

Integrating Question Answering Technology into E-Health Education: A Computer-Supported Learning Approach

Wiwin Suwarningsih*

Research Center for Data and Information Sciences, National Research and Innovation Agency, Bandung, Indonesia

Email: wiwi005@brin.go.id

ORCID iD: <https://orcid.org/0000-0003-0155-6329>

*Corresponding author

Nuryani

Research Center for Data and Information Sciences, National Research and Innovation Agency, Bandung, Indonesia

Email: nury012@brin.go.id

ORCID iD: <https://orcid.org/0000-0003-3159-4780>

Endang Suryawati

Research Center for Artificial Intelligence and Cyber Security, National Research and Innovation Agency, Bandung, Indonesia

Email: enda029@brin.go.id

ORCID iD: <https://orcid.org/0000-0001-9254-095X>

Received: November 18, 2025; Revised: January 10, 2026; Accepted: February 24, 2026; Published: August 8, 2026

Abstract: This study examines the integration of Question Answering (QA) technology into e-health education using a scaffolded method to address stunting prevention and understanding of early childhood nutrition in Indonesia. The research developed a computer-supported learning platform that combines adaptive QA systems with structured scaffolding techniques to support maternal knowledge development and informed nutrition practices. Data were collected from Indonesian stunting education programs and nutrition improvement initiatives targeting children under five years old. The platform implemented progressive learning modules that provided real-time responses to user queries while offering contextual support through scaffolded feedback mechanisms. Knowledge improvement was evaluated using a pre-test and post-test design administered before and after system interaction, with a follow-up assessment conducted four weeks later to measure retention. Results demonstrated improvements in knowledge retention scores (mean increase of 42%) and understanding of nutrition guidelines among participating mothers. The QA system helped clarify common misconceptions about complementary feeding practices while maintaining engagement through personalized learning paths. These findings indicate that scaffolded QA technology has potential as a supportive tool for e-health education in resource-limited settings, although broader implementation and long-term impact require further investigation.

Index Terms: E-Health, Nutrition Education, QA System, Stunting.

1. Introduction

E-health education has emerged as a transformative approach in public health, enabling scalable and accessible learning for caregivers in both urban and rural communities [1,2]. In Indonesia, stunting affected approximately 24% of children under five according to the Indonesian Nutritional Status Survey (SSGI) conducted by the Ministry of Health [3]. This high prevalence highlights the need for effective educational strategies, and digital platforms offer a valuable opportunity to improve maternal knowledge of early childhood nutrition and feeding practices [4,5]. However, many existing e-health education systems face significant challenges [6,7], including limited interactivity [8,9], low user engagement [10,11], and insufficient support for learners with varying levels of digital and health literacy [12,13]. Traditional e-learning models often deliver static content without adapting to individual learning needs, resulting in

suboptimal knowledge retention and application. Early studies have also shown that caregivers frequently rely on online sources for health information, although the quality, usability, and reliability of such platforms vary considerably, highlighting the need for trustworthy and user-centered digital health systems [10,14].

Despite advancements in digital health, there remains a critical gap in intelligent, responsive systems capable of understanding and answering user-generated questions in real time using natural language processing (NLP) [15,16]. While question answering (QA) technologies have demonstrated success in other domains [17], their application in public health education, particularly for nutrition in low-resource settings remains limited [18]. Few systems are designed to provide evidence-based, culturally relevant responses in a conversational format that supports self-paced, scaffolded learning [19]. This lack of adaptive support restricts the potential of computer-supported learning to guide caregivers through complex nutritional concepts progressively.

To address this gap, this study develops and integrates a QA system into an e-health education platform, specifically targeting stunting prevention and early childhood nutrition in Indonesia. The primary objective is to enhance maternal understanding through a computer-supported learning approach that combines NLP-driven QA with pedagogical scaffolding [20,21]. By enabling users to ask questions and receive tailored, real-time responses, the system supports personalized learning pathways and deeper cognitive engagement.

The methodology draws on educational materials from Indonesian stunting prevention and nutrition improvement programs as the knowledge base. A natural language processing (NLP)-based QA model was implemented to interpret and respond to user inquiries accurately. The system was evaluated through usability testing and pre- and post-learning assessments to measure improvements in knowledge and user satisfaction. Knowledge retention was further assessed through a delayed post-test conducted four weeks after the intervention to examine short-term learning persistence.

To provide a structured evaluation and ensure alignment between the study objectives and the experimental design, this study addresses the following research questions:

RQ1: Does the QA-based learning system improve users' knowledge of stunting and early childhood nutrition compared to conventional learning materials?

RQ2: How do users perceive the system in terms of usability and acceptance?

RQ3: Does the knowledge gained through the system persist over time?

RQ4: Is there a relationship between perceived system usability and users' learning outcomes?

Prior studies suggest that interactive and adaptive learning systems may improve user engagement and learning outcomes compared to static content delivery [22,23].

This study provides empirical evidence on the integration of QA technology within a scaffolded e-health learning environment in the Indonesian context [24-26]. Rather than proposing a generalizable solution, the findings provide an initial evaluation of its feasibility and learning benefits within a specific application domain.

The remainder of this paper is structured as follows: Section 2 presents the related work overview; Section 3 details the methods, including data sources and system design; Section 4 reports results on knowledge improvement and usability; Section 5 discusses implications, limitations, and future directions; and Section 6 concludes with recommendations for implementing QA-enhanced learning in public health programs.

2. Literature Review

2.1. E-Health Education: Concepts and Prior Studies

E-health education refers to the use of digital technologies to improve health literacy, promote preventive behavior, and support healthcare decision-making[1,2]. It integrates information technology, pedagogy, and health sciences to provide accessible, self-paced learning opportunities. In developing countries such as Indonesia, e-health education plays a vital role in addressing challenges like limited healthcare access and low nutrition awareness[27,28]. Earlier research has emphasized that usability, accessibility, and interaction quality are critical factors influencing user engagement and effectiveness in digital health information systems [29,30].

Previous research on e-health and online learning has demonstrated the effectiveness of mobile applications, web-based platforms, and community-based learning systems in improving health knowledge. For example, several studies have reported that mobile health (mHealth) programs enhance mothers' understanding of child nutrition and stunting prevention [31-33]. However, most e-health education systems rely on static content delivery, such as pre-defined modules or quizzes, which limits personalization and interactivity. As a result, many learners receive general information rather than adaptive or context-specific guidance [34,35].

2.2. Artificial Intelligence and Question Answering in Learning

Advancements in Artificial Intelligence (AI) have transformed educational practices by enabling personalized and interactive learning environments [36]. Among these technologies, Question Answering (QA) systems have gained attention for their ability to simulate human-like dialogue and provide direct answers to learner queries. QA-based

systems have been successfully applied in various fields, including language learning [37], science education [38], and computer science [39], to support inquiry-based learning and improve comprehension [40].

In educational contexts, QA systems allow learners to explore content in a self-directed manner by asking questions and receiving contextually relevant explanations [38,39]. Previous studies suggest that such interactive exchanges can support learner engagement, motivation, and knowledge retention, particularly when compared to passive content consumption. Despite these advantages, the application of QA technologies in e-health education remains limited. Only a small number of studies have explored their use for improving public health literacy, especially in domains such as child nutrition and stunting prevention, where culturally appropriate and personalized information is essential [41].

2.3. Research Gap, Contribution, and Novelty

While numerous studies have explored digital and AI-based learning, several research gaps remain [42]. Existing e-health education systems seldom integrate NLP-driven QA mechanisms that adapt to users' specific learning needs [29], [43]. Most prior works also lack empirical evaluation of how such systems affect knowledge improvement and usability among community learners[44]. Furthermore, few studies have examined the role of QA-based systems in promoting health education within developing countries [45,46].

