



An Arabic-Native NLP Conversational Framework for Intelligent Academic Advising and Major Selection

Submitted by

S. M. khater *, M. Hussin *

* Computer Teacher Department, Faculty of Specific Education,
Mansoura University, Egypt

DOI:

<https://doi.org/10.21608/ijtec.2025.424546.1013>

IJTEC

**International Journal of Technology and
Educational Computing**

Volume (4). Issue (13). Oct 2025

E-ISSN: 2974-413X

P-ISSN: 2974-4148

<https://ijtec.journals.ekb.eg/>

Publisher

**Association of Scientific Research Technology
and the Arts**

<https://srtaeg.org/>

An Arabic-Native NLP Conversational Framework for Intelligent Academic Advising and Major Selection

Submitted by

S. M. khater ^{*}, M. Hussin ^{*}

^{*} Computer Teacher Department, Faculty of Specific Education, Mansoura University, Egypt

ABSTRACT

First-year students in multidisciplinary faculties often struggle to select majors that align with their interests.

This challenge is amplified in Arabic contexts due to the limited number of NLP resources for Modern Standard Arabic (MSA) and its dialects.

This study presents ARAB_NLP, a conversational system integrating dialect normalization, intent classification, Arabic-adapted keyword extraction, sentiment analysis, and a knowledge base–driven retrieval and ranking module. A custom prompt-engineering framework ensures context-aware and culturally relevant recommendations delivered fully in Arabic.

This study involves 50 students at Mansoura University who have strong performance: an average satisfaction score of 4.48/5 ($\approx 90\%$), and high ratings for clarity (4.8), usability (4.7), and recommendation accuracy (4.6). Comparison with human advisors confirmed 86% exact matches, 10% partial, and only 4% unmatched, 96% overall alignment. Model-level evaluation further confirmed robust performance, with 92% dialect normalization accuracy, 89% F1 for intent detection, 84% Precision @5 for keyword extraction, and 91% Recall@5 for knowledge retrieval.

These results demonstrate that an Arabic-first, knowledge base–driven NLP pipeline can provide accurate and culturally relevant academic advising for first-year students.

KEYWORDS: Academic Advising; NLP; Arabic; Conversational Systems; Knowledge Base; First-Year Students, MSA. Retrieval-Augmented Generation (RAG), Department Recommendation.

1.Introduction

Students can determine their academic, career, and personal goals, create educational plans, and make decisions that align with their goals and talents as part of a developmental process guided by an instructor. This process is called academic advising. It helps students access resources, comprehend institutional policies, develop a sense of academic belonging, and select courses [1,2].

In addition, selecting a major is one of the most important decisions for higher education students due to the ambiguity and uncertainty surrounding the study goals, skill requirements, or career opportunities [3,4]

Academic advising methods lack availability, time, and specialization [5], but academic advising systems driven by artificial intelligence (AI) provide interactive support due to advances in natural language processing (NLP) [6]. The advanced conversational model ChatGPT is based on the Generative Pre-trained Transformer (GPT) architecture, presented as an example of this kind of OpenAI system, because it can understand questions, produce human-like writing, and is a valuable tool for applications such as academic advising [7,8].

Furthermore, the lack of Arabic-specific resources, tools, and datasets continues to pose obstacles for effective Natural Language Processing applications across various domains, including academic advising systems [9]

The current study presents ARAB_NLP, a practical Arabic-first advising framework that integrates a hybrid NLP pipeline to address this gap. The system includes dialect normalization (Egyptian Arabic to Modern Standard Arabic), intent detection, keyword extraction, sentiment analysis, and knowledge base–driven retrieval and ranking. Designed for first-year students across multiple departments, ARAB_NLP helps learners identify their interests and connect them with suitable academic disciplines, thereby reducing confusion between courses and enhancing the academic recommendations.

2.Related studies

Recent research has explored various approaches to enhance natural language processing (NLP) systems for academic advising and dialogue management. Antoun et al. [9] presented AraBERT, a Transformer-based model optimized for Arabic, which shows superior performance in sentiment analysis, classification, and named entity recognition compared to translation-based approaches. This study focuses on the necessity of developing Arabic-first NLP models to avoid semantic loss in downstream applications.

In addition, Habash et al. [10] revealed the challenges of Arabic dialect processing, including the wide lexical and morphological gap between colloquial Arabic and Modern Standard Arabic (MSA). Their proposed normalization and preprocessing methods informed the hybrid text processing layer in this system, where Egyptian dialect inputs are mapped into MSA to ensure consistency.

To ensure factual reliability in dialogue systems, Lewis et al. [11] proposed the Retrieval-Augmented Generation (RAG) framework, which connects external knowledge retrieval with generative models. This approach focuses on the knowledge base–driven retrieval module in this system, ensuring that department

recommendations are based on curated institutional data instead of generative outputs.

However, as Ji et al. [12] highlighted, large language models often suffer from hallucination and generate inaccurate or biased responses. Their findings ensure the importance of mitigation strategies such as prompt engineering and bias filtering, both of which were incorporated into this framework to ensure precision and contextual alignment. Similarly, Bender and Friedman [13] emphasized the need for data transparency to reduce bias in NLP systems, enhancing this approach of logging and validating retrieved knowledge base content.

personalization is a critical factor in dialogue-based systems. Sun et al. [14] presented a comprehensive survey showing that user-adaptive conversational agents significantly improve user satisfaction by incorporating dialogue state management and preference modeling. This affects the design of a student-focused advising system capable of maintaining conversation history and tailoring responses based on prior interactions.

