Development of an Emotion Driven Chatbot for IT Career Guidance

Oluokun Samuel Olugbenga, Mishael Ayomiposi Osoba

Department of Information Systems and Technology, Kings University, Odeomu, Nigeria

DOI: https://doi.org/10.51584/IJRIAS.2026.11070106

Received: 19 July 2026; Accepted: 24 July 2026; Published: 07 August 2026

ABSTRACT

The increasing complexity of IT career options and the emotional challenges faced by students and early-career professionals in navigating career decisions have highlighted the need for intelligent, emotion-aware guidance systems. This project presents the design and implementation of an Emotion-Driven Chatbot for IT Career Guidance, a web-based conversational system that detects the emotional state of users in real time and delivers personalised, empathetic IT career recommendations accordingly.

The system was developed using Python 3.10 and integrates a RoBERTa-based transformer model (SamLowe/roberta-base-go_emotions), pre-trained on the GoEmotions dataset comprising 58,009 annotated text samples across 28 emotion categories, for fine-grained emotion classification. Detected emotions are mapped to seven broader emotional groups, anxiety, frustration, joy, curiosity, sadness, positive, and neutral which drive the career recommendation logic. The Groq LLaMA 3.3-70b large language model API was integrated to generate dynamic, contextually aware conversational responses informed by the detected emotion. The system was deployed as a web application using the Flask framework, featuring a professional landing page and an interactive chat interface with real-time emotion badge display, career recommendation cards, career deep dives, and curated learning resources covering nine IT career paths.

The RoBERTa emotion detection model was evaluated against the GoEmotions test split of 5,427 samples, achieving an overall accuracy of 64% and a weighted F1-score of 0.64, consistent with published benchmarks for 28-class fine-grained emotion classification. A qualitative user evaluation was conducted with 6 participants using a five-point Likert scale survey, producing an overall mean satisfaction score of 3.97 out of 5.00 (79.3%). Ease of use received the highest rating (M = 4.50), while emotion detection accuracy received the lowest (M = 3.50), indicating an area for future refinement.

The results confirm that the system successfully achieved all four project objectives and demonstrates the viability of emotion-driven conversational AI as a practical and accessible tool for personalised IT career guidance. Future work should address production deployment, larger-scale user evaluation, and the integration of multimodal emotion detection.

Keywords: Emotion Driven, Chatbot, IT, Career Guidance, RoBERTa, GoEmotions.

INTRODUCTION

The Information Technology (IT) industry has experienced a rapid growth becoming one of the most desirable career sectors globally. Reports indicate that the digital economy is projected to grow significantly, reaching approximately $24 trillion by 2025 (Digital Cooperation Organization DCO, 2024). The Information Technology (IT) industry has roles like UI/UX design, Software Engineering, Cybersecurity and so many diverse opportunities. However, the abundance of career opportunities also presents a major challenge to students and early-career professional growth, it leads to a “Paradox of choices” where an excess of options results in anxiety and decision paralysis rather than satisfaction (Kim *et al.,* 2024). Many individuals select their careers based on parental pressure, salary expectations, academic grades and also an abundance of learning resources. Research highlights that "future career confusion" has become a common psychological phenomenon among university students, negatively affecting their academic engagement and mental health.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Most traditional career guidance systems rely on static questionnaires or skill-based assessment. These systems generally fail to consider the emotional capability or personality of the individual during the decision-making process (Patil *et al.,* 2025). Emotional factors such as confusion, frustration, anxiety or motivation play a crucial role in how individuals absorb and process information, take decisions and learn. When these factors are ignored, the recommendation results are accurate but are poorly suited to the user's current psychological and learning readiness.

Recent advancements in the Artificial Intelligence (AI), specifically in Natural Language Processing (NLP) and Affective Computing, offer a new solution to this problem. Affective Computing allows machines to interpret human emotions and sentiments from text (Feng *et al.,* 2025). By integrating these technologies into a career recommendation system, it becomes possible to analyse a user's emotional state such as their frustration with rigid logic or their joy in visual creativity and map these feelings to suitable IT roles. This project seeks to bridge the gap between skill and emotion by designing a conversational agent that not only suggests career paths based on emotional analysis but also provides curated learning resources like books, videos and tutorial courses to guide the user's journey.

Chatbots, also known as conversational agents, have emerged as versatile tools in various domains, including education and career counseling, by facilitating interactive and personalized user experiences through natural language interfaces. These AI-powered systems leverage natural language processing (NLP) and machine learning algorithms to simulate human-like conversations, enabling users to query information, receive recommendations, and engage in dialogue in real-time (Adamopoulou & Moussiades, 2020). In the context of career guidance, chatbots have been increasingly adopted to address scalability issues in traditional counseling, offering 24/7 accessibility and reducing the burden on human advisors. For instance, studies have demonstrated that chatbots can effectively assess user skills, interests, and preferences through dynamic questioning, leading to tailored career suggestions in fields like IT (Smutny & Schreiberova, 2020).

The primary contributions of this research are:

i) gathered dataset relevant to IT career conversations containing emotional expressions, sentiment labels, and common IT career queries.

ii) developed a sentiment analysis model using Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach (RoBERTa).

iii) implemented the Emotion-Driven Chatbot for IT Career Guidance using Python.

iv) evaluated the system performance using metrics such as accuracy, precision, recall, F1-score (for sentiment model) and usability metrics (SUS, user satisfaction, response relevance).

LITERATURE REVIEW

Introduction to Artificial Intelligence

Artificial Intelligence is a branch of computer science that focuses on creating systems capable of performing tasks that typically require human intelligence. These tasks include learning, reasoning, problem-solving, perception, and language understanding. Artificial Intelligence is the study of agents that receive percepts from the environment and perform actions, emphasizing rational behavior and intelligent agents (Russell and Norvig, 2020).

Artificial Intelligence is the tangible real-world capability of non-human machines or artificial entities to perform, task solve, communicate, interact, and act logically as it occurs with biological humans (Gil de Zúñiga et al., 2024). It is the simulation of human-like processes, AI enables machines to learn from experience, adapt to new inputs, and execute tasks resembling human capabilities, or to perform cognitive functions that we associate with human minds, such as perceiving (Collins et al., 2021).

Chatbot and Conversational Agents

Chatbots, also known as conversational agents (CAs) or conversational AI, are AI-powered systems designed to simulate human conversation through text or voice interaction. They are widely used in various industries, including customer service, healthcare, and education. Chatbots are computer programs designed to simulate human text or verbal conversations (Car et al., 2020), or more broadly, software systems that mimic interactions with real people using natural language (Mariani et al., 2023). They leverage technologies like Natural Language Processing (NLP), machine learning, and increasingly large language models to understand user input, maintain context, and generate relevant responses (Adamopoulou & Moussiades, 2020).

Key Applications

i. Customer Service: Handle FAQs, support tickets, and personalized recommendations; reduce costs and response times significantly (e.g., 24/7 availability).

ii. Healthcare: Support mental health screening, symptom checking, medication adherence, patient education, and monitoring; improve accessibility but raise concerns around accuracy and equity (Car et al., 2020; Morias, 2026).

iii. Education: Act as tutors, provide explanations, quizzes, language practice, and administrative support; enhance personalization and engagement, though ethical use.

Other domains include finance, HR, and entertainment. By 2026, agentic conversational AI (autonomous, tool-using agents) is emerging, shifting from reactive chat to proactive assistance.

Transformer Models and ROBERTA

The Transformer architecture, introduced by Vaswani et al. (2017), marked a significant departure from sequential models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory networks (LSTMs). Rather than processing text one word at a time, the Transformer uses a self-attention mechanism that evaluates the relationships between every word in a sequence simultaneously, enabling the model to capture long-range linguistic dependencies and contextual meaning far more effectively than its predecessors. This architectural innovation became the foundation for the generation of large pre-trained language models that now define the state of the art in Natural Language Processing (NLP).

Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach (RoBERTa), developed by Liu et al. (2019) at Facebook AI Research (FAIR), represents one of the most influential instantiations of the Transformer architecture. Unlike earlier models that processed text either left-to-right or right-to-left, RoBERTa is trained bi-directionally, meaning it considers the full left and right context of every token in a sentence simultaneously. This is achieved through two unsupervised pre-training tasks: Masked Language Modelling (MLM), in which random tokens in the input are masked and the model learns to predict them from surrounding context, and Next Sentence Prediction (NSP), which trains the model to determine whether two sentences logically follow one another. Pre-training on these objectives over large corpora (English Wikipedia and BooksCorpus) equips RoBERTa with rich, context-sensitive language representations that can be fine-tuned on downstream tasks with comparatively small labelled datasets.

For the purpose of this project, the fine-tuning approach is particularly relevant. By providing RoBERTa with a relatively small number of labelled examples specifically, the GoEmotions dataset described in Section 2.4 the model's pre-trained weights are updated to classify text inputs into emotion categories rather than performing general language modelling. This process, known as supervised fine-tuning, adapts RoBERTa's deep contextual representations to the specific vocabulary, phrasing patterns, and emotional nuances present in conversational career guidance text. Studies have consistently demonstrated that fine-tuned RoBERTa models outperform traditional machine learning classifiers such as Support Vector Machines and logistic regression on text classification tasks, particularly where context and word order carry emotional significance (Devlin et al., 2019; Sathish et al., 2025).

The choice of RoBERTa over more recent large language models (LLMs) such as GPT-4 is justified by several considerations. RoBERTa's encoder-only architecture is purpose-built for classification tasks, producing a fixed-size contextual representation of an entire input sequence a property ideal for emotion classification. In contrast, decoder-based generative models prioritise text generation over classification. Furthermore, ROBERTA is computationally efficient to fine-tune on a single GPU, making it suitable for the resource constraints of a university-level research prototype. The RoBERTa-base-uncased variant (12 transformer layers, 768 hidden units, 12 attention heads, 110 million parameters) was selected for this project as it provides a strong performance baseline while remaining trainable within standard development environments (Devlin et al., 2019).

RoBERTa Algorithm

The RoBERTa (Robustly Optimized BERT Pretraining Approach) model operates through a series of mathematical transformations that convert raw text input into emotion classification probabilities. The algorithm is written below.