This study addresses these gaps by developing and evaluating a QA-based learning system tailored for stunting prevention and early childhood nutrition education in Indonesia. The system integrates NLP-driven question answering within a scaffolded learning framework to support interactive and context-aware information access. Rather than proposing a broadly generalizable solution, this study provides an initial empirical evaluation of the feasibility, usability, and learning impact of QA-supported e-health education within a specific public health application context. The findings contribute to the ongoing discussion on digital health literacy and the practical integration of AI-based learning tools in resource-constrained environments.

3. Methodology

3.1. Proposed Research Design

The proposed research introduces an AI-driven Question Answering (QA) system designed to enhance digital health education, particularly in the domain of stunting prevention and early childhood nutrition in Indonesia. This framework combines Natural Language Processing (NLP) techniques and computer-supported learning principles to create an adaptive and interactive learning environment. The system enables users primarily mothers, health educators, and students to ask questions in natural language and receive accurate, context-based explanations derived from verified health resources.

The research framework follows an iterative learning cycle, integrating data retrieval, model reasoning, user interaction, and feedback adaptation. This approach not only supports immediate knowledge delivery but also allows the system to evolve dynamically through continuous data refinement and user feedback. The main process flow is illustrated in Fig. 1.

Fig. 1 illustrates the end-to-end workflow of the proposed system. The process starts with user input, where learners submit health-related questions in natural language form. The system then performs NLP parsing, which includes text preprocessing, tokenization, and intent recognition to understand the user's query.

Subsequently, the data retrieval module searches for relevant information from multiple trusted sources such as educational content on stunting and nutrition, Indonesian health guidelines, and previous user interaction logs. The retrieved content is then processed by the Question Answering System, which employs NLP models and AI reasoning to extract or generate the most suitable response.

The response generation stage produces a human-readable answer, which is displayed to the user through the e-learning interface. Following each interaction, the feedback collection component gathers user reactions, satisfaction ratings, and interaction data. These inputs are used in the learning and improvement cycle to refine both the dataset and model parameters.

The framework demonstrates a continuous learning loop, ensuring that the system not only provides accurate and reliable answers but also adapts to user behaviour and evolving educational needs. This architecture supports sustainable improvement in digital health literacy and knowledge dissemination within Indonesia's e-health education ecosystem.

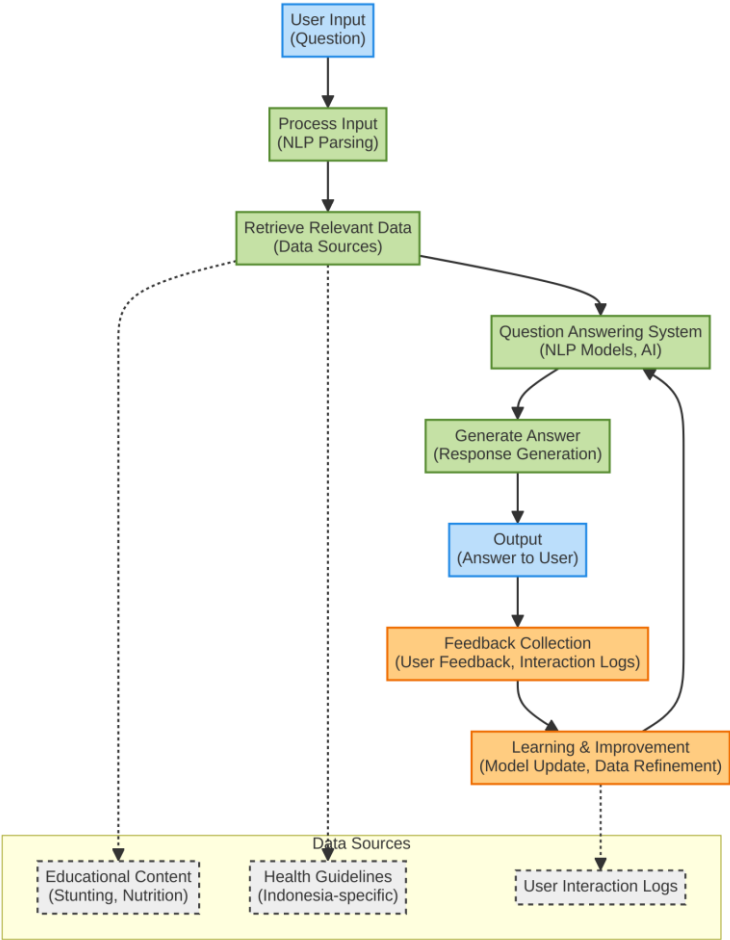


Fig. 1. Proposed System Flow of the QA-Based E-Health Education Framework.

3.2. Materials and Dataset Description

3.2.1. Data Sources and Collection

The knowledge base for this QA system is constructed from authoritative health education sources to ensure accuracy and reliability of information. Primary data sources include educational modules from the Indonesian Ministry of Health, which serve as the primary reference for national context, as well as World Health Organization (WHO) guidelines and UNICEF child nutrition recommendations used as complementary global references to ensure alignment with international health standards, along with existing Indonesian e-learning materials on child nutrition. These sources provide comprehensive coverage of evidence-based information about stunting prevention, child growth monitoring, nutritional interventions, and public health policies.

The selection criteria for source materials emphasized relevance to Indonesian context, accessibility for lay audiences, and alignment with national health priorities. Materials published after 2018 were prioritized to ensure current information, while foundational documents were retained for their enduring value in understanding the epidemiology and prevention strategies for childhood stunting.

Additional publicly accessible sources were also utilized to capture frequently asked questions and practical health communication patterns, including materials from the National Stunting Reduction Program (<https://stunting.go.id>) and publicly available health information platforms such as detikHealth and Halodoc. All collected materials were reviewed to ensure consistency with official health guidelines and evidence-based recommendations.

The construction of question–answer (QA) pairs was conducted through a semi-manual annotation process. The collected materials were first segmented into key informational units representing core concepts related to stunting prevention and child nutrition. Questions were then formulated to reflect natural information needs commonly expressed by parents, caregivers, and community health users. Answers were authored by summarizing, simplifying, and restructuring the corresponding source content into clear and concise Indonesian while preserving the original medical meaning. All QA pairs were written in original form through paraphrasing and content reorganization, and no source text was copied verbatim.

Each QA pair was independently reviewed by two annotators with backgrounds in public health and health communication. Discrepancies between annotators were resolved through discussion and, when necessary, consultation with a licensed healthcare professional. Inter-annotator agreement was assessed using percentage agreement, which reached 91%, indicating high consistency and reliability of the validated content.

The dataset used in this study was compiled from publicly available health education materials. Due to copyright and licensing considerations associated with several source platforms, the compiled QA dataset is not publicly distributed. However, it can be made available for academic research purposes upon reasonable request to the corresponding author.