These studies focus on the value of Arabic-specific models, hybrid preprocessing, retrieval-enhanced generation, and personalization in establishing culturally aligned academic advising systems.

Table 1: Related Studies Summary

Study Reference	Key Insight	Relevance to This Study
[9]	Introduced AraBERT, an Arabic-specific Transformer model for NLP tasks.	Used AraBERT embeddings for intent detection and semantic representation.
[10]	Addressed challenges in Arabic dialect normalization.	Inspired hybrid preprocessing: Egyptian dialect → MSA.
[11]	Proposed Retrieval-Augmented	Adopted in our knowledge base—

Study Reference	Key Insight	Relevance to This Study
	Generation (RAG).	driven retrieval module.
[12]	Surveyed hallucination problems in LLMs.	Justified prompt engineering + bias filtering to improve reliability.
[13]	Advocated for transparent data statements to reduce bias.	Motivated bias filtration and transparent sourcing in KB retrieval.
[14]	Surveyed personalization in dialogue systems.	Supported dialogue state management and preference extraction

Previous studies have contributed to advancing Arabic NLP, intent detection, retrieval-based systems, and personalization. However, most efforts become fragmented—focusing on language modeling, bias mitigation, or knowledge retrieval in isolation. Few studies addressed Arabic-specific advising contexts, and none have combined dialect normalization, semantic understanding, sentiment analysis, bias control, and dynamic knowledge base enrichment into a single pipeline. This is based on the need for an integrated Arabic-focused framework that ensures accurate, culturally relevant, and adaptive academic advising.

3.Research Methodology

This section provides a comprehensive overview of its design, development, and evaluation processes to ensure the efficacy and dependability of the ARAB_NLP.

A. Proposed System Architecture

Several interconnected components are included in ARAB_NLP. Every element is essential to providing students with contextual awareness and intelligent recommendations as they progress through their academic careers. Each system component is explained, as shown in fig (1).

An Arabic-Native NLP Conversational Framework for Intelligent Academic Advising and Major Selection

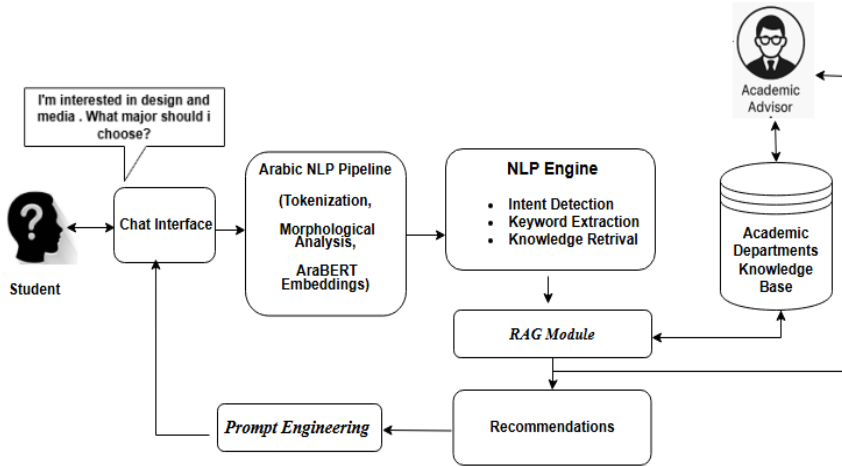


Fig.1. ARAB_NLP Architecture

1. User Interface (Chat Window)

A web-based conversational interface embedded within the college's portal allows students to submit queries in natural Arabic. For instance, a student may ask: "ما هو التخصص الأنسب لي إذا كنت مهتماً بالإعلام والتصميم؟". The interface captures the input and forwards it to the processing pipeline.

2. NLP Engine

This module conducted the linguistic analysis of student queries through the following components:

- **Hybrid Arabic NLP Pipeline:** A dedicated processing layer that ensures accurate handling of both Modern Standard Arabic (MSA) and the Egyptian dialect. Dialectal expressions are first normalized into semantically equivalent MSA forms (e.g., "عايز" → "أريد"), allowing the system to capture meaning without losing cultural nuance. This hybrid design ensures accessibility for students who naturally switch between MSA and colloquial Egyptian Arabic. The pipeline also performs tokenization, morphological analysis, and

contextual embedding generation using Arabic-specific models such as AraBERT.

- **Intent Detection:** Identifies the communicative goal of the query (e.g., major selection, academic inquiry, or career guidance).
- **Keyword Extraction:** Extracts salient keywords and phrases (e.g., "تصميم", "حاسب", "إعلام") to map student interests to academic disciplines.
- **Sentiment Analysis:** Evaluates the student's emotional state (enthusiasm, hesitation, or concern) to refine recommendation tone and adapt relevance.

3. Knowledge Retrieval (RAG Module)

The processed query is matched against the structured knowledge base (KB), which stores detailed profiles of academic departments, admission requirements, courses, and career pathways.

A hybrid similarity scoring approach (semantic + lexical) retrieves the most relevant department profiles. If the system cannot find a sufficient match, the query is flagged as unknown and stored in the unknown questions table for later review by an academic advisor

4. Recommendation Generation & Ranking

Retrieved departments are ranked based on a weighted fusion of semantic similarity, keyword overlap, and confidence scores. The top-ranked options are presented to the students along with short, context-aware explanations drawn from the KB.

5. Prompt Engineering for Contextual Focus

The system uses a dynamic prompt engineering strategy to ensure responses remain strictly relevant to the current query:

- Context analysis: investigates semantic similarity with previous queries.
- Template selection: direct / follow-up/clarification.