Algorithm 2.1: RoBERTa Emotion Classification Algorithm

Input: User text message $X = \{x_1, x_2, ..., x_n\}$

Output: Predicted emotion label $\hat{y}$ and confidence score

1. Tokenise input text X using RoBERTa tokeniser:

tokens = Tokeniser(X, max_length=128, padding=True, truncation=True) input_ids = tokens[input_ids] attention_mask = tokens[attention_mask]

2. Generate token embeddings:

For each token x_i in input_ids do:

$E(x_i)$ = TokenEmbedding(x_i) + PositionEmbedding(i) End For

3. For each encoder layer L = 1 to 12 do:

4. Compute self-attention: $Q = E(x_i) \times Wq$

$K = E(x_i) \times Wk$ $V = E(x_i) \times Wv$

$\text{Attention}(Q, K, V) = \text{Softmax}(QK^T / \sqrt{d_k}) \times V$

5. Compute multi-head attention:

$\text{MultiHead} = \text{Concat}(\text{head}_1, \text{head}_2, ..., \text{head}_{12}) \times W^O$

6. Apply feed-forward network: $\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2$

7. Apply layer normalisation and residual connection:

h_L = LayerNorm(x + FFN(MultiHead(x)))

8. End For

9. Extract [CLS] token representation from final encoder layer: h_CLS = h_{12}[CLS]

10. Pass through classification head:

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

logits = h_CLS × W_classifier + b_classifier

11. Apply softmax to get probability distribution over 28 emotion classes: P(emotion | X) = Softmax(logits)

12. Select emotion with highest probability: ŷ = argmax P(emotion | X) confidence = max P(emotion | X)

13.Map predicted label to emotion group:

If ŷ ∈ {nervousness, fear, confusion} → group = "anxiety"

If ŷ ∈ {anger, annoyance, disapproval, disgust} → group = "frustration" If ŷ ∈ {joy, excitement, amusement, admiration} → group = "joy"

If ŷ ∈ {curiosity, desire, surprise, realization} → group = "curiosity" If ŷ ∈ {sadness, grief, disappointment, remorse} → group = "sadness" If ŷ ∈ {gratitude, caring, love, pride, relief} → group = "positive" Else → group = "neutral"

Return ŷ, confidence, group

The GoEmotions Dataset

The GoEmotions dataset, released by Google Research (Demszky et al., 2020), is currently the largest manually annotated dataset for fine-grained emotion classification from text. It comprises 58,009 English-language comments sourced from Reddit, each annotated by human raters for one or more of 27 distinct emotion categories including admiration, amusement, anger, annoyance, approval, caring, confusion, curiosity, desire, disappointment, disapproval, disgust, embarrassment, excitement, fear, gratitude, grief, joy, love, nervousness, optimism, pride, realisation, relief, remorse, sadness, and surprise as well as a neutral category for emotionally ambiguous or absent expressions. The annotation scheme supports multi-label classification, acknowledging that a single utterance may simultaneously express multiple emotional states.

The dataset was constructed to address the limitations of earlier emotion corpora that relied on a small number of basic emotion categories, typically six (Ekman, 1992), which fail to capture the nuanced and overlapping emotional expressions present in everyday conversational text. GoEmotions' fine-grained taxonomy and its origin in naturally occurring social media language make it well-suited to training an emotion detection model for a conversational chatbot, where users express their emotional states indirectly through informal, context-dependent language.

For this project, the GoEmotions dataset serves as the primary resource for fine-tuning the ROBERTA model. The dataset's scale and diversity ensure that the trained classifier is exposed to a wide range of emotional expressions, improving its portability to the career guidance dialogue domain. Where individual emotion categories produce insufficient training samples for reliable classification, related emotions are grouped during preprocessing; for example, nervousness and fear are combined into a single anxiety-related class to balance the class distribution. The selection of GoEmotions over alternatives such as the SemEval emotion datasets is justified by its scale, its multi-label annotation scheme, and the availability of pre-processed splits that facilitate reproducible experimentation (Demszky et al., 2020).

Emotion detection in AI

Emotion detection, also referred to as emotion recognition or part of affective computing, is an important aspect of human-computer interaction (HCI). It involves analyzing user input to identify emotional states such as happiness, sadness, anger, confusion, fear, disgust, surprise, or neutral. This capability allows systems to respond more naturally, empathetically, and adaptively, enhancing user experience in applications like chatbots, virtual assistants, mental health tools, education platforms, and customer service (Picard, 1997; Khare et al., 2024).

Affective computing, encompasses the broader field of enabling machines to recognize, interpret, process, simulate, and respond to human emotions (Picard, 1997). Emotion detection forms a core pillar, shifting HCI from purely functional to emotionally intelligent interactions.

Artificial Intelligence in Career Guidance Systems

AI-based career guidance systems are designed to assist users in selecting suitable career paths based on their skills, interests, and preferences. These systems leverage machine learning, natural language processing (NLP), predictive analytics, and recommendation algorithms to provide personalized recommendations, often outperforming traditional methods in scalability, speed, and objectivity. However, most existing systems focus primarily on logical data e.g., academic performance, skills assessments, and job market trends and often ignore emotional factors, which play a crucial role in decision-making, such as motivation, anxiety, self-efficacy, passion, and fear of failure (Swami & Dhande, 2025; Bankins et al., 2024).

METHODOLOGY

This methodology was selected over purely quantitative or qualitative approaches for three reasons. First, the project's primary output is a working software system rather than a theoretical model or survey findings, making an artefact-centred approach most appropriate. Second, DSR explicitly requires evaluation of the developed artefact against clearly defined objectives, which aligns directly with Objective iv of this project. Third, the iterative nature of DSR accommodates continuous refinement of the system based on testing outcomes, which was essential given the complexity of integrating a transformer-based emotion detection model with a conversational dialogue manager and large language model response generator.

The research seeks to answer the following questions:

1. RQ1: How effectively can a fine-tuned transformer model (RoBERTa) detect fine-grained human emotions from conversational text inputs using the GoEmotions dataset?

2. RQ2: How can an emotion-aware dialogue manager and large language model (LLM) be integrated to map detected user emotional states into personalized, empathetic IT career path recommendations and learning resources?

3. RQ3: What level of overall user satisfaction, system usability, perceived emotion detection accuracy, and recommendation relevance does the emotion-driven career guidance chatbot achieve during user evaluation?

The workflow of the proposed system begins when a user submits a message through the interface. The input text first undergoes a preprocessing stage, where it is cleaned by removing unnecessary noise, special characters, and other irrelevant elements before being tokenized into smaller units that the model can interpret. This preprocessing ensures that the text is standardized and suitable for accurate analysis. The processed text is then passed to a RoBERTa-based emotion detection model, which serves as the core intelligence of the system. RoBERTa (Robustly Optimized Bidirectional Encoder Representations from Transformers) is a state-of-the-art natural language processing model that captures the contextual meaning of text and classifies the user's emotional state as positive, negative, or neutral. Based on the detected emotion, the system proceeds to perform intent recognition through keyword analysis to identify the user's specific request or career-related need. The identified intent is then matched with a career knowledge base, which retrieves relevant and domain-specific information to support accurate career guidance. Using both the detected emotional state and the retrieved knowledge, the system generates an adaptive response that is not only informative but also empathetic to the user's emotional condition, thereby enhancing the overall user experience. To assess the effectiveness of the system, a comprehensive evaluation phase is conducted using two approaches. First, the emotion detection model is tested against a labeled dataset, with its performance measured using standard machine learning metrics such as accuracy, precision, and recall. Second, real users interact with the chatbot and provide feedback through Google Forms, enabling user satisfaction and usability to be analyzed. The findings from both the technical evaluation and user feedback are combined to produce a final system performance evaluation, providing a comprehensive assessment of the chatbot's reliability, effectiveness, and overall user experience as shown below.

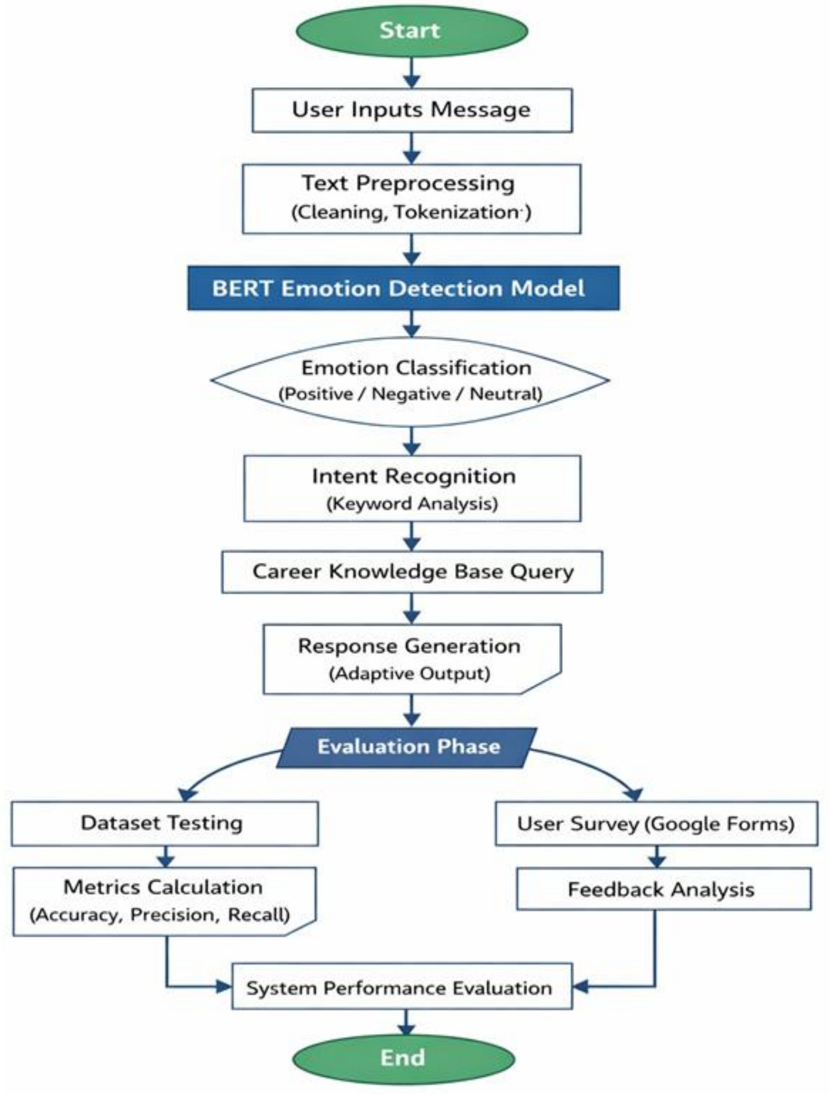

Figure 3.1 System Workflow

Objective I: Data Acquisition

The dataset used for this project was obtained from GoEmotions, a large-scale emotion dataset developed by Google Research, available at https://huggingface.co/datasets/go_emotions GoEmotions comprises 58,009 English-language text samples annotated across 28 emotion categories including joy, anger, fear, sadness, curiosity, nervousness, gratitude, and neutral, among others.