3.2.2. Dataset Structure and Categories

The finalized dataset consists of 2,847 validated question–answer (QA) pairs, systematically organized into eight thematic categories reflecting key dimensions of *stunting prevention* and *child nutrition*. These categories encompass:

- (1) *Definition and Basic Concepts* (312 entries), covering the fundamental understanding of stunting, growth standards, and child development milestones;
- (2) *Prevention Strategies* (456 entries), detailing primary prevention measures such as balanced nutrition, sanitation practices, and healthcare access;
- (3) *Nutritional Requirements* (398 entries), describing the essential macronutrient and micronutrient needs for children under five;
- (4) *Impact and Consequences* (287 entries), outlining both short- and long-term developmental effects of stunting;
- (5) *Detection and Monitoring* (356 entries), focusing on growth assessment techniques, measurement tools, and early identification indicators;
- (6) *Policy and Programs* (312 entries), documenting government and institutional efforts in national stunting reduction programs;
- (7) *Micronutrition* (426 entries), presenting information on vitamin and mineral supplementation; and
- (8) *Community Interventions* (300 entries), describing community-level initiatives and social support strategies related to child nutrition.

Each QA-pair was validated by certified healthcare professionals to ensure factual accuracy, contextual appropriateness, and adherence to evidence-based health education standards. The questions were formulated using natural Indonesian language patterns drawn from actual communication by parents and healthcare workers, while the answers were curated to maintain scientific validity yet remain understandable to the general public.

Table 1 presents a structured overview of the educational dataset on stunting, comprising five columns: ID, context, question, answer, and category. The dataset integrates key thematic areas such as the causes and prevention of stunting, nutritional requirements for infants and toddlers, and relevant public health policies. Each entry reflects evidence-based educational content emphasizing the interrelated roles of nutrition, health education, sanitation, and micronutrient supplementation in addressing child growth and development challenges. This table exemplifies how textual data were transformed into question–answer pairs for model training and evaluation within the proposed QA-based e-health education framework.

Table 1. Question–Answer Example Pairs in Bahasa Indonesia.

ID	Context	Question	Answer	Category
QA001	Stunting is a growth failure condition in children caused by chronic malnutrition, especially during the first 1,000 days of life. Children with stunting have a lower height than the standard for their age.	<i>Apa penyebab utama stunting pada anak?</i> (in English: What is stunting?)	<i>Kekurangan gizi kronis pada 1000 hari pertama.</i> (In English: Chronic malnutrition during the first 1,000 days of life.)	Definition & Causes
QA002	Adequate nutrition during pregnancy is essential to prevent stunting. Pregnant women should consume balanced foods containing protein, iron, and folic acid.	<i>Mengapa ibu hamil harus makan makanan bergizi seimbang?</i> (in English: How can pregnant women prevent stunting in their babies?)	<i>Gizi seimbang mendukung pertumbuhan janin dan mencegah stunting.</i> (In English: Balanced nutrition supports fetal growth and prevents stunting.)	Prevention
QA003	Children over six months old should receive complementary foods (MP-ASI) that contain carbohydrates, animal protein, fat, and vitamins to support growth.	<i>Apa yang harus diberikan kepada anak setelah usia 6 bulan?</i> (in English: What kind of food should be given to children over six months old?)	<i>MP-ASI bergizi seimbang dengan karbohidrat, protein hewani, lemak, dan vitamin.</i> (In English: Complementary foods with carbs, animal protein, fat, and vitamins.)	Child Nutrition

QA004	Stunting can be prevented by giving exclusive breastfeeding for the first six months, followed by appropriate complementary feeding.	<i>Bagaimana cara mencegah stunting pada bayi?</i> (in English: How does exclusive breastfeeding help prevent stunting?)	<i>ASI eksklusif dan MP-ASI bergizi sesuai usia.</i> (In English: Exclusive breastfeeding and continued nutritious complementary feeding.)	Prevention
QA005	Stunted children are at risk of impaired brain development, poor academic achievement, and reduced productivity in adulthood.	<i>Apa dampak jangka panjang stunting?</i> (in English: What are the long-term effects of stunting?)	<i>Gangguan otak dan produktivitas rendah.</i> (In English: Impaired brain development and lower productivity.)	Impacts
QA006	The national stunting prevention program includes specific interventions such as nutrient supplementation, immunization, and community nutrition education.	<i>Apa saja program pemerintah untuk mencegah stunting?</i> (in English: What programs are included in Indonesia's stunting prevention initiative?)	<i>Suplementasi, imunisasi, dan edukasi gizi.</i> (In English: Supplementation, immunization, and nutrition education programs.)	Policy
QA007	Iron deficiency in children can cause anemia, which worsens stunting conditions.	<i>Apa hubungan anemia dengan stunting?</i> (in English: How does iron deficiency affect stunting?)	<i>Kekurangan zat besi menyebabkan anemia dan memperparah stunting.</i> (In English: Iron deficiency causes anemia that worsens stunting.)	Micronutrients
QA008	Education for mothers and families is essential to promote understanding of healthy eating patterns and good environmental sanitation.	<i>Mengapa edukasi keluarga penting dalam mencegah stunting?</i> (in English: Why is family education important in preventing stunting?)	<i>Keluarga menjaga gizi, kebersihan, dan tumbuh kembang anak.</i> (In English: Families ensure nutrition, hygiene, and child growth support.)	Education
QA009	Poor environmental sanitation can cause repeated infections that inhibit nutrient absorption.	<i>Bagaimana hubungan sanitasi dengan stunting?</i> (in English: How does poor sanitation contribute to stunting?)	<i>Sanitasi buruk menyebabkan infeksi dan gangguan penyerapan gizi.</i> (In English: Poor sanitation causes infections that hinder nutrient absorption.)	Environment
QA010	Toddlers need nutritious food daily with a variety of menus to meet vitamin and mineral requirements.	<i>Seberapa sering anak harus mendapat makanan bergizi?</i> (in English: How can toddlers' nutritional needs be met daily?)	<i>Variasi menu harian memenuhi kebutuhan gizi anak.</i> (In English: Daily varied meals meet children's nutritional needs.)	Dietary Patterns

3.2.3. Data Preprocessing Pipeline

The data preprocessing pipeline was designed to ensure that all textual materials meet both technical compatibility and linguistic clarity requirements of the QA system. The workflow comprises five main stages: text cleaning, tokenization, semantic annotation, question generation, and quality assurance.

In the text cleaning stage, raw materials collected from diverse sources were standardized through the removal of formatting inconsistencies, normalization of Indonesian character encodings, and elimination of redundant or non-informative content. Particular care was taken to preserve medical terminology and to simplify syntactic complexity, ensuring that the resulting text remains accurate yet readable for non-specialist audiences.

The tokenization and segmentation process employed Indonesian-specific linguistic rules to divide the text into semantically coherent units. This step addressed the morphological richness of Bahasa Indonesia, handling variations such as affixation and reduplication that influence contextual meaning and downstream model interpretation.

Subsequently, a semantic annotation layer was added, tagging each segment with metadata related to topic relevance, difficulty level, and intended audience. These annotations enable the QA model to deliver context-sensitive and age-appropriate responses by aligning system outputs with user profiles and query intent.

The question generation phase transformed declarative content into natural-language questions using a combination of template-based and neural generation methods. Each generated question was manually reviewed by human annotators to ensure linguistic naturalness, contextual accuracy, and alignment with frequently asked health-related topics. This process resulted in a diverse query set that enhances the model's ability to generalize across varied user inputs.

Finally, the quality assurance stage implemented a multi-level validation protocol combining automated checks and expert evaluation. The dataset was subjected to consistency verification, medical content review by healthcare professionals, and readability assessment using standardized linguistic metrics. This comprehensive validation ensured that the final corpus adheres to both technical robustness and educational reliability standards required for health information dissemination.

3.3. Question Answering Model Development

3.3.1. Model Architecture Selection

The core Question Answering (QA) engine was developed using BERT-Indo, a variant of the Bidirectional Encoder Representations from Transformers (BERT) model that has been pretrained on extensive Indonesian language corpora. This model was chosen for its strong contextual comprehension in Indonesian, supporting accurate interpretation of medical queries and nuanced phrasing within health information retrieval tasks.