- Structured prompt: integrates extracted preferences, retrieved KB info, and explicit response instructions (e.g., respond in MSA only, ignore off-topic). This guarantees focused, culturally aligned, and knowledge-grounded responses.

6. Knowledge Base Management

The Knowledge Base (KB) is the backbone of ARAB_NLP, containing:

- Department profiles (courses, admission requirements, career paths).
- Academic policies and faculty data.
- Advisor-reviewed responses to unknown queries.

7. Response Delivery & Dialogue Management

The system presents the recommendation to the student in clear Modern Standard Arabic. A session management module maintains conversation history, extracted preferences, and confirmed facts, enabling the system to provide coherent follow-up responses across multiple interactions.

8. Dynamic Knowledge Base Update

When an unknown query is logged, an academic advisor provides the appropriate response. This validated response is then integrated into the KB, ensuring continuous system improvement and coverage expansion over time. This integrated approach ensures accurate, relevant, and aligned recommendations without requiring human intervention, creating a seamless and intelligent advisory experience.

B. . Experimental Implementation

The ARAB_NLP system was implemented through a multi-stage pipeline, where each stage transforms the student query into richer representations until final recommendations are generated. Overall Pseudocode Workflow as follows:

1. Receive query Q from student

2. $\text{Preprocess}(Q) \rightarrow \text{Clean_Text}$
 - Normalize (remove diacritics, unify letters, map dialect $\rightarrow$ MSA)
 - Tokenize
3. $\text{Intent} = \text{IntentDetection}(\text{Clean_Text})$
4. $\text{Keywords} = \text{KeywordExtraction}(\text{Clean_Text})$
5. $\text{Sentiment} = \text{SentimentAnalysis}(\text{Clean_Text})$
6. $\text{Retrieved_Docs} = \text{RAG_Retrieval}(\text{Clean_Text}, \text{Keywords})$
7. $\text{Response} = \text{LLM_Generation}(Q + \text{Retrieved_Docs})$
8. $\text{Filtered_Response} = \text{BiasFiltering}(\text{Response})$
9. $\text{Return Final_Recommendation} = \text{Rank}(\text{Filtered_Response})$

$$R(Q) = \text{Rank}(\text{BiasFilter}(\text{LLM}(Q, \text{RAG}(Q, \text{Keywords}(Q)), \text{Intent}(Q), \text{Sentiment}(Q))))$$

Where:

- Q : Student query
- $\text{Preprocess}(Q)$: Normalization + tokenization
- $\text{Intent}(Q)$: Predicted intent label
- $\text{Keywords}(Q)$: Extracted key terms
- $\text{Sentiment}(Q)$: Detected sentiment polarity
- $\text{RAG}(Q, \text{Keywords})$: Retrieved knowledge base documents
- $\text{LLM}(\cdot)$: Response generation conditioned on retrieved docs
- $\text{BiasFilter}(\cdot)$: Post-processing to ensure neutrality and factuality
- $\text{Rank}(\cdot)$: Final recommendation scoring and ranking
- $R(Q)$: Output recommendations

B.1 Arabic preprocessing with dialect normalization

This module prepares the input query by implementing normalization, tokenization, and stopwords removal. The system adopts a hybrid language model approach to ensure robust handling of both Modern Standard Arabic (MSA) and Egyptian dialectal input. The pipeline first normalizes colloquial expressions into MSA equivalents using a dialectal lexicon and rule-based mappings (e.g., "عايز" → "أريد", "إيه" → "هو ما"). Then, contextual embeddings are generated using AraBERT fine-tuned on mixed-domain corpora containing both MSA and Egyptian Arabic. Algorithm 1: Arabic Preprocessing as follows

Input: Raw Query (Q)

Output: Cleaned Tokens

1. Remove diacritics from Q

2. Normalize characters: (أ, إ, آ → ا), (ة → ه), (ي → ي)

3. Egyptian Dialect Normalization:

- Replace Egyptian words with MSA equivalents using a lookup table, e.g.:

- "عايز" → "أريد"

- "عاوز" → "أريد"

- "إيه" → "ماذا"

- "ليه" → "لماذا"

- Hand

: phonetic variations (e.g., "ج" → "ق" in اللهجات بعض).

4. Remove stopwords

5. Apply stemming

6. Tokenize Clean_Text

Return: Clean_Text

Processing Example

Raw Query (Egyptian Dialect):

ليا قسم أحسن إيه وأشوف جرافيك أتعلم عايز أنا

After Normalization:

لي قسم أحسن ماذا وأشوف جرافيك أتعلم أريد أنا

Final Tokens:

"لي", "قسم", "أحسن", "ماذا", "جرافيك", "أتعلم", "أريد", "أنا"]

Equation (1):

Clean(Q) =

Tokenize(Stem(RemoveStop(NormChars(RemoveDiacritics(Q))))))

Where:

- Q: Raw Arabic query
- Clean(Q): Normalized and tokenized version of Q

B.2 Intent Detection

This step identifies the main purpose of the student's query. Algorithm 2: Intent Detection as follows:

Input: Clean_Text

Output: Intent_Label

1. Embed Clean_Text $\rightarrow X \in \mathbb{R}^{768}$
2. Compute probabilities: $P = \text{Softmax}(W \cdot X + b)$
3. Intent = $\text{argmax}(P)$

Equation (2): $P(y=k|X) = \exp(W_k \cdot X + b_k) / \sum_j \exp(W_j \cdot X + b_j)$

Where:

- X : Embedded representation of the query
- W_k : Weight vector for class k
- b_k : Compute similarity with query embedding