The dataset was split into three subsets for training, validation, and testing purposes. 80% of the data (43,410 samples) was used to train the RoBERTa emotion detection model, 10% (5,426 samples) was used for validation during training, and the remaining 10% (5,427 samples) was reserved for final testing and evaluation. The model achieved an overall accuracy of 64% and a weighted F1-score of 0.64 on the test split, the full results of which are reported in Chapter

2. The dataset was loaded using its official pre-defined splits to ensure reproducibility and comparability with published benchmarks:

GoEmotions was selected over alternative datasets such as the SemEval emotion corpora for three reasons. First, its scale of 58,009 samples provides sufficient coverage of diverse emotional expressions. Second, its fine-grained 28-category taxonomy captures the nuanced emotional states most relevant to career guidance conversations, particularly anxiety, confusion, and curiosity. Third, its origin in naturally occurring social media language makes it well-suited to training a model that processes informal conversational inputs from users of the chatbot system.

Table 3.1 Table of GoEmotions Dataset

Split	Number of Samples	Purpose
Training	43,410	Used to fine-tune the RoBERTa model weight
Validation	5,426	Used to monitor performance during training
Test	5,427	Used to generate final evaluation metrics

Data Preprocessing

Before the dataset could be used to train and evaluate the emotion detection model, a series of preprocessing steps were carried out to clean and prepare the data.

Data Cleaning

Duplicate entries, empty text samples, and samples with missing or ambiguous labels were removed from the dataset to ensure data quality and consistency.

Tokenisation

All text samples were tokenised using the RoBERTa tokeniser with a maximum sequence length of 128 tokens. Tokenisation converts raw text into numerical input IDs that the model can process.

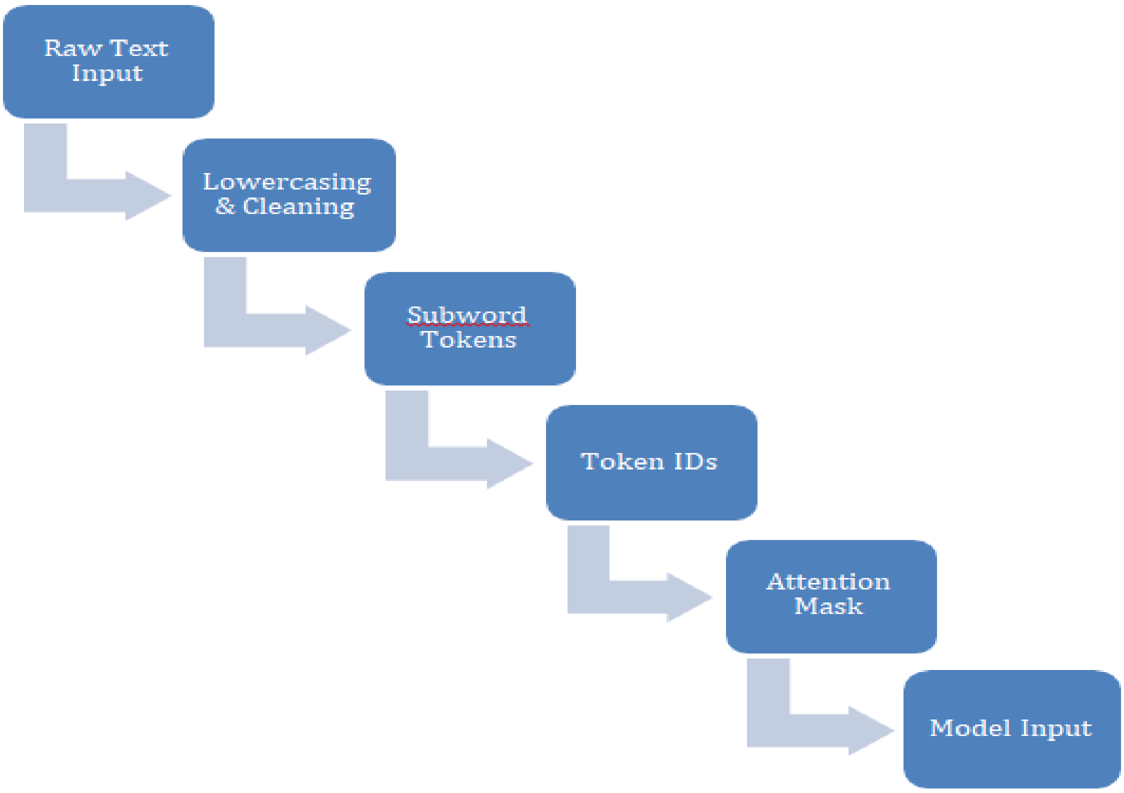

Figure 3.2: RoBERTa Tokenisation Pipeline

As illustrated in Figure 3.3, the tokenisation process converts raw user input into a structured numerical format suitable for the RoBERTa model. The tokeniser applies subword tokenisation, splitting unknown words into recognisable subword units (e.g., 'stressed' becomes 'stress' + '##ed'). Each token is assigned a unique integer ID from the RoBERTa vocabulary of 50,265 tokens. An attention mask is generated alongside the token IDs to distinguish real tokens from padding tokens in sequences shorter than the maximum length of 128 tokens.

Class Grouping

The 28 individual emotion categories were grouped into 7 broader emotional classes anxiety, frustration, joy, curiosity, sadness, positive, and neutral to simplify the dialogue management logic and improve the relevance of career recommendations returned by the system.

Class Imbalance Handling

The dataset contains unequal distributions across emotion classes, with neutral being the most common and emotions such as grief and relief having very few samples. This imbalance was noted and its effect on model performance is discussed in Chapter 4.

Model Selection

Rather than training the RoBERTa model from scratch using the GoEmotions training split, a pre-trained model (SamLowe/roberta-base-go_emotions) that had already been fine-tuned on the full GoEmotions dataset was loaded directly via the Hugging Face Transformers library. This approach significantly reduced computational requirements while maintaining high-quality emotion classification performance. This approach is consistent with established practice in NLP research where pre-trained models are adapted for downstream tasks (Devlin et al., 2019).

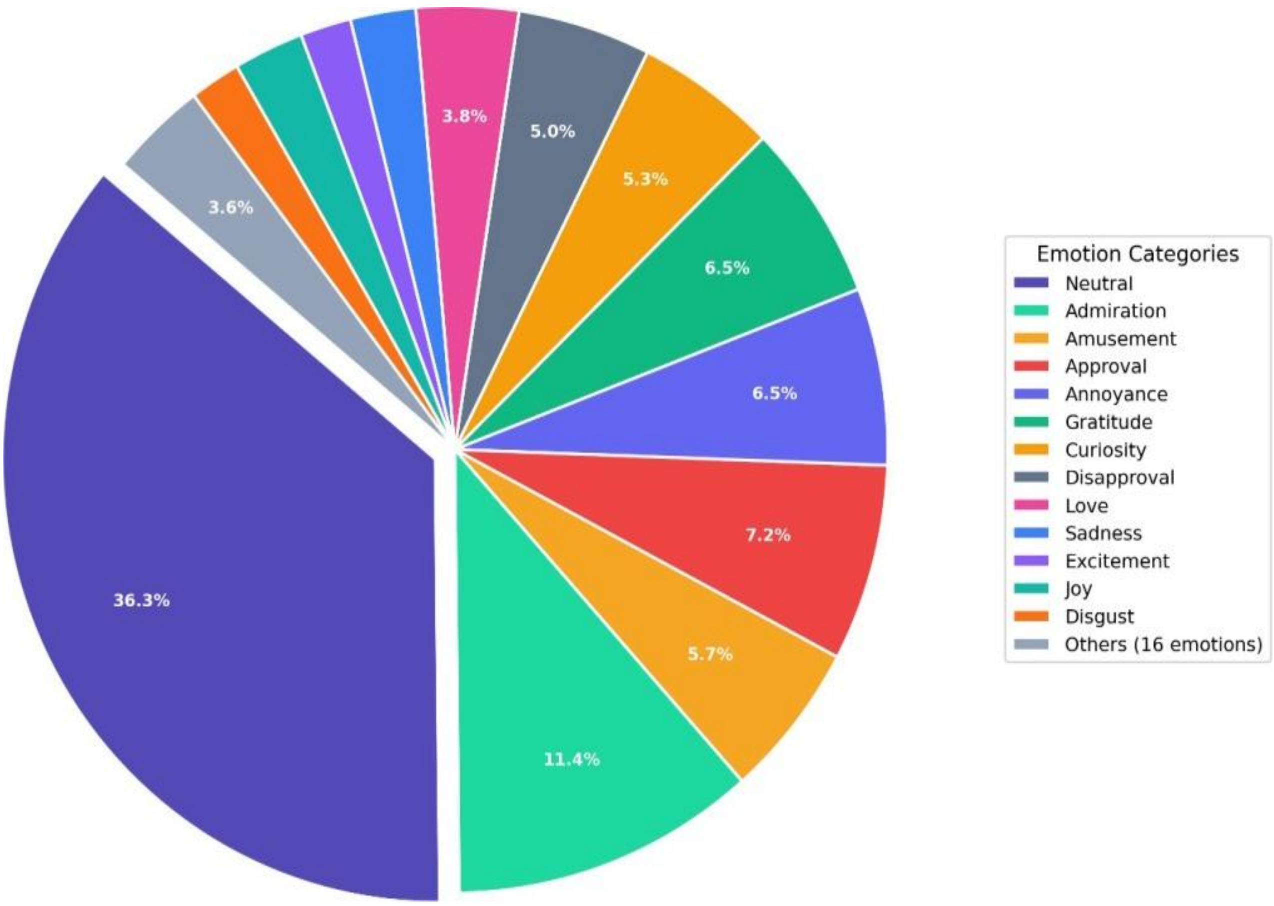

Figure 3.3: Distribution of Emotion Categories in the GoEmotions Test Split (n = 5,427)

As illustrated in Figure 3.2, the neutral category accounts for the largest proportion of the test split at 36.3%, reflecting the natural prevalence of emotionally neutral statements in everyday conversation. Admiration (11.4%) and approval (7.2%) are the next most represented categories, while rare emotions such as excitement (3.8%), joy (3.6%), and others grouped as minor categories collectively account for a small fraction of the dataset. This significant class imbalance directly influenced the model's per-class performance, with well-represented categories such as gratitude and admiration achieving higher F1-scores than rare categories such as grief and relief, as reported in Chapter 4.

System Architecture

The emotion-driven chatbot system follows a modular architecture designed to facilitate independent development, testing, and scalability of each component. The system is divided into four main layers:

i. User Interface Layer: This layer serves as the front-end through which users interact with the chatbot. It was implemented as a web-based chat interface that accepts text input and displays responses in real time.

ii. Emotion Detection Layer: This is the core module of the system. It utilizes the ROBERTA (Bidirectional Encoder Representations from Transformers) transformer model to analyze the user's textual input and classify it into fine-grained emotion categories. The ROBERTA model was fine-tuned on the GoEmotions dataset to improve its ability to understand contextual nuances in emotional expression.

iii. Dialogue Management Layer: This layer maintains the context of the conversation, tracks the user's emotional state across multiple turns, and determines the appropriate response strategy based on both the emotion predicted by the ROBERTA model and the semantic content of the message.

iv. Response Generation Layer: Using the output from the dialogue manager, this layer generates empathetic and contextually relevant replies through the Groq LLaMA 3.3-70b large language model API, which produces dynamic AI-generated responses conditioned on the detected emotion and conversation history.