Several architectural modifications were introduced to enhance the base BERT-Indo framework. First, domain adaptation was performed through further fine-tuning on health-related Indonesian datasets, including medical journals, educational resources, and patient discussions. This adaptation strengthened the model’s ability to process domain-specific terminology and contextual semantics.

Second, a multi-task learning strategy was implemented, enabling the model to jointly learn *question answering*, *text classification*, and *semantic similarity* tasks. This multi-objective setup was intended to support generalization across varied query types. Lastly, attention mechanism refinement was applied by integrating custom medical-focused attention layers designed to emphasize key biomedical terms and concepts. This enhancement was designed to increase the model’s focus on medically relevant content during answer identification and generation. A summary of the architectural configuration is presented in Table 2.

The BERT-Indo model was fine-tuned using the AdamW optimizer with a learning rate of 2×10^{-5} and a batch size of 16 for four training epochs. The dataset was split into training (80%), validation (10%), and testing (10%) subsets. Early stopping based on validation loss was applied to prevent overfitting.

Table 2. Summary of the BERT-Indo–based Question Answering Model Architecture.

Component	Description
Base Model	BERT-Indo
Adaptation Method	Fine-tuning on Indonesian health corpora
Learning Strategy	Multi-task learning (QA, classification, semantic similarity)
Attention Enhancement	Custom medical term-focused attention

3.3.2. Retrieval-Augmented Generation (RAG) Implementation

To balance factual accuracy with linguistic fluency, the system employs a Retrieval-Augmented Generation (RAG) architecture that integrates dense retrieval and generative modelling. This hybrid setup is intended to reduce the limitations of retrieval-only systems, which may be rigid, and generation-only systems, which may produce fluent but inconsistent responses.

The document indexing stage converts all knowledge base entries into dense vector embeddings using *sentence-transformers* trained on Indonesian health-related text. These embeddings capture semantic similarity beyond keyword overlap, allowing concept-level retrieval.

During the retrieval process, the system executes approximate nearest neighbour (ANN) searches to locate the most semantically relevant passages. Multiple relevance indicators are used, including semantic proximity, medical terminology overlap, and content recency. For each query, the system retrieves the top-5 most relevant passages (top-k = 5), which are used as contextual input for the generation stage.

In the response generation phase, retrieved text segments are supplied as conditioning context to a fine-tuned T5-Indo model. The generative model is constrained to use information strictly derived from the retrieved content, aiming to improve factual consistency while maintaining conversational naturalness. The T5-Indo model was fine-tuned using a learning rate of 3×10^{-5} with a batch size of 8 for three epochs. Early stopping based on validation performance was applied to maintain generalization.

Finally, a quality filtering module evaluates generated responses using automated metrics covering relevance, coherence, and domain accuracy. Outputs failing to meet thresholds are reprocessed or escalated for human validation to maintain reliability. The core components of the implemented RAG framework are summarized in Table 3 to provide a structured overview of the retrieval, generation, and validation processes.

Table 3. Key Components of the Retrieval-Augmented Generation (RAG) Framework.

Component	Description
Retrieval Encoder	Sentence-transformer (Indonesian health text)
Search Mechanism	Approximate Nearest Neighbour (ANN)
Generative Model	T5-Indo (fine-tuned)
Validation Module	Automated quality filtering (relevance, coherence, accuracy)

3.3.3. Technical Implementation Details

The system was implemented using Python 3.9, supported by several specialized libraries for natural language processing (NLP) and machine learning tasks. The Hugging Face Transformers and TensorFlow 2.8 frameworks were employed for model development and fine-tuning, while FAISS was integrated to enable high-performance similarity searches on dense vector representations. For semantic encoding and retrieval, Sentence Transformers provided efficient document embeddings.

Model training and inference were executed on an NVIDIA Tesla V100 GPU environment equipped with 64 GB of RAM, allowing for large-scale processing of textual data. A distributed training architecture using Horovod was adopted to ensure scalability and efficient utilization of computational resources.

To enhance deployment efficiency, multiple optimization strategies were implemented. Quantization was applied to minimize model size with minimal loss of accuracy, while knowledge distillation produced lightweight models suitable for edge devices. Additionally, caching mechanisms were incorporated to reduce latency by pre-storing frequently accessed embeddings and query results. A consolidated summary of the implementation environment, supporting libraries, hardware configuration, and applied optimization techniques is presented in Table 4.

Table 4. Summary of Technical Implementation Environment and Optimization Strategies.

Component	Description
Programming Language	Python 3.9
Core Libraries	Transformers, TensorFlow 2.8, FAISS, Sentence Transformers
Hardware	NVIDIA Tesla V100 GPU, 64 GB RAM
Distributed Framework	Horovod
Optimization	Quantization, Knowledge Distillation, Caching Strategy

3.4. System Integration and Learning Platform

3.4.1. System Architecture Overview

The learning platform is built on a microservices-based architecture that prioritizes scalability, modularity, and ease of maintenance. Each subsystem operates independently and communicates through well-defined APIs, ensuring that updates or failures in one component do not disrupt the entire platform. This architectural approach enables continuous improvement, streamlined deployment, and robust system resilience across diverse operational conditions.

The presentation layer is implemented using React.js to support dynamic content rendering and responsive interaction. Material-UI provides a coherent and adaptable design system, while Progressive Web App features extend usability through offline access and mobile-friendly performance. These choices ensure that the user interface remains consistent, accessible, and reliable across devices ranging from smartphones to desktop environments, providing users with a seamless experience regardless of their connectivity conditions.

At the service layer, the backend is developed using FastAPI to leverage asynchronous request handling and automated API documentation. Core modules include user authentication, natural language query processing, answer generation, and analytic monitoring, each deployed inside isolated Docker containers for predictable scaling and fault containment. The data layer employs a hybrid storage strategy combining PostgreSQL for structured information, SQLite for lightweight local operations, and Pinecone as a vector database to manage high-dimensional embeddings for semantic search. Together, these components create a responsive and efficient system capable of handling varied workloads and supporting advanced retrieval capabilities.

3.4.2. Learning Management Features

The learning management component expands the platform’s capabilities beyond simple question answering, offering a structured and adaptive educational environment. Its design supports guided learning, continuous monitoring, and community-driven engagement, enabling users to deepen their understanding of child nutrition and stunting prevention in a systematic and personalized manner. By integrating interactive content, performance analytics, and collaborative tools, the platform delivers a holistic learning experience that aligns with diverse user needs.

The platform provides interactive learning modules that organize material into progressive levels of complexity. Each module blends multimedia elements such as instructional videos, infographics, and hands-on assessments to support different learning styles and improve long-term retention. Alongside these modules, the system implements detailed progress tracking, using built-in analytics to analyse engagement patterns and identify individual learning gaps. These insights drive personalized content recommendations, ensuring that users receive reinforcement where it is most needed.

Assessment tools are embedded throughout the learning journey to evaluate comprehension at multiple stages. A dynamic testing engine adjusts question difficulty based on user performance, maintaining an appropriate level of challenge and sustaining motivation. Complementing these evaluative features, the platform includes community-based

components such as moderated discussion forums and peer-support channels. These spaces allow users to exchange experiences, validate understanding, and cultivate a collaborative learning culture while ensuring the accuracy and reliability of shared information.

