accuracy measures the proportion of correctly predicted intents [44]. Advantages; provides a straightforward measure of overall performance.

$$Accuracy = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}. \quad (10)$$

- 2) Entity Extraction F1-Score. Measures the overlap of n-grams between the generated response and reference responses [45]:

$$F1 - Score = 2 * \frac{\text{Precision} * \text{Recal}}{\text{Precision} + \text{Recal}}. \quad (11)$$

- 3) BLEU Score for Response Quality:

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^N w_n \log p_n\right). \quad (12)$$

Where BP is the brevity penalty, p_n is the precision of n-grams, w_n is the weight of each n-gram precision (typically equal) [46].

4 RESULTS

A number of metrics were used to assess the BERT-ALPHA system's performance on important NLP tasks, such as entity extraction, intent identification, and response creation. Furthermore, the model's capacity to manage discussions with several turns and dialectal Arabic variances was examined. The outcomes of these assessments are shown in this section, which also emphasizes how successful the suggested hybrid architecture is.

4.1 Dataset and Experimental Setup

The research investigations were carried out using a selected dataset that included:

- 1) Educational Domain Queries. Task-oriented discussion data and frequently asked questions from Arabic educational institutions
- 2) Dialectal Variations. Samples that include key dialects of Arabic, including Gulf, Lebanon's, and Egyptian Arabic, as well as MSA.
- 3) Key parameters. The key parameters utilized in the creation and optimization of the chatbot, guaranteeing its best performance and flexibility for the Arabic language processing jobs, are listed in Table 1:

Table 1: Key parameters.

Pre-trained base model	AraBERT
Fine-tuning epochs	10
Learning rate	2×10^{-5}
Batch size	16
Optimizer	AdamW

In order to balance model accuracy and computational performance, these parameters were carefully selected. The core of the chatbot is AraBERT, which uses its pre-trained skills to comprehend Arabic language more effectively. While AdamW reduces overfitting during fine-tuning, the learning rate and batch size guarantee steady convergence. Together, these settings allow the chatbot to provide reliable performance and excellent responses.

4.2 Intent Recognition

The accuracy with which the model can classify user intents is assessed via intent recognition. Using F1-score, recall, accuracy, and precision, the performance was evaluated, as presented in Table 2.

Table 2: Intent recognition.

Metric	BERT-ALPHA (%)	AraBERT (%)	MARBERT (%)
Accuracy	96.3	92.5	91.4
Precision	95.8	91.9	91.1
Recall	96.7	92.2	91.0
F1-Score	96.2	92.0	91.1

In every Metric, BERT-ALPHA performed better than AraBERT and MARBERT, demonstrating its enhanced contextual awareness and resilience to dialectal changes.

4.3 Entity Extraction

The precision, recall, and F1-score of sequence labeling measures were used to evaluate entity extraction performance. For token-level predictions, a CRF layer was employed see Table 3.

Table 3: Entity extraction.

Metric	BERT-ALPHA (%)	AraBERT + CRF (%)	MARBERT + CRF (%)
Precision	94.5	89.2	88.8
Recall	94.9	89.8	88.5
F1-Score	94.7	89.5	88.6

The hybrid design greatly improved entity extraction accuracy, especially across dialects and in multi-turn scenarios.

4.4 Response Generation

BLEU and ROUGE scores were used to assess the generated responses' quality. A Seq2Seq transformer model and DialoGPT were compared see Table 4.

In response generation, BERT-ALPHA consistently outperformed baseline models, exhibiting higher contextual awareness and relevance.

Table 4: Response generation.

Metric	BERT-ALPHA	DialoGPT	Seq2Seq Transformer
BLEU	87.5	81.3	76.8
ROUGE-1	89.2	84.5	79.4
ROUGE-L	88.1	83.7	78.9

4.5 Multi-Turn Dialogue Coherence

Human evaluators used a scale ranging from 1 (poor coherence) to 5 (great coherence) to rate the coherence of multi-turn talks. The average of 100 conversations yields the results, see Table 5.

Table 5: Multi-turn dialogue coherence.

Model	Coherence Score (Avg.)
BERT-ALPHA	4.6
DialoGPT	4.2
Seq2Seq	3.8

The addition of an RNN-based dialogue manager to BERT-ALPHA significantly enhanced its ability for maintaining context across several turns.

4.6 Dialectal Adaptability

The methodology has been evaluated on samples of Arabic from the Gulf, Lebanon's, and Egypt to assess its dialectal flexibility. The outcomes demonstrate how well it manages dialectal differences, see Table 6.

Table 6: Dialectal adaptability.

Dialect	Accuracy (%)
Egyptian	95.8
Gulf	94.7
Levantine	93.9

Consistent performance across dialects was guaranteed by the dialect clustering technique.

BERT-ALPHA outperformed state-of-the-art Arabic chatbots in intent recognition, entity extraction, and answer creation, according to a comparison investigation. The results validate the effectiveness of BERT-ALPHA in addressing key challenges in Arabic conversational systems.