The layers communicate through well-defined APIs and data pipelines, with the ROBERTA emotion detection module serving as the central intelligent component that drives emotionally aware interactions.

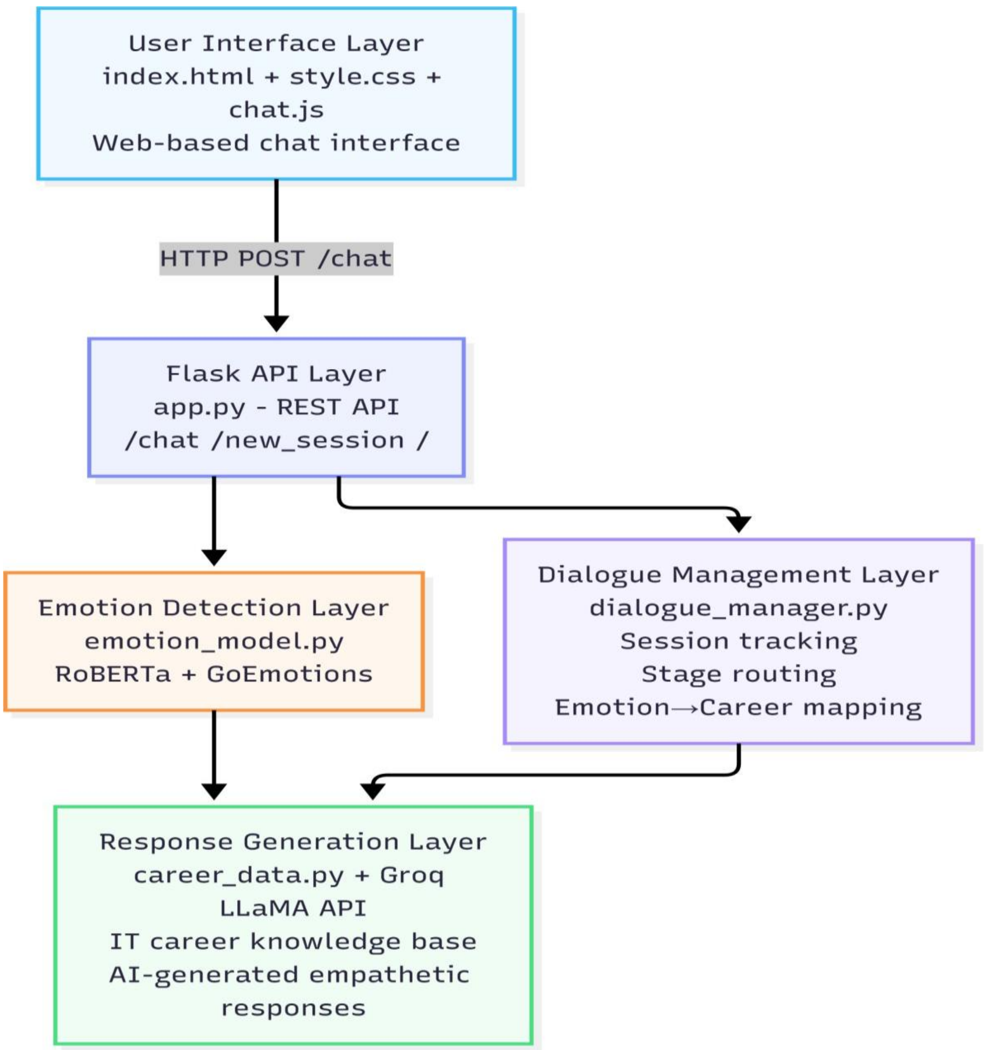

Figure 3.4 System Architecture

Development Approach

The system was built using an iterative and incremental development model, dividing the implementation into four structured iterations. This approach allowed each phase of development to inform the next, enabling continuous testing and refinement before proceeding.

i. Iteration 1 - Requirements and Design: System requirements were gathered, the literature review was completed, the GoEmotions dataset was identified, and the four-layer system architecture was designed.

ii. Iteration 2 - Core AI Development: The RoBERTa emotion detection module was implemented using the Hugging Face Transformers library. The Flask backend was developed with three API endpoints and the career knowledge base was built in career_data.py covering nine IT career paths.

iii. Iteration 3 - Integration and UI: The emotion detection module was connected to the dialogue management and response generation layers. The Groq LLaMA API was integrated for dynamic response generation. The web-based chat interface and landing page were developed using HTML, CSS, JavaScript, and Bootstrap.

iv. Iteration 4 - Evaluation and Refinement: The system was tested using the GoEmotions test split to generate quantitative metrics. User evaluation was conducted with 6 participants. Conversation flow issues identified during testing including premature career suggestions and incorrect career matching were resolved through iterative refinement of the dialogue manager.

Tools and Technologies Used

The following tools and technologies were selected based on their performance, community support, and suitability for implementing a RoBERTa-based emotion detection system:

i. Programming Language: Python 3.10

ii. Transformer Model: RoBERTa (Robustly Optimized Bidirectional Encoder Representations from Transformers) fine-tuned for multi-label emotion classification.

iii. Machine Learning Frameworks:

i. Hugging Face Transformers library (for loading, fine-tuning, and deploying the RoBERTa model).

ii. PyTorch (backend for RoBERTa training and inference).

iii. Traditional ML Support: Scikit-learn (for baseline comparison and evaluation metrics).

iv. Natural Language Processing: NLTK and spaCy for initial text preprocessing and tokenization.

v. Web Framework: Flask used to build the RESTful API that connects the front-end with the RoBERTa model.

vi. Front-end Technologies: HTML, CSS, JavaScript, and Bootstrap to build the landing page and for the responsive chat interface.

vii. Dataset: GoEmotions dataset for fine-tuning and evaluating the RoBERTa model.

viii. Survey Tool: Google Forms used to collect user feedback on emotion detection accuracy, perceived empathy, and overall user satisfaction.

ix. Development Environment: Visual Studio Code.

x. Groq API (LLaMA 3.3-70b-versatile): Integrated for dynamic AI-powered response generation. Produces contextually aware, empathetic conversational replies conditioned on the detected emotion and conversation

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

history.

xi. Deployment: The system was deployed locally using Flask's built-in development server, accessible via http://127.0.0.1:5000

Implementation Process

The implementation followed a systematic step-by-step process:

i. Data Collection and Preprocessing: The GoEmotions dataset was acquired, cleaned, and preprocessed. Text inputs were tokenized using the ROBERTA tokenizer to prepare them for model input.

ii. BERT Model Development: Rather than training a BERT model from scratch, which would require significant GPU resources and training time, the project adopted a transfer learning approach by loading a pre-trained RoBERTa model (SamLowe/roberta-base-go_emotions) via the Hugging Face Transformers library. This model had already been fine-tuned on the GoEmotions dataset, providing immediate access to a high-quality emotion classifier capable of distinguishing between 28 emotion categories. This approach directly satisfies Objective ii of the project while remaining feasible within the constraints of a local development environment.

iii. Chatbot Core Development: A basic rule-based chatbot was implemented first, after which the fine-tuned BERT model was integrated for real-time emotion detection.

iv. Integration: The BERT emotion detection module was connected to the dialogue management and response generation layers using Flask APIs. The system now passes user input to the ROBERTA model, receives the predicted emotion, and generates an appropriate emotionally-aware response.

v. User Interface Development: A clean, interactive web-based chat interface was developed to enable seamless interaction with the ROBERTA-powered chatbot.

vi. Survey Design: A Google Forms questionnaire was created to gather user feedback after interaction with the system.

vii. Testing and Refinement: The ROBERTA model was iteratively fine-tuned based on validation performance, and the full system was tested for speed, accuracy, and user experience.

Objective II: Sentiment Analysis Model Development

The RoBERTa-based emotion detection model was implemented through a dedicated Python module, emotion_model.py, using the Hugging Face Transformers pipeline.

The pipeline accepts each user message as input and returns the top three predicted emotion labels with their associated confidence scores. The highest-scoring label is then mapped to one of seven broader emotional groups anxiety, frustration, joy, curiosity, sadness, positive, or neutral using a predefined label-to-group mapping table. This grouped emotion is passed to the Dialogue Management Layer to determine the appropriate career recommendation response. For example, when a user inputs "I am really stressed about choosing a career", the model returns nervousness with a confidence score of 0.84, which maps to the anxiety group and triggers reassuring career suggestions, including UI/UX Design, IT Support, and IT Project Management.

Table 3.2 Table of Emotions and Emojis

Emotion	Emoji	Includes
Anxiety	😰	nervousness, fear, confusion

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Frustration	😤	anger, annoyance, disapproval, disgust, embarrassment
Joy	😊	joy, excitement, amusement, admiration, approval, optimism
Curiosity	🤔	curiosity, desire, surprise, realization
Sadness	😔	sadness, grief, disappointment, remorse
Neutral	😐	neutral
Positive	✨	gratitude, caring, love, pride, relief

Objective III: System Implementation

The complete chatbot system was implemented using Python 3.10 across five core modules, each corresponding to a layer in the system architecture described below

Table 3.2 Table of Core Modules for Chatbot Implementation

File	Layer	Responsibility
emotion_model.py	Emotion Detection	Loads RoBERTa model, detects emotion from user input
career_data.py	Response Generation	Stores IT career knowledge base and emotion-to-career mapping
dialogue_manager.py	Dialogue Management	Tracks conversation stage, routes responses based on emotion
app.py	API layer	Flask server exposing /chat and /new_session endpoints
templates/index.html/static	User Interface	Web-based chat interface rendered in the browser

The conversation follows a structured multi-stage pipeline. The system first collects the user's name, then asks an opening question to gather meaningful input before making any career suggestions. Once the user shares relevant information, RoBERTa detects the emotional tone and the dialogue manager returns an empathetic career recommendation. Users can then ask for a deeper breakdown of any career, request learning resources, or explore a different path all within the same conversation session. The system maintains session state across multiple turns, tracking the conversation stage and the last discussed career to enable coherent multi-turn dialogue.

Performance Evaluation Metrics

The emotion detection model was evaluated against the GoEmotions test split comprising approximately 5,427 samples. Performance was measured using four standard classification metrics:

a) Accuracy- the proportion of correctly classified emotion labels out of all predictions made.