3.4.3. Integration Strategies

The platform incorporates a set of integration mechanisms designed to ensure secure, reliable, and scalable interaction with external systems. These strategies enable the learning environment to operate within a broader digital ecosystem, maintaining consistent performance while supporting real-time, context-aware functionality. By centralizing communication, leveraging dynamic data sources, and streamlining user access, the system strengthens its operational flexibility and relevance in practical health education settings.

A centralized API Gateway coordinates all inbound and outbound communication, handling authentication, request routing, and rate limiting. This structure provides a unified control point for managing interoperability with third-party platforms, ensuring that external integrations follow the same security and access control protocols as internal services. Beyond this communication layer, the platform connects to external data sources such as national nutrition surveillance feeds and contextual datasets including weather indicators that influence food availability. These integrations enable the system to deliver recommendations that are both timely and regionally tailored.

To support seamless professional access, the system implements Single Sign-On mechanisms that integrate with existing authentication services used by healthcare workers. The federated identity approach allows multiple authentication providers to coexist, simplifying user onboarding while maintaining strict privacy and security guarantees. This strategy ensures that authorized personnel can access the platform efficiently without compromising data protection standards.

3.5. Educational Evaluation Methods

System performance was evaluated using a combination of technical, expert-based, usability, and educational metrics appropriate for domain-specific health question answering systems. The quality of generated responses was primarily assessed based on factual correctness and contextual relevance. A subset of system responses was independently evaluated by two healthcare experts to determine medical accuracy, clarity, and actionability. Based on expert annotations, Precision, Recall, and F1-score were calculated to quantify response correctness.

To complement expert evaluation, automatic text similarity metrics were computed on a validation subset using expert-written reference answers. BLEU and ROUGE-L were used to measure lexical overlap, while BERTScore was applied to assess semantic similarity between generated and reference responses. These automatic metrics were treated as secondary indicators, as factual correctness is more critical than lexical similarity in health education contexts. System efficiency was evaluated using average response time during real-time interaction.

Usability testing was conducted as a separate evaluation involving 50 participants (30 healthcare workers and 20 parents). This phase applied think-aloud protocols and the System Usability Scale (SUS) to assess perceived usability. Post-interaction questionnaires were also used to measure user satisfaction, perceived usefulness, ease of use, and the relevance of system responses based on Technology Acceptance Model (TAM) constructs.

Educational effectiveness was evaluated through a controlled experimental study involving 120 participants. Participants were divided into an experimental group using the QA-based learning system and a control group using conventional learning materials. The control group received structured text-based learning materials covering the same stunting prevention and child nutrition topics as the experimental group. These materials were compiled from official health education sources and presented in a static format without interactive features. Both groups were given equal learning duration and were instructed to complete the learning session under the same time and study conditions to ensure comparable exposure.

Knowledge improvement was assessed using a standardized pre-test and post-test design. Participants completed a baseline assessment prior to system use and an immediate post-test after the learning session. To measure knowledge retention, a delayed post-test was administered four weeks after the intervention.

The assessment consisted of validated multiple-choice items covering key concepts in stunting prevention and child nutrition. All scores were normalized to a 0–100 scale. Learning improvement and retention were calculated using gain scores between pre-test and post-test results. Statistical analysis included significance testing to compare learning outcomes between groups, effect size estimation to assess the magnitude of differences, and confidence interval analysis to evaluate the reliability of the observed results.

3.6. Ethical Considerations

All data derive from open-access educational sources; no personal medical records are used. Anonymization, encryption, and minimal data retention ensure compliance with Indonesian and international privacy standards. Participants provide informed consent under institutional ethics approval, following the Declaration of Helsinki. The system includes disclaimers and referral mechanisms for medical emergencies to uphold beneficence principles. Bias mitigation involves dataset balance checks and fairness evaluation across demographics. Culturally adapted language and WCAG-compliant interfaces ensure inclusivity and accessibility.

3.7. Limitations and Mitigation

Dialectal variations may challenge language models; regional fine-tuning addresses this. Automated content monitoring maintains data currency. Model compression and edge computing reduce computational load. Current coverage focuses on stunting and nutrition; future extensions will include maternal and infectious health. Progressive enhancement supports complex query handling, while disclaimers and hotline links manage emergency cases. Limited sample size constrains generalizability; multi-region trials are planned. Current findings emphasize short-term learning; longitudinal studies will assess sustained outcomes. Cultural and socioeconomic adaptations will further refine context alignment.

4. Results and Discussion

The evaluation of the NLP-based Question Answering (QA) system for stunting-related health education was conducted through two separate phases: (1) a controlled experimental study to measure learning effectiveness and (2) an independent usability study to assess user experience. The overall goal of these evaluations was to examine the system’s effectiveness in supporting learning outcomes and user interaction.

The experimental study focused on knowledge improvement using pre- and post-tests administered before and after system interaction. System performance within the learning environment was also monitored based on answer accuracy, response time, and the relevance of the provided information.

The learning effectiveness evaluation involved 120 participants, consisting of healthcare workers (30%), medical students (25%), mothers of toddlers (35%), and general users (10%). Participants were divided into a control group using conventional educational materials and an experimental group using the QA-based learning system.

The usability evaluation was conducted as a separate study involving an additional 50 participants. Usability was assessed using the System Usability Scale (SUS), the Technology Acceptance Model (TAM), and user satisfaction measures to evaluate perceived usefulness, ease of use, and overall interaction quality.

4.1. System Performance Evaluation

The performance of the QA system was evaluated to assess response accuracy, relevance, semantic quality, and operational efficiency in a health education context. A total of 300 randomly selected question–answer interactions were independently reviewed by two healthcare experts. Each response was evaluated for medical correctness and contextual appropriateness. Disagreements were resolved through discussion to produce final labels.

Based on expert validation, Precision, Recall, and F1-score were calculated. In addition, the proportion of responses judged as relevant and actionable was recorded. To provide complementary text quality assessment, automatic evaluation metrics were computed on a validation subset using expert-written reference answers. BLEU and ROUGE-L measure lexical similarity, while BERTScore reflects semantic similarity between generated and reference responses. System efficiency was measured using average response time during live interaction. Table 5 presents the overall system performance metrics.

Table 5. System Performance Metrics.

Metric	Value
Answer Accuracy (Expert)	85.6%
Precision	86.2%
Recall	87.3%
F1-score	86.7%
Relevant Responses (Expert-rated)	88.4%
BLEU	0.62
ROUGE-L	0.74
BERTScore (F1)	0.89
Average Response Time	1.8 seconds

The results indicate that the system provides responses that are factually accurate and contextually appropriate for most user queries. The relatively moderate BLEU and ROUGE scores reflect expected paraphrasing in educational responses, while the high BERTScore suggests strong semantic alignment with expert references. The average response time of 1.8 seconds supports real-time interaction requirements for e-health learning applications.

4.2. Knowledge Improvement Results

4.2.1. Pre-test and Post-test Analysis

The evaluation results showed a substantial improvement in participants' knowledge after using the QA system. A digital learning evaluation framework commonly applied in health education research was used to assess the magnitude of learning gains [47]. A summary of the pre-test, post-test, and gain score results for both groups is provided in Table 6.

Table 6. Pre-test and Post-test Scores of the Experimental and Control Groups.

Group	N	Pre-test Mean (SD)	Post-test Mean (SD)	Gain Score	p-value
Experimental (QA)	60	42.3 (12.5)	78.6 (8.9)	+36.3	< 0.001
Control (Conventional)	60	43.1 (11.8)	61.2 (10.3)	+18.1	< 0.001
Difference	–	- 0.8	17.4	18.2	< 0.001

The experimental group demonstrated significantly higher learning gains compared to the control group ($p < 0.001$). The QA system produced an average improvement of 36.3 points (85.8%), while the control group improved by 18.1 points (42.0%). These results suggest that the QA system was associated with greater learning gains compared to the conventional approach within the study setting.