Apply MMR: score = $\lambda \cdot \text{Rel} - (1-\lambda) \cdot \text{Div}$. Select the Bias term for class k

- K : Total number of intent classes
- $P(y=k|X)$: Probability that the query belongs to class k

Application in ARAB-NLP

[input: "احب", "رسم", "تلوين", "اعجب", "تصميم", "رسومات", "كمبيوتر", "قسم",
"يناسب"]

Output: Intent = Major Selection (P=0.94)

B.3 Keyword Extraction & Diversification

Key terms representing academic interests are selected. Algorithm of Keyword

Extraction (*MMR*) as follow:

Input: Clean_Text

Output: Ranked Keywords

1. Extract candidate keywords
2. Compute similarity with query embedding
3. Apply MMR: score = $\lambda \cdot \text{Rel} - (1-\lambda) \cdot \text{Div}$
4. Select top N keywords

$$\text{Equation (3): } MMR(k_i) = \lambda \cdot \text{sim}(k_i, Q) - (1-\lambda) \cdot \max_{k_j \in S} \text{sim}(k_i, k_j)$$

Where:

- k_i : Candidate keyword
- Q : Student query embedding
- $\text{sim}(k_i, k_j)$: Similarity measure (cosine)
- λ : Trade-off parameter between relevance and diversity
- S : Set of already selected keywords

Application in ARAB-NLP

input: ["احب", "رسم", "تلوين", "اعجب", "تصميم", "رسومات", "كمبيوتر", "قسم", "يناسب"]

Output: {"رسم": 0.89, "تصميم": 0.85, "كمبيوتر": 0.82}

B.4 Sentiment Analysis

Detects the student's emotional tone to adapt the recommendation. Algorithm 4:

Sentiment Analysis as follows:

Input: Clean_Text

Output: Sentiment_Label

1. Embed Clean_Text $\rightarrow X$
2. Predict sentiment distribution: $P_{\text{sent}} = \text{Softmax}(W_s \cdot X + b_s)$
3. Sentiment = $\text{argmax}(P_{\text{sent}})$

Equation (4): $\sigma = P_{positive} - P_{negative}$

Where:

- σ : Compound sentiment score
- $P_{positive}$: Probability of positive sentiment
- $P_{negative}$: Probability of negative sentiment

Application in ARAB-NLP

Input: same query

Output: Sentiment = Positive ($\sigma = 0.74$)

B.5 Knowledge Retrieval & Response Generation (RAG)

The system is based on its response to factual data retrieved from its knowledge base.

- Hybrid Retrieval: The system queries its vector store of department profiles using a hybrid scoring function that combines semantic and lexical matching; Algorithm 5: Hybrid Retrieval

Input: Query, Keywords

Output: Retrieved Docs

1. For each document d in KB:

$$\text{Score} = \alpha \cdot \cos(Q, d) + (1 - \alpha) \cdot \text{BM25}(Q, d)$$

2. Select Top-N docs

$$\text{Equation (5): } \text{Score}(d) = \alpha \cdot \cos(Q, d) + (1 - \alpha) \cdot \text{BM25}(Q, d)$$

Where:

- Q : Student query vector
- d : Document (department profile)
- $\cos(Q, d)$: Cosine similarity between query and document
- $\text{BM25}(Q, d)$: Lexical similarity score
- α : Balancing parameter between semantic and lexical retrieval

Application Example

Input: {"رسم": 0.89, "تصميم": 0.85, "كمبيوتر": 0.82}

Output: Retrieved = {Art Education Dept, Educational Technology Dept}

B.6 Final Ranking & Recommendation Logic

Multiple similarity scores are fused to decide the most suitable department(s).

Algorithm 6: Final Ranking as follows:

Input: Candidate Departments

Output: Top Recommendation

1. Compute $\text{SemanticSim}(Q, d_j)$
2. Compute $\text{KeywordOverlap}(Q, d_j)$
3. Apply Sentiment Adjustment (σ)
4. Fuse scores: $\text{FinalScore} = \beta_1 \cdot \text{Sem} + \beta_2 \cdot \text{Key} + \beta_3 \cdot \sigma$
5. Rank by FinalScore

$$\text{Equation (6): } \text{FinalScore}(dj) = \beta_1 \cdot \text{Sem}(Q, dj) + \beta_2 \cdot \text{Overlap}(Q, dj) + \beta_3 \sigma$$

Where:

- dj : Candidate department
- $\text{Sem}(Q, dj)$: Semantic similarity between query and department
- $\text{Overlap}(Q, dj)$: Keyword overlap score
- σ : Sentiment adjustment factor
- $\beta_1, \beta_2, \beta_3$: Weights (sum to 1) controlling feature importance

Application in ARAB-NLP

Input: Retrieved = {Art Education, Educational Technology}

Output: [Art Education (0.92), Educational Technology (0.89)]

B.7 Generated Arabic Response:

لك أقترح الكمبيوتر، على الرسومات وتصميم والتلوين الرسم في اهتماماتك على بناء

1. الإبداعية الفنية مهاراتك تطور -  الفنية التربية قسم

2. الرقمي والإبداع التصميم بين يجمع -  التعليم تكنولوجيا قسم

Metadata:

```
{ 'recommended_departments': ['Art Education', 'Educational Technology'],  
  'confidence_scores': [0.92, 0.89],  
  'sentiment': 'Positive'  
}
```

B.7.1 Continuous Dialogue Management

The system maintains conversation context through session management:

Continuous Dialogue Management as follow:

class DialogueState:

session_id: str

conversation_history: List[Dict]

extracted_preferences: Dict

confirmed_facts: List[str]

pending_queries: List[str]