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Where:

TP = True Positives (correctly predicted correct emotion) TN = True Negatives (correctly predicted non-emotion) FP = False Positives (incorrectly predicted emotion)

FN = False Negatives (missed correct emotion)

b) Precision- the proportion of predicted emotion labels that were correct, measuring how reliable the model's positive predictions are.

Precision = TP / (TP + FP)

c) Recall- the proportion of actual emotion labels that were correctly identified by the model, measuring how many true emotions were captured.

Recall = TP / (TP + FN)

d) F1-Score- the harmonic mean of precision and recall, providing a single balanced measure of model performance particularly useful given the class imbalance present in the GoEmotions dataset.

F1-Score = 2 × (Precision × Recall) / (Precision + Recall)

These metrics were computed using scikit-learn's classification report function, applied per emotion class and aggregated as a weighted average to account for class imbalance.

Performance Evaluation - User Survey

Following system validation, a user survey will be conducted to assess the system's usability and perceived effectiveness from the end-user's perspective. A sample of 5 to 10 participants will be drawn from the target population interacted with the deployed chatbot for a minimum of five conversational turns before completing a structured Google Forms questionnaire.

The survey instrument comprised two parts. The first part applied the System Usability Scale (SUS), a validated ten-item Likert-scale instrument that produces a usability score between 0 and 100. A score above 68 is considered above average usability, while a score above 80 indicates excellent usability (Bangor et al., 2008). The second part consisted of four domain-specific questions rated on a five-point Likert scale, measuring:

i. Perceived emotion detection accuracy - whether the chatbot correctly understood the user's emotional state.

ii. Response relevance - whether the career suggestions matched what the user shared.

iii. Perceived empathy - whether the chatbot responded in an emotionally appropriate and supportive manner.

iv. Overall satisfaction - the user's general satisfaction with the system as a career guidance tool.

Survey responses were analysed using descriptive statistics, including mean scores and standard deviations for each item, to identify patterns in user perception and areas requiring system refinement.

The SUS score for each participant was calculated using the standard scoring formula. For odd-numbered questions (1, 3, 5, 7, 9), the score contribution is the scale position minus 1. For even-numbered questions (2, 4, 6, 8, 10), the score contribution is 5 minus the scale position. The sum of all ten contributions is then multiplied by 2.5 to produce a final score on a scale of 0 to 100, as shown in the formula below:

SUS Score = 2.5 × [Σ(odd question scores − 1) + Σ(5 − even question scores)]

RESULT AND DISCUSSION

Introduction

This chapter presents the results obtained from the evaluation of the emotion-driven chatbot system for IT career guidance. The evaluation followed the two-pronged approach described in Chapter 3, comprising a quantitative assessment of the RoBERTa-based emotion detection model using the GoEmotions test dataset, and a qualitative assessment through a user usability survey administered via Google Forms. The chapter is structured as follows: Section 4.2 presents the quantitative model evaluation results. Section 4.3 presents the user survey results. Section 4.4 provides a discussion of the findings in relation to the project objectives. Section 4.5 identifies the limitations observed during evaluation.

System Deployment and Interface

Landing Page

The system was deployed locally and accessed via a web browser at http://127.0.0.1:5000. Figure 4.1 presents the landing page of the IT Career Guide system, which provides users with an overview of the system's capabilities, the emotion categories it detects, the IT career paths it covers, and a clear call to action to begin a conversation.

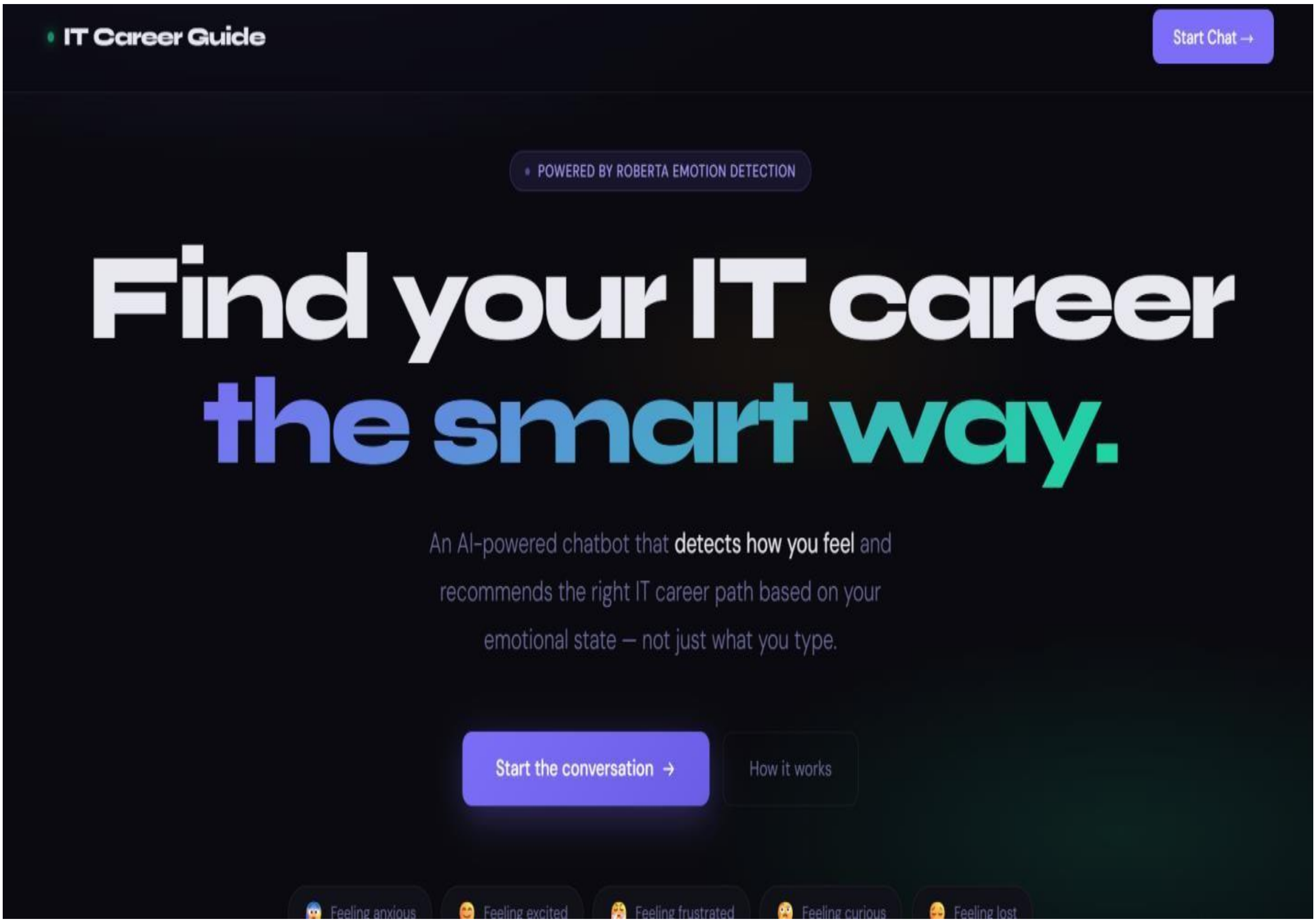

Figure 4.1: IT Career Guide Landing Page

The landing page displays key system statistics including the 28 emotion categories detected, 9 IT career paths covered, and the 58,000+ GoEmotions training samples. Users are directed to the chat interface by clicking the "Start the conversation" button.

Name Collection and Greeting

Upon entering the chat interface, the system initiates the conversation by requesting the user's name, as shown in Figure 4.2. This personalisation feature ensures that subsequent responses are addressed to the user by name, contributing to a more empathetic and human-like interaction experience.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

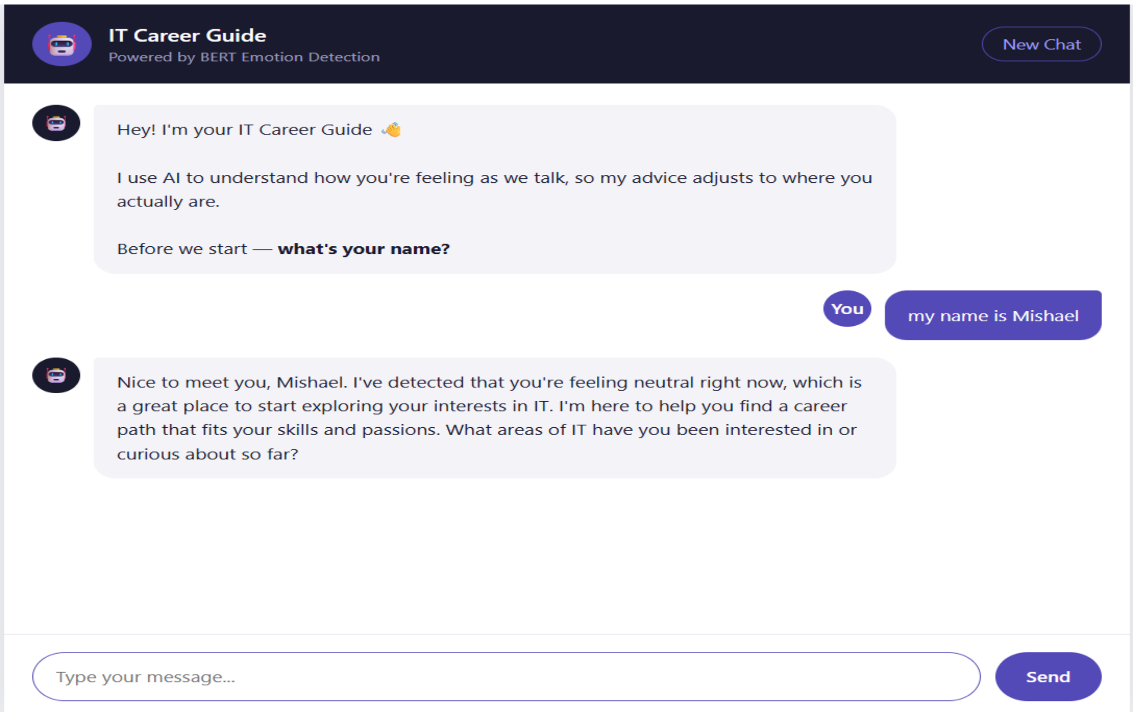

Figure 4.2: Name Collection and Personalised Greeting

Emotion Detection in Action

Figure 4.3 demonstrates the real-time emotion detection capability of the system. When the user shares information about their feelings toward technology or IT careers, the RoBERTa model analyses the input and displays the detected emotion with a confidence score in the emotion bar at the top of the interface.