4.2.2. Domain-Specific Knowledge Improvement

A deeper analysis revealed variations in knowledge improvement across different stunting-related domains. The QA system consistently enhanced understanding in all topic areas, with the largest improvement observed in prevention and intervention content. The detailed distribution of knowledge gains across domains is presented in Table 7.

Table 7. Domain-Specific Knowledge Gains of the Experimental and Control Groups.

Topic Domain	Pre-test (%)	Post-test (%)	Improvement (%)
Definition of Stunting	52.3	91.7	+39.4
Causes and Risk Factors	38.5	85.2	+46.7
Prevention and Interventions	35.8	88.9	+53.1
Toddler Nutrition	41.2	86.4	+45.2
Long-term Impacts	29.7	79.3	+49.6
Policies and Programs	31.4	72.8	+41.4
Early Detection	44.6	83.5	+38.9
Micronutrient Intake	37.9	81.6	+43.7

The highest improvement occurred in the “Prevention and Interventions” domain (+53.1%), indicating that the QA system was particularly effective in delivering practical, action-oriented content that users could apply directly. In addition, all domains showed consistent upward trends, suggesting that the system supported broad improvements across multiple topic areas rather than isolated gains. This pattern demonstrates the system's ability to strengthen both foundational knowledge and applied health concepts across diverse user groups.

4.2.3. Interpretation of Findings

The combined results from the overall scores and domain-specific analysis indicate that the QA system contributed to measurable improvements in users' understanding of stunting. The learning gains suggest that interactive question–answer mechanisms may support more active engagement with health content compared to conventional text-based materials. The consistent improvements across all domains also show that the system supports diverse user backgrounds without performance degradation. Overall, these findings indicate the potential usefulness of NLP-driven QA systems for supporting health knowledge acquisition.

4.2.4. Effect Size Analysis

To evaluate the magnitude of the intervention's impact, Cohen's d was computed using the post-test means and the pooled standard deviation of both groups. The calculations were as follows:

- Mean difference: $78.6 - 61.2 = 17.4$
- Pooled standard deviation: ≈ 9.69

- Cohen's d : $17.4 / 9.69 = 1.79$

A Cohen's d value of 1.79 indicates a large observed effect within the study context. Although this value exceeds the commonly reported range for educational interventions, the magnitude should be interpreted with caution. The observed effect may be influenced by contextual factors, including the novelty of the interactive QA-based learning system, increased user engagement compared to conventional materials, and the short-term nature of the assessment. Therefore, the reported effect reflects the controlled experimental setting, and further studies involving longer-term evaluation and more diverse populations are required to confirm the stability and generalizability of the findings.

4.2.5. Retention Analysis

A follow-up assessment was conducted four weeks after the initial evaluation to measure knowledge retention. The experimental group retained 82.4% of the knowledge gained, while the control group retained only 61.3%. These findings suggest improved short-term retention among users who engaged with the QA system, in line with prior research on self-directed learning and memory consolidation. Studies have shown that self-directed learning promotes deeper cognitive engagement and that techniques like spaced practice and retrieval practice significantly enhance long-term memory retention [48,49]. Moreover, biological evidence from systems consolidation models further supports that memory becomes more stable over time [50].

4.3. Usability Evaluation Results

4.3.1. System Usability Scale (SUS) Results

The usability of the system was assessed using the 10-item System Usability Scale (SUS). The overall SUS score was 82.6 (SD = 11.3), which is categorized as "Excellent" based on SUS benchmarks [51]. Table 8 presents the SUS scores across the different user segments.

Table 8. System Usability Scale (SUS) Results Across User Segments.

User Segment	N	SUS Mean (SD)	Category
Health Workers	18	85.2 (9.8)	Excellent
Medical Students	15	84.7 (10.2)	Excellent
Mothers of Toddlers	21	79.8 (12.1)	Good
General Users	6	76.3 (13.5)	Good
Total	60	82.6 (11.3)	Excellent

These results indicate that users across various backgrounds generally perceived the system as easy to use, with the majority falling within the Excellent range. Health workers and medical students demonstrated the strongest usability ratings, suggesting smooth interaction even among users with demanding information needs. The consistent performance across segments also implies that the system design accommodates both expert and nonexpert users effectively.

4.3.2. Technology Acceptance Model (TAM) Analysis

The Technology Acceptance Model was used to evaluate user perceptions of usefulness, ease of use, intention to use, and actual usage behaviour. All constructs scored above the threshold of 4.0 on a 5-point scale, indicating strong acceptance [52]. Table 9. provides an overview of the TAM results.

Table 9. Technology Acceptance Model (TAM) Results.

Construct	Mean	SD
Perceived Usefulness (PU)	4.21	0.68
Perceived Ease of Use (PEOU)	4.08	0.72
Behavioural Intention to Use (BI)	4.15	0.65
Actual System Use	3.97	0.71

These results show that users perceived the system as both beneficial and easy to operate, reinforcing its overall practical value in supporting health-related information tasks. The strong behavioural-intention score indicates that users are highly likely to continue using the system, reflecting solid acceptance and perceived long-term utility. Although the actual-use score is slightly lower than the intention metrics, it remains robust and likely reflects

differences in users' prior experience with digital health platforms. Collectively, these scores indicate positive user acceptance across the key TAM constructs within the study sample.

4.3.3. User Satisfaction Survey

User satisfaction was measured using a structured 5-point Likert questionnaire to capture users' perceived quality of interaction with the system. This assessment focused on usability, clarity, response performance, and overall experience. The results provide a clear picture of how well the system meets user expectations across practical and informational aspects. Table 10. summarizes the satisfaction metrics for each indicator.

Table 10. User Satisfaction Indicators.

Indicator	Mean (SD)	% Very Satisfied / Satisfied
Ease of Use	4.23 (0.71)	89.3%
Clarity of Information	4.31 (0.68)	91.7%
Response Speed	4.18 (0.74)	87.5%
Answer Relevance	4.25 (0.69)	90.2%
Interface Design	4.12 (0.77)	85.8%
Overall Satisfaction	4.22 (0.70)	88.9%

The high levels of satisfaction across all indicators demonstrate that users valued both the functionality and the interaction quality offered by the system. Clarity of information received the highest rating, indicating that the system effectively delivers understandable and meaningful content. Overall satisfaction levels above 88% suggest that the system was well received by participants in terms of usability and information quality.

4.3.4. Task Completion Rate

The high completion rate indicates that the system workflow is logical and requires minimal user effort. The faster completion time supports the notion that the QA system reduces the cognitive and operational burden typically encountered in conventional learning methods. These results collectively demonstrate that the system enables efficient, streamlined user interaction.

Formula Used:

$$\text{Completaion Rate} = \frac{N_{\text{completed}}}{N_{\text{total}}} \times 100\% \quad (1)$$

$$\text{Average Time} = \frac{\sum_{i=1}^N T_i}{N} \quad (2)$$

Task performance was evaluated through task completion success and the time required to finish each assigned task. To ensure transparency of the reported metrics, the computation basis is presented in Table 11. These values form the source of the 94.2% completion rate and the 3.2-minute mean completion time reported in this subsection.

Table 11. Task Performance Summary.