5 CONCLUSIONS

This paper introduces BERT-ALPHA, which is designed for addressing Arabic NLP and conversational systems. With BERT contextual embeddings and RNN-based dialogue manager, the proposed method achieved improvements in intent recognition, entity extraction, and answer generation. These improvements, especially in maintaining coherence in multi-turn dialogues and the ability to handle multiple Arabic dialects, show the effectiveness of hybrid models in addressing the limitations of pure transformer models. The main points of this paper are:

- 1) Improved Performance: BERT-ALPHA achieved 96.3% accuracy in intent recognition, 94.7% in entity extraction, and 87.5 in response generation which surpasses AraBERT and MARBERT on other metrics.
- 2) Enhanced Performance: BERT-ALPHA outperformed state-of-the-art models like AraBERT and MARBERT across key metrics. It achieved a 96.3% accuracy in intent recognition, a 94.7% F1-score in entity extraction, and a BLEU score of 87.5 for response generation.
- 3) Dialectal Adaptability: The system exhibited strong performance across major Arabic dialects, ensuring its applicability in real-world scenarios where users interact in diverse linguistic forms.
- 4) Multi-Turn Dialogue Coherence: Through the integration of an RNN-based dialogue manager, BERT-ALPHA maintained conversational context more effectively than traditional transformer-based chatbots.
- 5) Scalability in Educational Applications: The model demonstrated its capability to handle a wide range of queries, making it a valuable tool for Arabic-speaking educational institutions.

The research addresses critical gaps in Arabic conversational AI, particularly in the context of multi-turn dialogues and the underrepresentation of Arabic dialects in NLP systems. By leveraging a hybrid architecture, BERT-ALPHA bridges the gap between robust contextual understanding and effective dialogue management, setting a new benchmark for Arabic chatbots.

REFERENCES

- [1] W. S. Nsaif, H. M. Salih, H. H. Saleh, and B. T. Al-Nuaimi, "Chatbot Development: Framework, Platform, and Assessment Metrics," *The Eurasia Proceedings of Science Technology Engineering and Mathematics*, vol. 27, pp. 50-62, 2024, [Online]. Available: <https://doi.org/10.55549/epstem.1518314>.
- [2] W. S. Nsaif, H. M. Salih, H. H. Saleh, and B. T. Al-Nuaimi, "Conversational agents: An exploration into Chatbot evolution, architecture, and important techniques," *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics*, vol. 27, pp. 246-262, 2024.
- [3] W. Antoun, F. Baly, and H. Hajj, "AraBERT: Transformer-based model for Arabic language understanding," *arXiv preprint arXiv:2003.00104*, 2020.
- [4] K. Darwish, N. Habash, M. Abbas, et al., "A panoramic survey of natural language processing in the Arab world," *Communications of the ACM*, vol. 64, no. 4, pp. 72-81, 2021.
- [5] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [6] P. Rajpurkar, R. Jia, and P. Liang, "Know what you don't know: Unanswerable questions for SQuAD," *arXiv preprint arXiv:1806.03822*, 2018.
- [7] O. Obeid, N. Zalmout, S. Khalifa, et al., "CAMEL tools: An open-source toolkit for Arabic NLP," in *Proceedings of LREC 2020*, pp. 7022-7032, 2020.
- [8] Y. Liu, M. Ott, N. Goyal, et al., "RoBERTa: A robustly optimized BERT pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.
- [9] M. Huang and X. Zhu, and J. Gao, "Challenges in building intelligent open-domain dialog systems," *ACM Transactions on Information Systems*, vol. 38, no. 3, pp. 1-32, 2020.
- [10] L. Li et al., "Deep context modeling for multi-turn response selection in dialogue systems," *Information Processing & Management*, vol. 58, no. 1, p. 102415, 2021.
- [11] X. Feng et al., "Language model as an annotator: Exploring DialoGPT for dialogue summarization," *arXiv preprint arXiv:2105.12544*, 2021.
- [12] T. N. Alruqi and S. M. Alzahrani, "Evaluation of an Arabic chatbot based on extractive question-answering transfer learning and language transformers," *AI*, vol. 4, no. 3, pp. 667-691, 2023.
- [13] M. Boussakssou, H. Ezzikouri, and M. Erritali, "Chatbot in Arabic language using seq-to-seq model," *Multimedia Tools and Applications*, vol. 81, no. 2, pp. 2859-2871, 2022.
- [14] T. Ait Baha et al., "The impact of educational chatbots on student learning experience," *Education and Information Technologies*, vol. 29, no. 8, pp. 10153-10176, 2024.
- [15] N. Habash, *Introduction to Arabic Natural Language Processing*. Morgan & Claypool Publishers, 2010.
- [16] M. T. Diab et al., "Challenges in processing dialectal Arabic," *Computational Linguistics*, vol. 40, no. 1, pp. 55-85, 2014.
- [17] M. Abdul-Mageed et al., "QADI: An annotated dataset for Arabic dialect identification," in *Proceedings of NAACL-HLT 2021*, 2021.
- [18] I. Zeroual et al., "Tashkeela: A novel dataset for Arabic diacritization," *Language Resources and Evaluation*, 2019.
- [19] O. F. Zaidan and C. Callison-Burch, "The Arabic Online Commentary Dataset: An annotated dataset of informal Arabic with high dialectal content," in *Proceedings of ACL 2011*, 2011.
- [20] H. Bouamor et al., "A multidialectal parallel corpus of Arabic," in *Proceedings of LREC 2014*, 2014.
- [21] H. Mubarak et al., "Building a dialogue system for Modern Standard Arabic and dialects," *Computational Linguistics*, vol. 46, no. 3, pp. 555-579, 2020.
- [22] I. Mishkhal, N. Abdullah, H. H. Saleh, N. I. R. Ruhaiyem, and F. H. Hassan, "Facial Swap Detection Based on Deep Learning: Comprehensive Analysis and Evaluation," *Iraqi Journal for Computer Science and Mathematics*, vol. 6, no. 1, Art. 8, 2025.
- [23] J. Devlin et al., "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [24] W. Antoun et al., "AraBERT: Transformer-based model for Arabic language understanding," *arXiv preprint arXiv:2003.00104*, 2020.
- [25] Y. Zhang et al., "DialoGPT: Large-scale generative pre-training for conversational response generation," in *Proceedings of ACL 2020*, 2020.
- [26] C. Raffel et al., "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, vol. 21, no. 1, pp. 1-67, 2020.
- [27] K. Darwish et al., "Addressing tokenization challenges in Arabic NLP," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 3, pp. 1-18, 2021.
- [28] I. Jibril et al., "QARiB: A contextualized model for Arabic NLP tasks," in *Proceedings of the 2022 International Conference on NLP and AI*, 2022.
- [29] I. Sutskever et al., "Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems*, 2014.
- [30] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proceedings of EMNLP 2014*, 2014.
- [31] S. Alhumoud et al., "Enhancing Arabic chatbots using Transformer-RNN hybrid architectures," *Journal of Artificial Intelligence Research*, vol. 68, no. 4, pp. 115-132, 2023.
- [32] A. Farghaly and K. Shaalan, "Arabic natural language processing: Challenges and solutions," *ACM Transactions on Asian Language Information Processing*, vol. 8, no. 4, pp. 1-22, 2009.
- [33] T. Al-Hanai et al., "Optimizing Arabic dialogue systems using reinforcement learning and transformer architectures," *Computational Intelligence*, vol. 38, no. 2, pp. 312-330, 2022.
- [34] A. A. Hussein, K. M. Hussein, H. H. Saleh, and I. H. Farhan, "Survey towards a sustainable information and communication technologies (ICT) in Iraq," *Journal of Physics: Conference Series*, vol. 1530, no. 1, p. 012089, May 2020.