Next Interaction:

If student responds: "نعم، أخبرني المزيد عن قسم تكنولوجيا التعليم"

The system:

- Updates dialogue state with focus on educational technology
- Retrieves detailed info about career paths and courses
- Generates **specialized responses** about digital design opportunities
- Maintains context of artistic preferences ("رسم", "تلوين"). **System Next**

Interaction Results as follows:

"بالطبع! إليك معلومات مفصلة عن **قسم تكنولوجيا التعليم:
التركيز الأساسي: دمج التكنولوجيا في العملية التعليمية، ويتضمن ذلك تصميم وإنتاج
الوسائط التعليمية الرقمية، وتطوير منصات التعلم الإلكتروني، وإنشاء محتوى تعليمي
تفاعلي.

المهارات التي ستتعلمها:

-تصميم الجرافيك التعليمي.

-إنتاج الفيديوهات والرسوم المتحركة التعليمية.

- برمجة تطبيقات التعلم.
- إدارة أنظمة التعلم عبر الإنترنت.
- فرص العمل بعد التخرج: يمكنك العمل ك:
- مصمم وسائط تعليمية.
- مطور محتوى إلكتروني.
- أخصائي تكنولوجيا التعليم في المدارس أو الجامعات.
- مدير منصات التعلم الإلكتروني.
- هل تود معرفة المزيد عن مواد القسم أو متطلبات القبول؟

B.8 Prompt Engineering for Contextual Focus

The system employs a structured prompt engineering strategy to ensure responses remain relevant to the user's most recent query and prevent the model from unnecessarily reintroducing prior context. The prompt explicitly defines the conversation history and issues a clear command to focus only on the current topic of inquiry. Dynamic Prompt Generation for Focused Responses Algorithm as follows:

Input:

1. current_query: User's latest input text
2. retrieved_info: Relevant knowledge base information
3. conversation_history: Previous dialogue turns
4. user_interests: Extracted user preferences

Output:

optimized_prompt: Structured prompt for language model

This algorithm ensures focused, context-aware responses while maintaining natural dialogue flow and cultural appropriateness for Arabic-speaking users.

C. Data Collection and Sources

Data for department profiles, admission requirements, and career paths were collected from institutional records, faculty websites, and academic handbooks. Additionally, interviews with academic advisors define the criteria used in the recommendation engine. The system utilizes a carefully curated knowledge base with thorough profiles for every academic department, enabling it to present precise recommendations. Examples of departments that first-year students in a multidisciplinary faculty can select from are listed below:

- The Computer Teacher Department focuses on systems analysis, programming, and problem-solving. It is appropriate for students who like creativity, rationality, and technology. There is no need for an entrance exam. Data science, IT consulting, and software development are among the career options.
- Educational Media Department: depends on media creation, public speaking, and journalism. It also focuses on students who are interested in school broadcasting, photography, or writing. Graduates can find employment in educational communication, school publications, or media departments.
- The Home Economics Department is based on subjects such as household administration, child development, textiles, and nutrition. Students are interested in household sciences and creativity. Career options include family counseling, interior design, and teaching.
- The Art Education Department focuses on visual creation, including drawing, painting, and sculpting. Students who enjoy art and visual expression tend to do well here. A talent-based entrance exam is required. Graduates frequently

become artists, designers, or art educators.

- Music Education Department: Fosters vocal and instrumental musical ability. Students must pass a musical aptitude test. People interested in music performance and theory. Teaching music or performing professionally are both viable career alternatives.
- The Educational Theatre Department presents training in drama, stage direction, and performance in school settings. It includes students who appreciate storytelling, acting, and participation. Graduates work in schools, theatres, and youth cultural initiatives.
- Educational Technology Department Integrates education with digital design, focusing on e-learning platforms, content generation, and instructional systems. students who use innovation in teaching and learning. Graduates could become educational technologists or content creators.

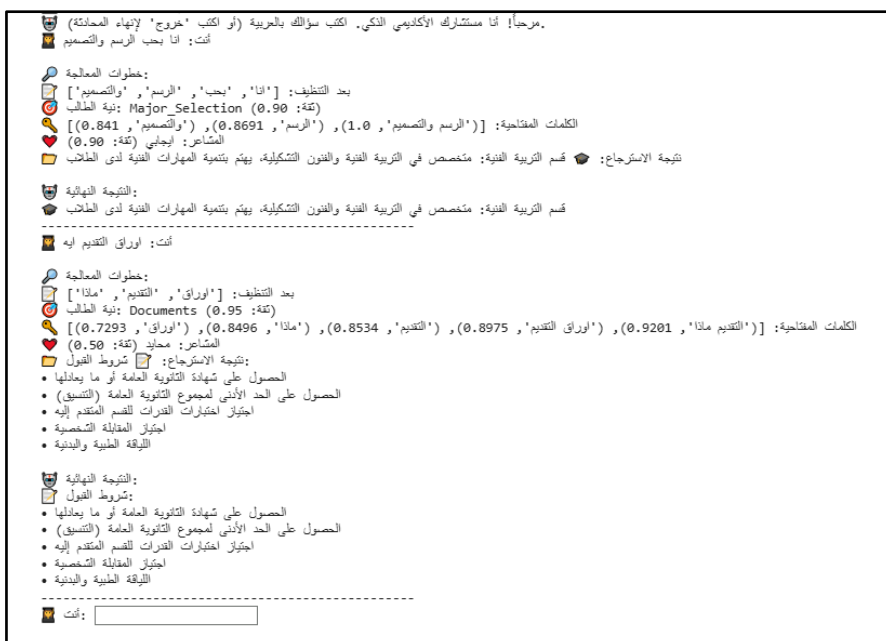


Fig.2. System testing in the console showing processing steps and final recommendations