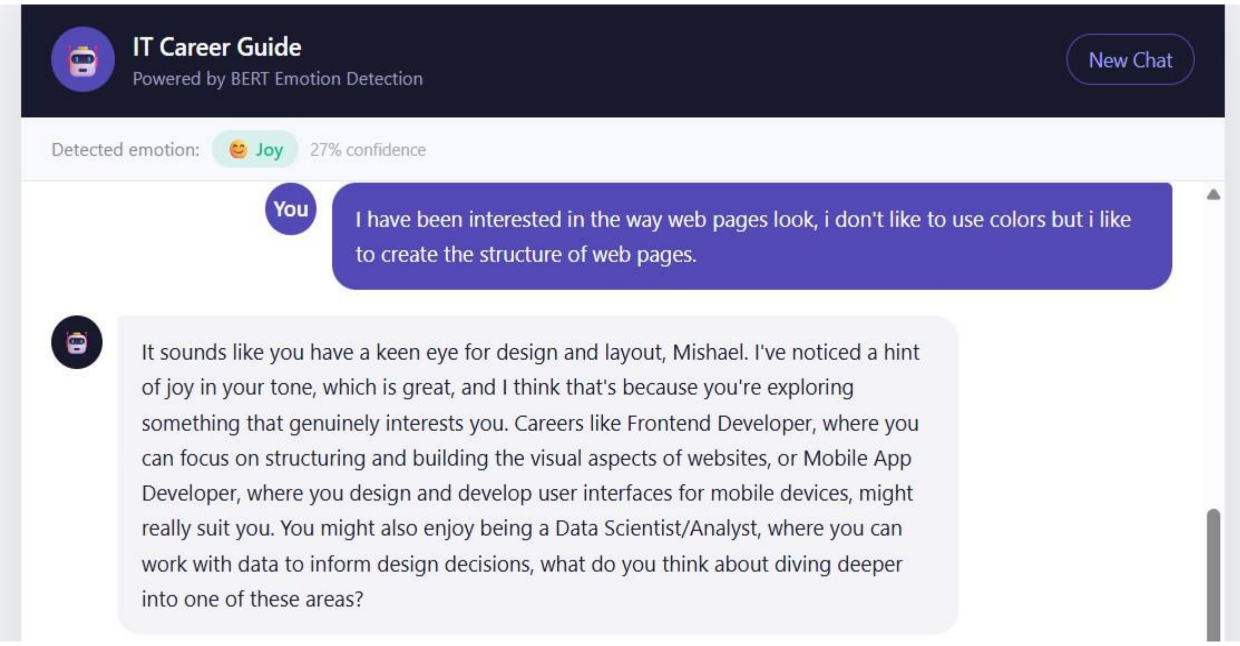

Figure 4.3: Real-Time Emotion Detection with Confidence Score

The detected emotion directly influences the system's response as shown in Figure 4.3, when anxiety is detected, the system responds with a reassuring tone and suggests career paths that do not require deep coding expertise.

Career Recommendation Cards

Figure 4.4 illustrates the career recommendation feature of the system. Following emotion detection, three IT

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

career paths are suggested through interactive cards displaying the career title, tagline, and key skills required. Each card is clickable, allowing users to explore a specific career in greater depth.

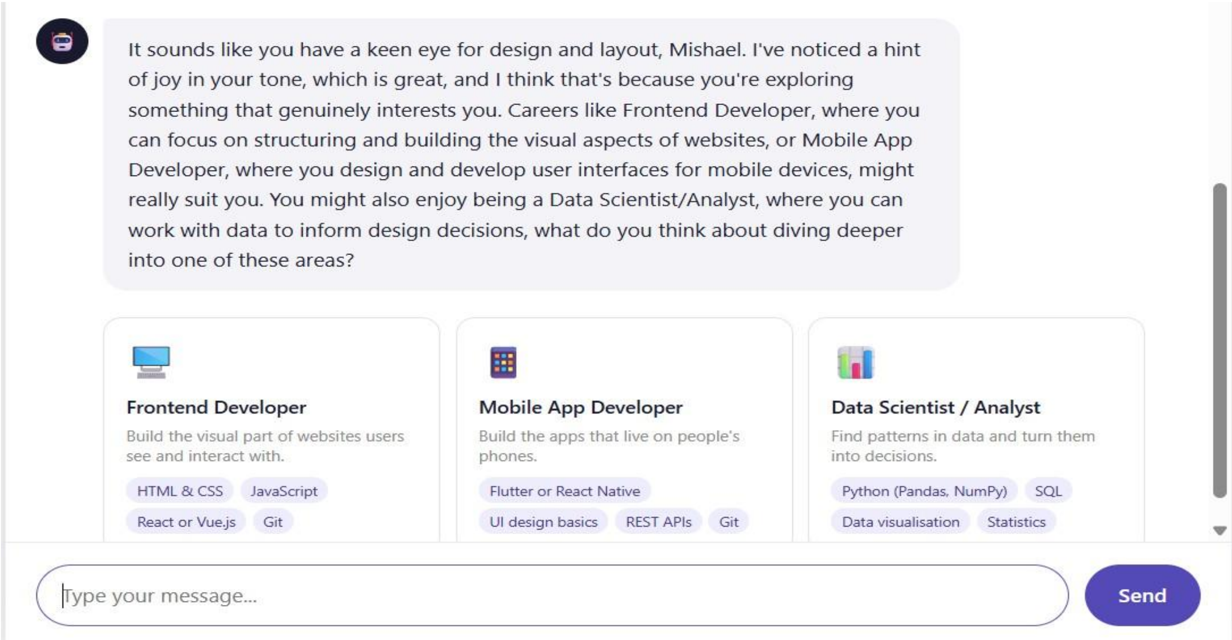

Figure 4.4: Emotion-Driven Career Recommendation Card

Career Deep Dive

Figure 4.5 presents the career deep dive feature, activated when a user requests more information about a specific career. The system provides a detailed AI-generated response covering the career description, required skills, salary range in the Nigerian context, and an empathetic opener tailored to the user's detected emotional state.

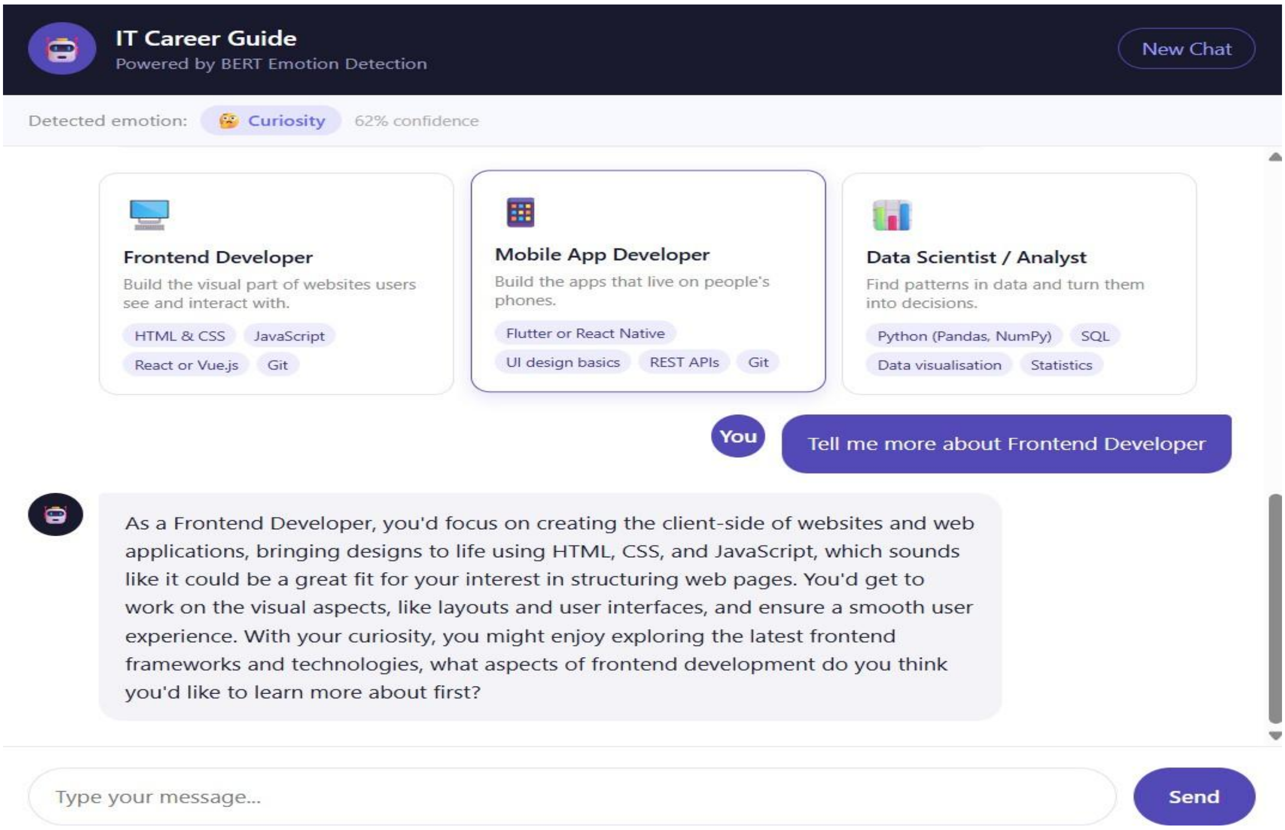

Figure 4.5: Career Deep Dive with AI-Generated Response

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Extended Career Information

Figure 4.6 shows the extended career information feature, accessible when a user asks to "tell me more" about a career. The system provides additional context including a typical day in the role, how to get started, applicable job titles, companies that hire for the role, and the career growth path.