- [35] T. N. Alruqi and S. M. Alzahrani, "Evaluation of an Arabic chatbot based on extractive question-answering transfer learning and language transformers," *AI*, vol. 4, no. 3, pp. 667-691, 2023.
- [36] J. Lafferty et al., "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of ICML*, 2001.
- [37] O. Obeid et al., "CAMEL tools: An open-source toolkit for Arabic NLP," in *Proceedings of LREC 2020*, 2020.
- [38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [39] F. Fernández-Martínez, C. Luna-Jiménez, R. Kleinlein, D. Griol, Z. Callejas, and J. M. Montero, "Fine-tuning BERT models for intent recognition using a frequency cut-off strategy for domain-specific vocabulary extension," *Applied Sciences*, vol. 12, no. 3, p. 1610, 2022.
- [40] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.
- [41] H. H. Saleh, W. Nsaif and L. Rashed, "Design and implementation of a web-based collaborative E-learning model: A case study — Computer Science Department Curriculum," in *2018 1st Annual International Conference on Information and Sciences (AiCIS)*, Nov. 2018, pp. 193-200.
- [42] C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," in *Proceedings of the Workshop on Text Summarization*, 2004.
- [43] I. Chalkiadakis, G. W. Peters, and M. Ames, "Hybrid ARDL-MIDAS-Transformer time-series regressions for multi-topic crypto market sentiment driven by price and technology factors," *Digital Finance*, vol. 5, no. 2, pp. 295-365, 2023.
- [44] D. Peras, "Chatbot evaluation metrics," *Economic and Social Development: Book of Proceedings*, pp. 89-97, 2018.
- [45] S. K. Yuwono, B. Wu, and L. F. D'Haro, "Automated scoring of chatbot responses in conversational dialogue," in *9th International Workshop on Spoken Dialogue System Technology*, pp. 357-369, Springer, Singapore, 2019.
- [46] S. M. Suhaili, N. Salim, and M. N. Jambli, "Service chatbots: A systematic review," *Expert Systems with Applications*, vol. 184, p. 115461, 2021.