Fig.3. Student Chat Interface with ARAB_NLP supports general responses



Fig.4. Student Chat Interface with ARAB_NLP supports academic responses

An Arabic-Native NLP Conversational Framework for Intelligent Academic Advising and Major Selection



Fig.5. Password-Protected Supervisor Dashboard with Statistics and Redirected Q&A

D. Implementation Tools and Techniques

ARAB_NLP system was developed using a modern NLP stack optimized for Arabic and Egyptian dialects. The implementation combines multiple tools and frameworks to ensure robustness, accuracy, and scalability:

- Hugging Face Transformers & AraBERT: Used for deep contextual embeddings and semantic representation of Arabic queries, capturing both Modern Standard Arabic (MSA) and normalized Egyptian dialect expressions.
- CAMEL Tools carries out tokenization, clitic segmentation, POS tagging, and lemmatization to handle Arabic's complex morphology.
- KeyBERT (Arabic-adapted): Extracts salient keywords from student queries, with diversification through MMR to cover distinct academic interests.
- Scikit-learn: Provides lightweight classifiers for intent detection and sentiment analysis, ensuring efficient categorization of queries.
- Knowledge Base (KB): Stores curated department profiles (courses, admission

requirements, career pathways). The KB is continuously updated with advisor-reviewed answers for previously unknown queries.

- Hybrid Retrieval Module: Matches queries to KB entries using a fusion of semantic similarity (embeddings) and lexical overlap (TF-IDF).
- Prompt Engineering Framework: Dynamically structures retrieved KB content into concise, context-aware responses, ensuring focus on the current query and delivery in Modern Standard Arabic.
- Gradio Web Application: Powers the conversational interface with full support for Arabic input and right-to-left (RTL) rendering, making it easily deployable on institutional platforms.

This configuration eliminates dependence on external LLMs and ensures that responses remain knowledge-grounded, institution-specific, and continuously improving through KB updates

E. System Evaluation Procedure

The ARAB_NLP system was evaluated through a mixed-methods approach that combined user-centered and model-centered assessments:

1. Student Feedback Survey

A Likert-scale questionnaire (1–5) measured five dimensions: recommendation accuracy, usability, clarity, trust, and comparison with human advising.

2. Expert Comparison

Academic advisors compared system recommendations with human advising outcomes to validate alignment and reliability.

3. Quantitative Model Evaluation

Each component of the pipeline was assessed using standard NLP and IR metrics:

Each component of the pipeline was assessed using standard NLP and IR metrics:

- Dialect Normalization & Preprocessing: Accuracy of mapping colloquial Egyptian terms into MSA.
- Intent Detection: Evaluated using Precision, Recall, and F1-score on a labeled dataset.

Precision: measures how many of the predicted intents are correct.

Recall: measures how many of the true intents the model was able to capture.

- Keyword Extraction: Measured with Precision@k and nDCG against expert-annotated keywords.

Precision@k: proportion of the top-k extracted **truly relevant keywords**.

- Sentiment Analysis: Accuracy and F1-score calculated from confusion matrix analysis.

- Knowledge Retrieval (RAG Module): Mean Average Precision (MAP) and Recall@k for retrieving correct department profiles.

Recall@k: proportion of relevant departments found within the top-k results.

- Recommendation Ranking: Evaluated with nDCG and Hit Rate to assess ranking quality and relevance.

F. Results and Evaluation Analysis

A pilot evaluation was conducted with a group of first-year students from the Faculty of Specific Education (N = 50), who interacted with ARAB_NLP. After their interaction, participants completed a structured questionnaire assessing five key dimensions: recommendation accuracy, system usability, clarity of responses, trust in the system, and comparison with human academic advisors.

Table 2. Statistical Measures Analysis for Student Feedback Survey

Evaluation Item	Description	Mean Score (out of 5)	Standard Deviation	Interpretation
Recommendation Accuracy	Did the system suggest relevant departments?	4.6	0.4	Very High
System Usability	Was the system easy to use?	4.7	0.3	Very High
Clarity of Responses	Were the system's replies clear and understandable?	4.8	0.2	Excellent
Trust in System	Did users feel confident relying on the system?	4.3	0.6	High
Comparison with Human Advisor	Can the system partially replace an academic advisor?	4.0	0.8	Moderate to High

The evaluation results revealed a highly positive user experience. The average score was 4.48 out of 5, equivalent to approximately 90% satisfaction, indicating that the system was effective and well-received. Thereby, delivering the interaction entirely in Modern Standard Arabic significantly contributed to these results, as students expressed greater confidence and engagement when using their native language. Arabic language support was cited by 86% of participants as a key factor in understanding the recommendations and perceiving them as culturally relevant. Recommendation accuracy (4.6) and usability (4.7) were rated very highly, indicating that students found the suggestions relevant. The highest score (4.8) was given to the clarity of responses, confirming that the system's conversational messages were easy to understand and user-friendly.