Metric	Value
Total Assigned Tasks	312
Tasks Successfully Completed	294
Task Completion Rate	94.2%
Mean Completion Time	3.2 min
Baseline Completion Time	5.8 min

The results show that 94.2% of users successfully completed all assigned tasks. The average task completion time was 3.2 minutes, substantially faster than the conventional baseline of 5.8 minutes. The high completion rate indicates that the workflow is clear and requires minimal user effort. The faster completion time further suggests that the QA system reduces unnecessary cognitive and operational steps. Collectively, these results suggest that the system supports relatively efficient user interaction compared to the baseline condition.

4.3.5. Error Rate Analysis

User interaction errors were analyzed to assess system clarity and identify potential usability bottlenecks. Table 12 provides the underlying error counts and total user-interaction events used to derive the percentage values presented in this subsection.

Formula Used:

$$Error\ Type\% = \frac{E_{type}}{N_{total_actions}} \times 100\% \quad (3)$$

$$Total\ Error\ Rate = \frac{E_{nav} + E_{input} + E_{interp}}{N_{total_actions}} \times 100\% \quad (4)$$

Table 12. Error Summary and Computation Basis.

Error Type	Count	% of Total Actions
Navigation Errors	19	2.3%
Input Errors	34	4.1%
Interpretation Errors	15	1.8%
Total Errors	68	8.2%
Total User Actions Logged	830	–

The overall error rate was 8.2%, composed of navigation errors (2.3%), input errors (4.1%), and interpretation errors (1.8%). The relatively low error distribution indicates that users generally understood the interface structure and system feedback. Most errors were minor and did not disrupt task completion, reflecting a clear and well-designed interaction flow. These findings are consistent with usability research in health information systems, which shows that well-designed interfaces significantly reduce user error [51,52]. Moreover, contemporary evaluations highlight that low error occurrence is a key indicator of effective usability methods in digital health applications [53,54].

4.4. Discussion

4.4.1. Learning Performance and Usability Insights

The relationship between usability and learning outcomes in this study indicates a positive association. Higher usability scores were strongly associated with greater knowledge gains, indicating that users who experienced smoother interactions were better able to absorb the educational content. This pattern aligns with Cognitive Load Theory, which argues that well-structured interfaces minimize extraneous cognitive demands and allow users to process essential information more effectively. These findings suggest that usability may play an important role in supporting learning performance in digital health environments.

The magnitude of knowledge improvement observed in this evaluation appears higher than ranges reported in several previous studies, which commonly document increases of 45–70%. This outcome is consistent with literature showing that interactive question–answer systems support engagement and comprehension in health education contexts. The results provide preliminary evidence that NLP-driven QA systems may contribute to improved understanding within the evaluated context.

The results also carry important theoretical and practical implications. Theoretically, the study reinforces Constructivist Learning Theory by showing that active inquiry enabled through QA exchanges promotes higher-order cognitive processing. The findings also validate the relevance of the Technology Acceptance Model (TAM) for assessing user perceptions of e-health systems in Indonesia. Further, the evaluation contributes to emerging research by providing empirical evidence on the educational effectiveness of NLP-based QA tools. Practically, the system may be considered for further exploration within public health learning initiatives, telemedicine workflows, and community education programs. Its modular design also provides a basis for potential extension to related health domains such as maternal health, nutrition, and chronic disease management.

Cross-segment analysis highlights notable differences in system engagement and learning patterns. Health workers showed moderate learning gains but the highest interaction frequency, reflecting their familiarity with digital health resources. Mothers of toddlers achieved the largest improvement scores and reported strong satisfaction, indicating that the system effectively serves users with immediate personal health needs. Medical students displayed balanced outcomes across engagement and learning metrics, consistent with their intermediate domain expertise. These patterns demonstrate that the system accommodates users with varying health literacy levels and supports learning across diverse demographic groups.

Taken together, the findings indicate that the system was effective within the study context and demonstrated acceptable usability across user groups. The integration of NLP-based question-answering mechanisms supported consistent learning performance across segments, while the usability profile suggests accessibility for users with varying technical experience. The results emphasize the importance of user-centered design in shaping learning outcomes and engagement. Further evaluation is required before broader deployment in public health or telehealth environments.

4.4.2. System Outcomes, Limitations, and Research Directions

The evaluation produced several clear outcomes regarding the system's overall performance. Knowledge improvement was substantial, with the experimental group achieving significantly higher gains than the control group (85.8% vs. 42.0%). Usability ratings were consistently strong across user segments, as reflected in the average SUS score of 82.6. Retention outcomes were also robust, with users retaining 82.4% of their acquired knowledge after four weeks, and task completion rates reached 94.2%. The positive and statistically significant correlation between usability and knowledge gains ($r = 0.68$, $p < 0.001$) suggests that interaction quality may influence learning effectiveness.

The study's hypotheses were supported by the empirical findings. The first hypothesis, stating that the QA system enhances user knowledge, was validated with strong statistical significance. The second hypothesis, which proposed that the system would exhibit high usability and acceptance, was confirmed through both quantitative and qualitative indicators. The third hypothesis, predicting a positive association between usability and learning outcomes, was also validated, supporting theoretical expectations that more intuitive interfaces enable more efficient cognitive processing.

Despite these strengths, several limitations warrant attention. The Policy and Program content area produced lower relative gains, suggesting that the instructional materials in this domain may require restructuring, clearer examples, or additional contextual explanations. Users with limited digital literacy may also benefit from enhanced onboarding mechanisms, such as guided tutorials or simplified navigation cues. The inclusion of multimodal learning elements, such as images, animations, or scenario-based microlearning could strengthen user engagement and support broader comprehension across diverse audiences.

The findings open several pathways for future research. Longitudinal studies are needed to explore whether sustained system use leads to behavioral changes or improvements in real-world decision-making. Expanding the system to support multiple languages will improve accessibility across Indonesia's culturally and linguistically diverse population. Integrating the QA system with Electronic Health Records (EHR) could enable personalized content delivery based on user history and health status. Comparative evaluations using contemporary generative AI models will also help clarify the relative strengths of QA-centric architectures and inform the development of next-generation digital health learning solutions.

5. Conclusion

This study developed and empirically evaluated an NLP-based Question Answering (QA) system designed to enhance public health education on stunting. The findings indicate that the system demonstrated acceptable performance within the study context across accuracy, usability, and learning-related measures. First, the QA engine delivered precise and contextually relevant responses, achieving an accuracy of 85.6%, with a precision of 86.2% and recall of 87.3%. These results suggest that the system was able to interpret user questions and retrieve information from the curated knowledge base with reasonable reliability. Second, the integration of IndoBERT with a Retrieval-Augmented Generation (RAG) pipeline was designed to constrain responses to retrieved reference content, helping reduce the risk of unsupported information in health education settings. Third, the user-centered interface contributed to a positive interaction experience, reflected in an 89% satisfaction rate and a usability score of 4.2 out of 5, indicating that the system was generally acceptable to participants with varying levels of technical experience.

5.1. Limitations

Despite the overall success, several limitations must be acknowledged. The current dataset covers only eight primary topic categories, limiting the system's ability to address highly specific or complex clinical scenarios. Expanding the dataset to include edge cases and detailed subtopics will improve domain coverage. The learning effectiveness evaluation involved 120 participants, while usability testing was conducted with a separate group of 50 participants; larger and more diverse samples are needed to better represent real-world populations and usage conditions. Additionally, the system currently supports only Indonesian, which restricts adoption in multilingual regions across the country. Incorporating local dialects and regional languages will improve accessibility and inclusiveness.