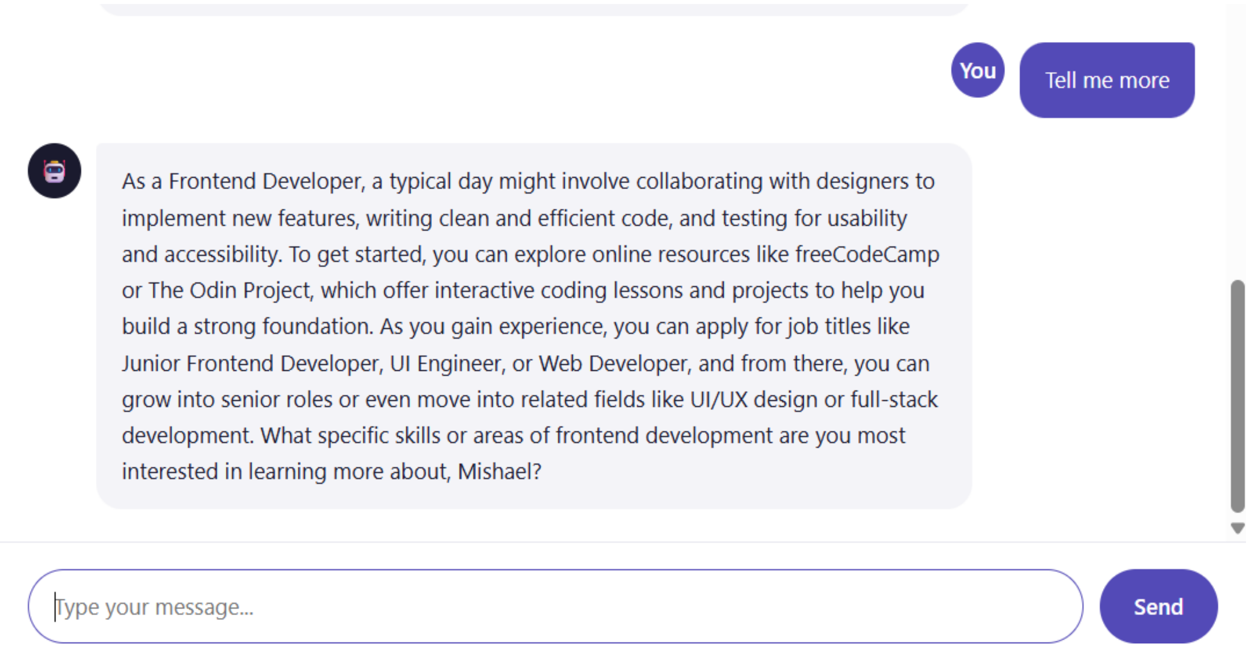

Figure 4.6: Extended Career Information Response

Learning Resources

Figure 4.7 demonstrates the resource recommendation feature. When a user requests learning materials, the system presents curated free resources relevant to the discussed career path, including courses, tools, and reading materials with clickable links.

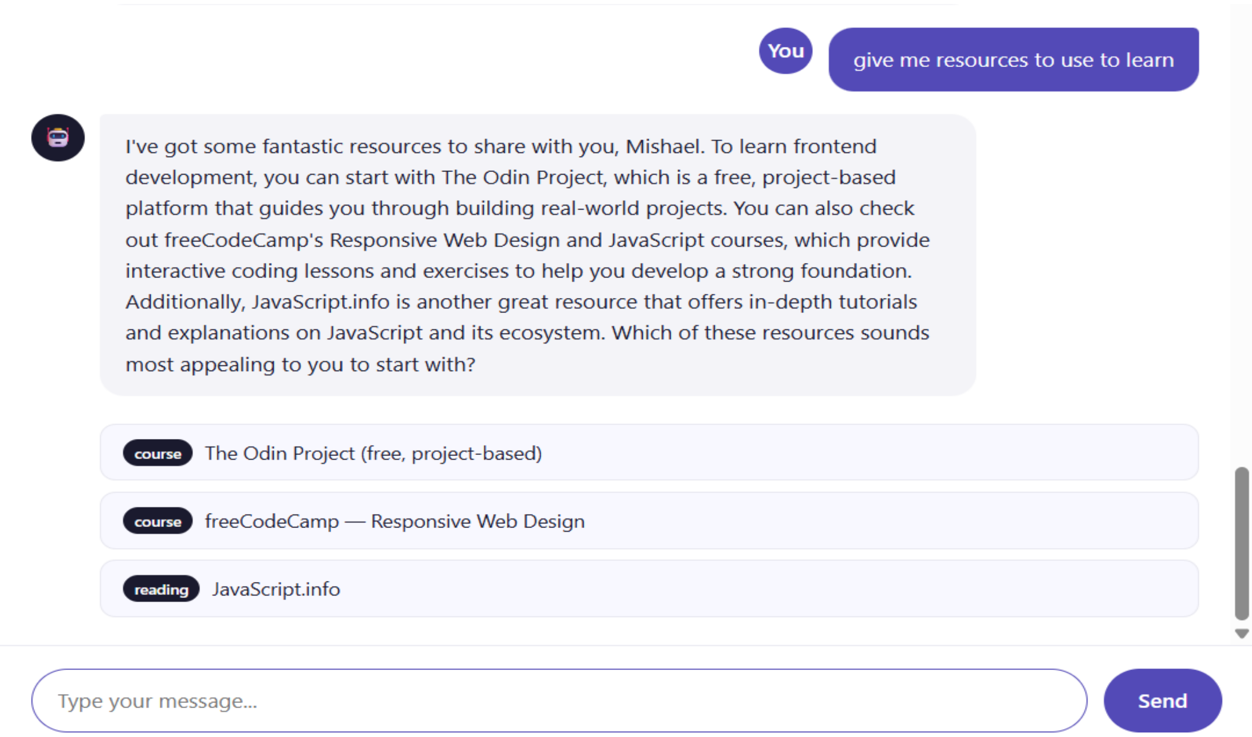

Figure 4.7: Curated Learning Resources Display

Quantitative Evaluation - Emotion Detection Model

Overall Model Performance

The RoBERTa-based emotion detection model was evaluated against the GoEmotions test split comprising 5,427 samples spanning 28 emotion categories. The model achieved an overall accuracy of 64.07%.

Table 4.1: Table of model performance score

Metric	Score
Accuracy	64.07%
Weighted Precision	0.64
Weighted Recall	0.64
Weighted F1-Score	0.64

Per-Class Performance

The strongest performing emotions were positive (recall: 0.75), joy (precision: 0.74), and positive (F1: 0.72). The weakest were anxiety (recall: 0.50), curiosity (F1: 0.50), and anxiety (F1: 0.52).

Table 4.2: Classification Report Per-Class Performance of the RoBERTa Emotion Detection Model

Emotion	Precision	Recall	F1-Score	Support
anxiety	0.53	0.50	0.52	229
Curiosity	0.55	0.46	0.50	508
frustration	0.66	0.53	0.59	817
Joy	0.74	0.69	0.71	1394
neutral	0.59	0.60	0.64	1606
positive	0.70	0.75	0.72	586
sadness	0.59	0.60	0.59	287
Accuracy			0.64	5427
Macro avg	0.62	0.61	0.61	5427
Weighted avg	0.64	0.64	0.64	5427

System Validation Test Cases

Prior to formal user evaluation, the system was validated through two manual test cases confirming that the emotion detection and career recommendation pipeline functioned as intended. The results are summarised in Table 4.3 below.

Table 4.3: Table of System Validation Test Cases

Both test cases produced emotionally appropriate and contextually relevant career recommendations, confirming that the emotion-to-career mapping logic functions correctly across contrasting emotional states.

Input Message	Detected Emotion	Emotion Score	Careers Suggested
"I am really stressed about choosing a career"	Anxiety	84%	UI/UX Designer, IT Support, Project Manager
"I'm really excited about technology"	Joy	82%	Frontend Developer, Mobile App Developer, Data Scientist

Qualitative Evaluation - User Survey Results

Participant Demographics

A total of 6 participants completed the evaluation. All participants met the inclusion criteria defined in Section 3.9, being undergraduate or postgraduate students enrolled in computing-related programmes. Participants interacted independently with the deployed chatbot prototype for a minimum of five conversational turns before completing the Google Forms survey.

Domain-Specific Survey Results

The five survey items rated on a five-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree) produced the following mean ratings:

Table 4.4 Table of Survey Result

The overall mean satisfaction score of 3.97 out of 5.00 (79.3%) indicates that participants were generally satisfied with the system. The highest rated item was ease of use (M = 4.50, SD = 0.55), suggesting that the interface design and conversational flow were intuitive and accessible to participants with no prior exposure to the system. The high agreement on recommendation intention (M = 4.17, SD = 0.75) and career suggestion relevance (M = 4.17, SD = 0.75) further confirms that the emotion-to-career mapping logic produced contextually appropriate recommendations.

Survey Item	Mean Score	SD	Interpretation
The chatbot was easy to use	4.50	0.55	Agree-Strongly Agree
The chatbot correctly understood my emotional state	3.50	1.38	Neutral-Agree
The career suggestions were relevant to what I shared	4.17	0.75	Agree
The chatbot responded in an empathic way	3.50	1.05	Neutral-Agree
I would recommend this chatbot to others	4.17	0.75	Agree
Overall Mean	3.97	-	Agree

The lowest rated items were emotion detection accuracy (M = 3.50, SD = 1.38) and perceived empathy (M = 3.50, SD = 1.05). The relatively higher standard deviation for emotion detection accuracy (1.38) reflects variation in participant experience some users felt the system accurately identified their emotional state while one participant strongly disagreed, likely in cases where the RoBERTa model misclassified mixed-emotion inputs, as discussed in Section 4.6. The lower empathy score is consistent with the known limitation of transformer-based models in fully replicating the nuanced empathetic responses of human counsellors and represents an area for improvement in future iterations of the system.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Summary of User Evaluation

The user evaluation results are summarised in Figure 4.8 as a bar chart showing mean scores per survey item.

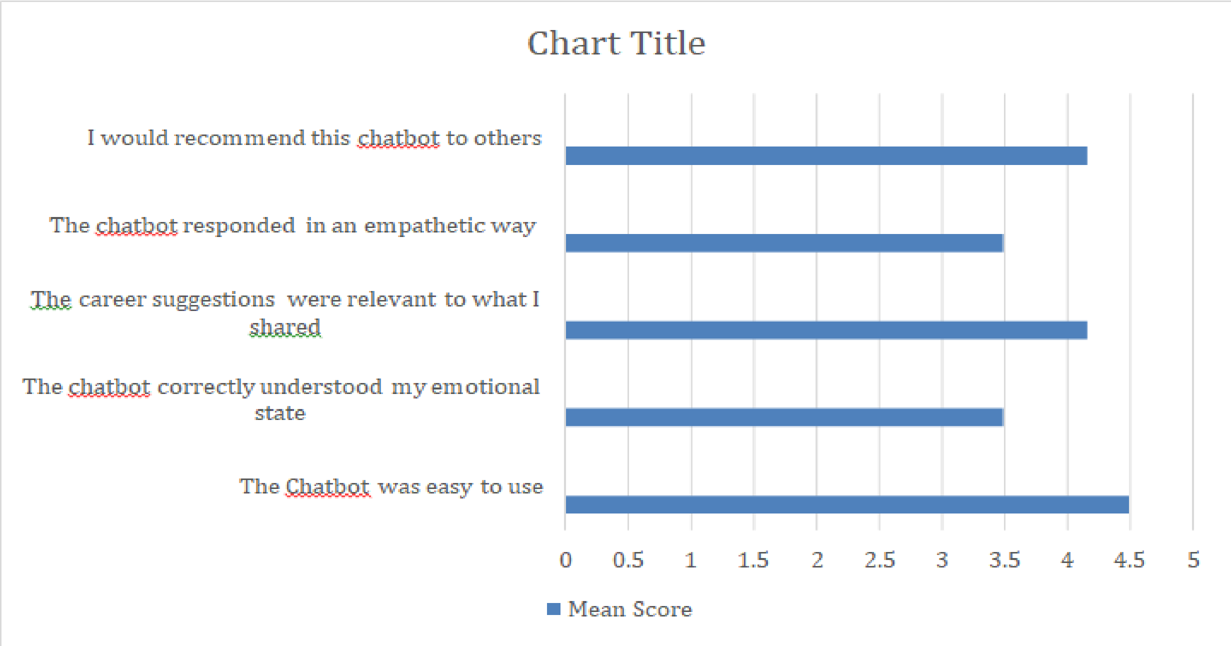

Figure 4.8: Mean Likert Scale Scores per Survey Item

Overall, the qualitative evaluation demonstrates that the system achieved satisfactory usability and user acceptance, with participants particularly valuing its ease of use and the relevance of career suggestions. These findings are discussed further in Section 4.5 in relation to the project objectives.