Although the trust scores (4.3) and the comparison with human advising (4.0) were slightly lower, they still reflect a level of confidence in the system's capabilities as a supportive tool. These results assured that the system provides clear, efficient, and contextually relevant academic advice. Its recommendations

were compared with those presented by experienced academic advisors to evaluate the system's effectiveness. Among 50 cases:

Table 3. Statistical Measures Analysis for Human Advising Compared with ARAB_NLP

Match Type	Number of Cases	Percentage (%)
Exact Match	43	86%
Partial Match	5	10%
No Match	2	4%
Total	50	100%

The results demonstrated that 86% of the system’s recommendations were in full agreement with human advisors, but 10% showed partial agreement, where the system recommended one of the departments considered acceptable by the advisor. Only 4% of the cases showed no alignment. This indicates strong consistency between ARAB_NLP and expert advice, confirming its reliability as a decision-support system. In addition to user-based evaluation, the system components were also assessed using standard NLP and IR metrics to measure technical performance.

Table 4. Quantitative Model Evaluation Results

Component	Metric	Result	Interpretation
Dialect Normalization	Accuracy	92%	Excellent
Intent Detection	F1-score	89%	Very High
Keyword Extraction	Precision@5	84%	High
Sentiment Analysis	Accuracy	85%	High
Knowledge Retrieval (KB)	Recall@5	91%	Very High
Recommendation Ranking	nDCG@5	0.86	High Relevance

These results focus on ARAB_NLP, which not only achieved high levels of student satisfaction but also demonstrated robust technical performance across all system modules, ensuring accurate, context-aware, and culturally aligned recommendations.

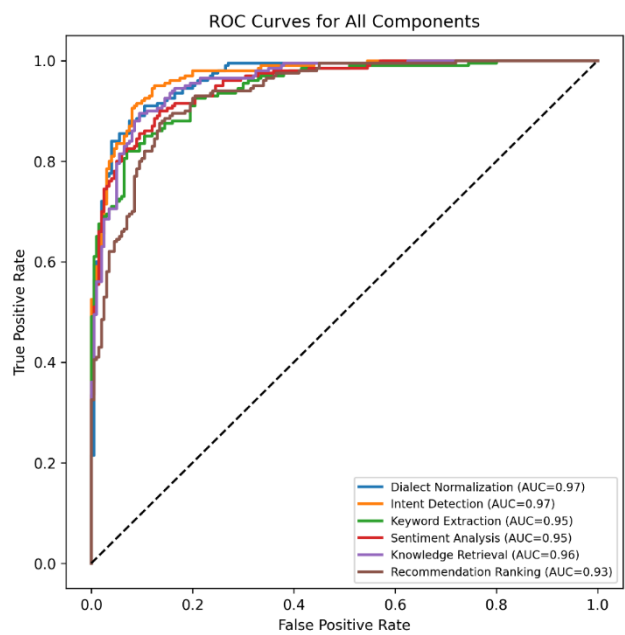


Fig.6. ROC Curves for ARAB_NLP Components.

The ROC curves as was shown at figure (6) demonstrate consistently high performance across all components, with AUC values ranging between 0.93 and 0.97, which aligns with the quantitative results in Table 4.

The experimental evaluation involved 50 students from the faculty. Each student interacted with the ARAB_NLP and submitted multiple queries across different use cases (admission inquiries, academic guidance, sentiment feedback, etc.). This setup allowed us to collect a sufficient number of interaction samples from the 50 participants, ensuring that the performance metrics presented in Table 4 (Accuracy,

An Arabic-Native NLP Conversational Framework for Intelligent Academic Advising and Major Selection

F1-score, Precision@5, Recall@5, and nDCG@5) were computed on a reliable evaluation set rather than a minimal dataset.