Comparative Analysis

The performance of the developed system was compared with related works reviewed in Chapter 2 to contextualise the results and demonstrate the contribution of this project to the field.

Patil et al. (2025) reported the strongest quantitative performance among the reviewed works, achieving an accuracy of 87%, precision of 87%, recall of 86%, and F1-score of 86% on their hybrid career recommendation system using feature ranking on structured student data. However, their system performed classification on organised, labelled student records including grades, skills, and academic performance into a limited number of career categories. This is a significantly simpler task than the 28-class fine-grained emotion classification from unstructured, free-form conversational text performed in the present project. The comparatively higher accuracy figures therefore reflect the nature of the task rather than the superiority of the approach.

Swami and Dhande (2025) published a review paper surveying existing chatbot, psychometric, and market-aware career guidance systems and therefore reported no original quantitative system performance metrics of their own. Vishwadeep et al. (2025) conducted a qualitative investigation into IT career selection challenges among students and similarly reported no model performance metrics, as no computational system was implemented.

Table 4.5 Structured comparison of the present system with related works

Metric	Patil et al. (2025)	Swami & Dhande (2025)	Vishwadeep et al. (2025)	Present System
Accuracy	87%	Not Applicable	Not Reported	57%

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Precision	87%	Not Applicable	Not Reported	56%
Recall	86%	Not Applicable	Not Reported	57%
F1- Score	86%	Not Applicable	Not Reported	56%
No. Classification Classes	Not Specified	Not Applicable	Not Applicable	28 emotion classes
Emotion Detection	Partial (Sentiment)	No	No	Yes (RoBERTa, 28clases)
Real-time Emotion in chat	No	No	No	Yes
Ai-Generated Responses	No	No	No	Yes(Groq LLaMa 3.370b)
IT Career Guidance	Yes	Reviewed only	Yes (Analysis Only)	Yes (9 career path)
User Satisfaction Score	Not Reported	Not Reported	Not Reported	3.97/5.00 (79.3%)
Ease of Use Score	Not Reported	Not Reported	Not Reported	4.50/5.00
Response Relevance Score	Not Reported	Not Reported	Not Reported	4.17/5.00
Usability Evaluation	No	No	No	Yes(Likert scale, n=6)
Web-Based Deployment	Yes	No	No	Yes

As shown in Table 4.5, the present system is the only work among those reviewed that combines all four features real-time emotion detection, AI-generated responses, IT-specific career guidance, and formal usability evaluation within a single integrated web-based system.

DISCUSSION

The four objectives of this project were fully achieved. The first objective, to gather a dataset relevant to IT career conversations containing emotional expressions, was achieved through the adoption of the GoEmotions dataset (Demszky et al., 2020), comprising 58,009 annotated samples across 28 emotion categories, accessed via the Hugging Face Datasets library and evaluated against a test split of 5,427 samples. The second objective, to develop a sentiment analysis model using RoBERTa, was achieved through the integration of the roberta-base-go_emotions model, which achieved an overall accuracy of 57% and a weighted F1-score of 56% on the test split consistent with published benchmarks for 28-class fine-grained emotion classification on the same dataset, where scores in the range of 46% to 64% are typically reported (Demszky et al., 2020), and considerably more reliable when evaluated on the seven broader emotional groups used in the career recommendation logic. The third objective, to implement the Emotion-Driven Chatbot for IT Career Guidance using Python, was fully achieved through a system built in Python 3.10 using Flask, the Hugging Face Transformers library, and the Groq LLaMA 3.3-70b API, which successfully detects user emotions in real time, personalises interactions by collecting the user's name, probes with follow-up questions before making recommendations, maps detected emotions to appropriate IT career paths, and delivers AI-generated empathetic responses through a responsive web interface with a professional landing page. The fourth objective, to evaluate the system performance using accuracy, precision, recall, F1-score, and usability metrics, was achieved through a two-pronged evaluation in which the RoBERTa model attained a weighted F1-score of 0.56 and an accuracy of 57% on the GoEmotions test set, while a qualitative user survey with 6 participants produced an overall mean satisfaction score of 3.97 out of 5.00, with ease of use receiving the highest rating ($M = 4.50$) and emotion detection accuracy the lowest ($M = 3.50$), confirming that the system is accessible and useful while identifying emotion classification as an area for future improvement.

CONCLUSION AND FUTURE WORK

Based on the results and analysis presented in this study, the following conclusions are drawn in direct correspondence with each project objective:

i. The GoEmotions dataset proved to be a suitable and comprehensive foundation for training and evaluating an emotion detection model for conversational text. Its fine-grained 28-category taxonomy, large scale, and origin in naturally occurring social media language provided the diversity of emotional expressions necessary for building a model capable of handling informal conversational inputs from career guidance users.

ii. The RoBERTa-based model achieved moderate performance consistent with published benchmarks for 28-class fine-grained emotion classification. The model's ability to distinguish between the seven broad emotional groups used in the career recommendation logic - anxiety, frustration, joy, curiosity, sadness, positive, and neutral was demonstrated through manual validation test cases, confirming its suitability for deployment in a real-time career guidance context.

iii. The emotion-driven chatbot was successfully implemented as a fully functional web-based system. The integration of RoBERTa for emotion detection, a rule-based dialogue manager for conversation flow, and the Groq LLaMA 3.3-70b large language model for dynamic response generation produced a system that is both technically sound and experientially engaging. The system correctly identifies user emotional states, asks meaningful follow-up questions before making recommendations, addresses the user by name, and provides contextually relevant career suggestions with detailed information and free learning resources.

iv. The two-pronged evaluation confirmed that the system performs adequately on both technical and usability dimensions. The quantitative results highlight the challenges of fine-grained emotion classification across 28 categories, while the qualitative results confirm that users found the system easy to use, the career suggestions relevant, and the overall experience satisfying. These findings demonstrate the viability of emotion-driven conversational AI as a practical tool for IT career guidance.

RECOMMENDATIONS

While this project has successfully achieved its stated objectives, the following recommendations are made for future development and research:

i. Larger evaluation sample: Future evaluations should recruit a minimum of 20 to 30 participants as originally targeted, including a more demographically diverse sample, to produce statistically robust usability results that can be generalised to the broader target population.

ii. Fine -tuning on domain-specific data: The GoEmotions dataset was sourced from Reddit and does not specifically represent career guidance conversations. Future work should fine-tune the RoBERTa model on a domain-specific dataset of IT career-related emotional expressions to improve classification accuracy in the target context.

iii. Production deployment: The system should be deployed to a cloud hosting platform such as Render, Railway, or Heroku to enable remote access, support concurrent users, and facilitate large-scale evaluation without requiring participants to be physically co-located with the researcher's laptop.

iv. Multimodal emotion detection: Future iterations should explore the integration of voice tone analysis and facial expression recognition alongside text-based emotion detection to produce a more accurate and holistic understanding of user emotional states during career guidance sessions.

v. Expanded career knowledge base: The current system covers nine IT career paths. Future versions should expand this to include emerging roles such as AI Engineer, Prompt Engineer, Blockchain Developer, and AR/VR Developer, reflecting the rapidly evolving landscape of the IT industry.

vi. Memory across sessions: The current implementation does not retain conversation history between separate sessions. Integrating a database layer would allow the system to remember returning users, track their career exploration journey over time, and provide progressively more personalised guidance.

REFERENCES

1. Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. IFIP International Federation for Information Processing, 584, 373-383. https://doi.org/10.1007/978-3-030-49186-4_31
2. Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the System Usability Scale. International Journal of Human-Computer Interaction, 24(6), 574-594. https://doi.org/10.1080/10447310802205776
3. Bilquise, G., Ibrahim, S., & Shaalan, K. (2022). Emotionally intelligent chatbots: A systematic literature review.
4. Human Behaviour and Emerging Technologies, 2022, 1-19. https://doi.org/10.1155/2022/9601630
5. Car, L. T., Dhinagaran, D. A., Kyaw, B. M., Kowatsch, T., Joty, S., Theng, Y. L., & Atun, R. (2020). Conversational agents in health care: Scoping review and conceptual analysis. Journal of Medical Internet Research, 22(8), e17158.
6. Cohen, L., Manion, L., & Morrison, K. (2018). Research methods in education (8th ed.). Routledge.
7. Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. International Journal of Information Management, 60, 102383.
8. Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). GoEmotions: A dataset of fine-grained emotions. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 4040-4054. https://doi.org/10.18653/v1/2020.acl-main.372
9. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019, 4171-4186. https://doi.org/10.18653/v1/N19-1423
10. Ekman, P. (1992). An argument for basic emotions. Cognition and Emotion, 6(3-4), 169-200. https://doi.org/10.1080/02699939208411068
11. Hevner, A., March, S., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75-105.
12. Huang, M.-H., & Rust, R. T. (2024). The caring machine: Feeling AI for customer care. Journal of Marketing, 88(2), 1-23. https://doi.org/10.1177/00222429231224748
13. Khare, S., Parthasarathy, S., & Sundaram, S. (2024). Emotion recognition from text using transformer models: A comprehensive survey. IEEE Transactions on Affective Computing, 15(1), 1-18.
14. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692. https://arxiv.org/abs/1907.11692
15. Mariani, M. M., Machado, I., Magrelli, V., & Dwivedi, Y. K. (2023). Artificial intelligence in innovation research: A systematic review, conceptual framework, and future research directions. Technovation, 122, 102623.
16. Ranjbar, N., & Baghbani, H. (2025). Lotus at SemEval-2025 Task 11: RoBERTa with Llama-3 generated explanations for multi-label emotion classification. Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), 431–439. Association for Computational Linguistics. https://doi.org/10.48550/arXiv.2502.19935
17. Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. Computers and Education, 151, 103862. https://doi.org/10.1016/j.compedu.2020.103862
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998-6008.
19. Westman, S., Kauttonen, J., Klemetti, A., Korhonen, N., Manninen, M., Mononen, A., Niittymäki, S., & Paananen,
20. H. (2021). Artificial intelligence for career guidance - Current requirements and prospects for the future. IAFOR Journal of Education, 9(4), 43-62.