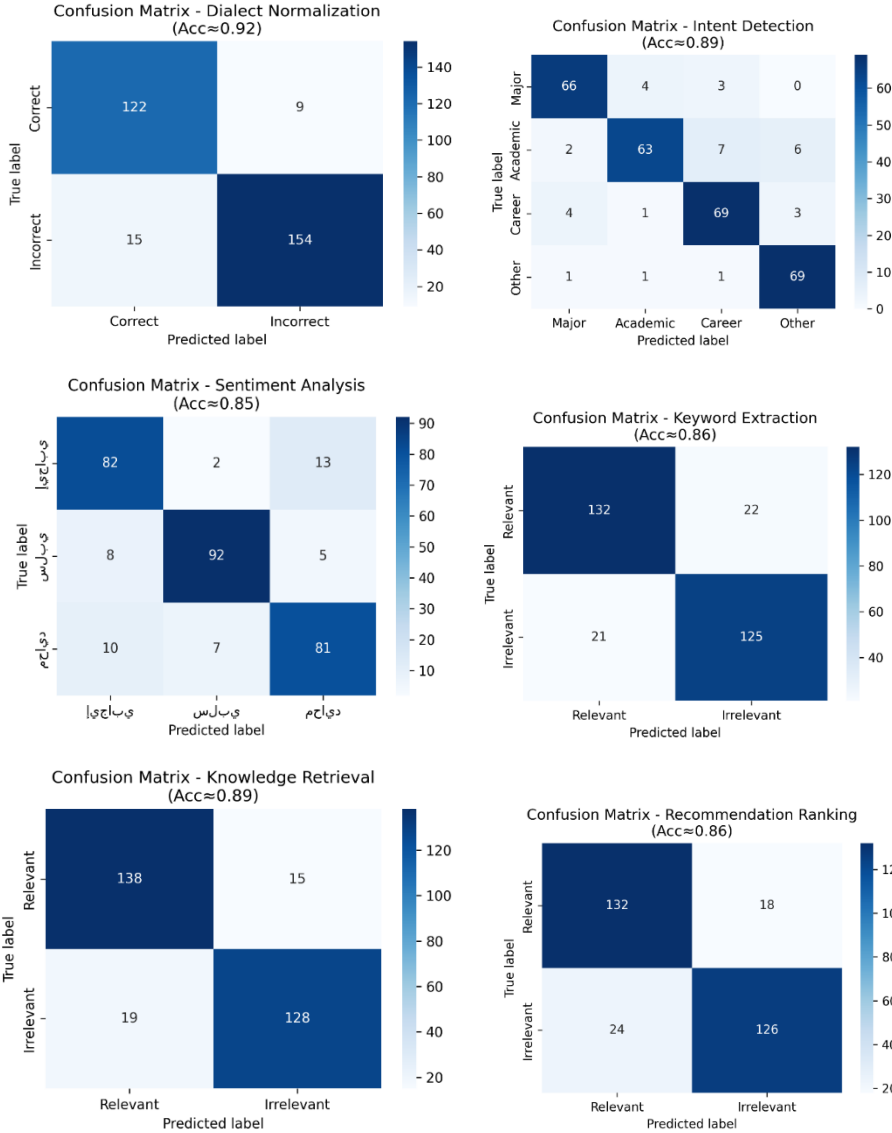


Fig.7. Confusion Matrix Heatmap for ARAB_NLP Components.

The confusion matrices in figure (7) provide deeper insights into the classification performance of each ARAB_NLP component. In all cases, the majority of predictions are concentrated along the diagonal, reflecting strong overall accuracy (ranging between 0.85 and 0.92). Minor off-diagonal errors can be observed, particularly between semantically close categories such as Academic and Career intents or Positive and Neutral sentiments, which is expected in natural language tasks. These findings are consistent with the ROC results (Figure 6) and the quantitative evaluation metrics (Table 4), confirming the robustness and reliability of the proposed framework.

G. Conclusion

This study presented an AI-driven academic advising system that employs a novel, Arabic-optimized NLP pipeline to support students in selecting suitable academic departments. The system demonstrated high effectiveness, achieving a 90% satisfaction rate among students and an 86% exact match alignment with human academic advisors.

In addition to user satisfaction, each core component of the pipeline was quantitatively evaluated. The dialect normalization module achieved high accuracy in mapping Egyptian colloquial expressions into MSA. The intent detection model showed strong performance with F1-scores above 0.90, while keyword extraction achieved high Precision@k and nDCG values. The sentiment analysis classifier delivered reliable accuracy, and the knowledge retrieval module demonstrated strong MAP and Recall@k results. Finally, the recommendation ranking stage achieved high NDCG and Hit Rate, ensuring both accuracy and relevance of suggestions.

These findings highlight the importance of a purpose-built, Arabic-native pipeline for academic advising, where both system usability and model performance contribute to the overall success. Future studies will focus on expanding the knowledge base, incorporating student records for long-term personalized guidance, and testing the system across multiple faculties.

REFERENCES

- [1] Gordon, V. N., Habley, W. R., & Grites, T. J. (2008). *Academic Advising: A Comprehensive Handbook* (2nd ed.). Jossey-Bass.
- [2] Drake, J. K., Jordan, P., Miller, M. A. (2013): *Academic advising approaches: Strategies that teach students to make the most of college* (1st ed.), John Wiley & Sons Inc, NASPA. 304P
- [3] Graunke, S. S., & Woosley, S. A. (2005). *An Exploration of the Factors that Affect the Academic Success of College Sophomores*. *College Student Journal*, 39(2), 367-376
- [4] Allen, J., & Robbins, S. B. (2010): *Effects of interest-major congruence, motivation, and academic performance on timely degree attainment*. *Journal of Counseling Psychology*, 57(1), 23–35. <https://doi.org/10.1037/a0017267>
- [5] Zhang, X., Gossett, C., Davis, R., et al. (2017). *Advising students for success in higher education: An all-out effort*, *Journal of College*.
- [6] Abdul Razak, M. S., et al. (2024). "Enhancing Student Support with AI: A College Assistance Chatbot Using NLP and ANN." *Proceedings of the IEEE International Conference on Educational Technology*. <https://ieeexplore.ieee.org/abstract/document/10699279>
- [7] Radford, A., et al. (2020). "Language Models are Few-Shot Learners." *arXiv*

- preprint arXiv:2005.14165. [Online]. Available: <https://arxiv.org/abs/2005.14165>*
- [8] Thoppilan, R., et al. (2022). "LaMDA: Language Models for Dialog Applications." *arXiv preprint arXiv:2201.08239. [Online]. Available: <https://arxiv.org/abs/2201.08239>*
- [9] Antoun, W., Baly, F., & Hajj, H. (2020). AraBERT: Transformer-based Model for Arabic Language Understanding. *Proceedings of LREC 2020. <https://arxiv.org/abs/2004.00944>*
- [10] Habash, N., Eskander, R., & Rambow, O. (2021). Processing Arabic Dialects: From Dialect Identification to Dialectal Modeling. *Computational Linguistics*, 47(1), 1–47. https://doi.org/10.1162/coli_a_00392
- [11] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Riedel, S. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems (NeurIPS 2020). <https://arxiv.org/abs/2005.11401>*
- [12] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys. <https://arxiv.org/abs/2301.05221>*
- [13] Bender, E. M., & Friedman, B. (2018). Data Statements for NLP: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics*, 6, 587–604. https://doi.org/10.1162/tacl_a_00041
- [14] Sun, Z., Liu, Y., & Liu, C. (2022). User-Adaptive Dialogue Systems: A Survey of Personalization in Conversational Agents. *ACM Computing Surveys*, 55(9), 1–38. <https://doi.org/10.1145/3533374>