Furthermore, the system was evaluated in a controlled study environment, and real-world deployment constraints were not examined. Practical factors such as internet connectivity, device availability, infrastructure reliability, and the workload of healthcare workers may influence system adoption and sustained use, particularly in low-resource settings. As a result, the scalability and deployment potential discussed in this study should be interpreted as preliminary, and field-based evaluations are required to assess feasibility and operational performance under real-world conditions.

5.2. Recommendations for Future Work

Future work should focus on improving both technical capability and real-world applicability. Enhancing the system toward multimodal functionality, such as processing images or visual health information (e.g., growth charts or feeding practices), may expand its educational value. Personalization based on user characteristics, including child age, risk factors, or regional context, could support more tailored and context-aware guidance. Improvements in multi-turn dialogue management are also needed to support longer interactions and maintain conversational coherence.

Beyond technical enhancements, future research should expand the system's content coverage to additional public health domains, including immunization, maternal care, and infectious disease prevention. Integration with telemedicine platforms or Electronic Health Records (EHR) could enable more personalized information delivery where appropriate. Developing a mobile application with offline functionality may also improve accessibility in low-connectivity settings.

Further evaluation should involve longitudinal studies to examine the system's longer-term influence on caregiver knowledge and behavior. Large-scale implementations within community or public health programs would provide more realistic operational insights. Finally, cross-regional or cross-country evaluations may help assess the adaptability and generalizability of the approach across different health contexts.

All the Declarations and Statements

Author Contributions Statement

Wiwin Suwarningsih – Conceptualization, Methodology, Formal Analysis, Writing – Original Draft, Writing – Review & Editing, and Supervision: Proposed the research concept, developed the methodological framework, conducted formal analysis, drafted the initial manuscript, contributed to manuscript revision, and supervised the overall research process.

Nuryani – Methodology, Software, Formal Analysis, Data Curation, and Writing – Review & Editing: Contributed to methodological design, implemented the software, performed formal analysis, managed data curation and preprocessing, and participated in manuscript revision.

Endang Suryawati – Validation, Data Curation, and Writing – Review & Editing: Conducted validation of results, contributed to data management, and reviewed and edited the manuscript for intellectual content.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflict of interest.

Funding Declaration

This research received no external funding.

Data Availability Statement

The dataset used in this study was constructed from publicly available health education sources, including materials from the Indonesian Ministry of Health, the National Stunting Reduction Program (<https://stunting.go.id>), WHO, UNICEF, and publicly accessible health information platforms. Due to copyright and licensing considerations associated with several source platforms, the compiled dataset is not publicly distributed but may be made available for academic research purposes upon reasonable request to the corresponding author.

Ethical Declarations

This study involved human participants in usability testing and learning evaluation. Participation was voluntary, and informed consent was obtained from all participants prior to data collection. The study was conducted in accordance with standard ethical research principles, and all collected data were anonymized and used solely for research purposes.

Acknowledgements

The authors would like to express their sincere appreciation to the reviewers for their valuable and constructive comments, which significantly contributed to improving the quality and clarity of this manuscript. The authors would like to thank the healthcare professionals who assisted in content validation and expert evaluation, as well as all participants who contributed to the usability testing and experimental study.

Declaration of Generative AI in Scholarly Writing

AI-assisted tools were used only for language editing and grammar refinement during manuscript preparation. All research design, analysis, interpretation, and final content were developed and reviewed by the authors, who take full responsibility for the integrity and accuracy of the work.

References

- [1] J. Dol, B. Richardson, T. Boates, and M. Campbell-Yeo, "Learning to parent from Google? Evaluation of available online health evidence for parents of preterm infants requiring neonatal intensive care," *Health Informatics J.*, vol. 25, no. 4, pp. 1265–1277, Dec. 2019, doi: 10.1177/1460458217752564.
- [2] J. Parga-Belinkie, B. Courchia, and D. Y. Barbeau, "Using media to transform education and public health messaging in the neonatal intensive care unit," *Semin. Perinatol.*, vol. 49, no. 7, p. 152150, Nov. 2025, doi: 10.1016/J.SEMPERI.2025.152150.
- [3] Ministry of Health Republic of Indonesia, "Indonesian Nutritional Status Survey (SSGI) 2022." Accessed: Feb. 12, 2026. [Online]. Available: <https://layanandata.kemkes.go.id/katalog-data/ssgi/ketersediaan-data/ssgi-2022>
- [4] Y. Abimulyani, Y. Y. Kainde, and M. T. Minggele, "Factors causing stunting in pregnant women and its prevention efforts: Public Health Perspectives," *Oshada*, vol. 1, no. 6, pp. 68–81, Dec. 2024, doi: 10.62872/AC0HF695.
- [5] G. Prayitno *et al.*, "Community-powered environmental pathways to reduce stunting: Food security, social capital, and open innovation in semi-urban Indonesia," *Environmental Challenges*, vol. 21, p. 101350, Dec. 2025, doi: 10.1016/J.ENVC.2025.101350.
- [6] J. P. McNulty *et al.*, "Professional challenges faced by student radiographers during COVID-19," *Radiography*, vol. 31, no. 4, p. 103000, Jul. 2025, doi: 10.1016/J.RADI.2025.103000.
- [7] S. Beaini, Y. Gharib, B. Kandil, C. Asmar, and N. Ziade, "ABS1178-HPR BUILDING RESILIENT HEALTHCARE SYSTEMS: INTEGRATING PLANETARY HEALTH EDUCATION IN THE FACE OF CLIMATE CHALLENGES – PERSPECTIVES FROM LEBANON," *Ann. Rheum. Dis.*, vol. 84, p. 1623, Jun. 2025, doi: 10.1016/J.ARD.2025.06.1031.
- [8] L. J. Morrison, S. Otis-Green, J. Bruno, P. N. Fordham, and E. C. Carey, "A Decade of Interactive Educational Exchange: Impacting Interprofessional Palliative Care Education," *J. Pain Symptom Manage.*, vol. 70, no. 6, pp. e351–e365, Dec. 2025, doi: 10.1016/J.JPAINSYMMAN.2025.08.001.
- [9] L. J. Morrison *et al.*, "Launching the Next Steps to Improve Hospice and Palliative Medicine Fellow Performance Assessment: A Look Back to the Initial Toolkit of Assessment Methods," *J. Pain Symptom Manage.*, vol. 61, no. 3, pp. 613–627, Mar. 2021, doi: 10.1016/j.jpainsymman.2020.10.014.
- [10] J. A. Greene, N. K. Choudhry, E. Kilabuk, and W. H. Shrank, "Online social networking by patients with diabetes: A qualitative evaluation of communication with Facebook," *J. Gen. Intern. Med.*, vol. 26, no. 3, pp. 287–292, Mar. 2011, doi: 10.1007/S11606-010-1526-3.
- [11] Y. Shang and X. (Robert) Luo, "An empirical investigation of users' engagement in online health communities for chronic diseases," *Information & Management*, vol. 62, no. 8, p. 104213, Dec. 2025, doi: 10.1016/J.IM.2025.104213.
- [12] L. Huang *et al.*, "Second Victim Experiences of Health Care Learners and the Influence of the Training Environment on Postevent Adaptation," *Mayo Clin. Proc. Innov. Qual. Outcomes*, vol. 8, no. 3, pp. 232–240, Jun. 2024, doi: 10.1016/J.MAYOCPIQO.2024.03.004.
- [13] V. Young, Y. Yan, P. J. Mason, L. El-Assad, and M. Mulvaney, "Preparing health care learners using an interprofessional virtual simulation: A team-based community intervention," *J. Interprof. Educ. Pract.*, vol. 37, p. 100729, Dec. 2024, doi: 10.1016/J.XJEP.2024.100729.
- [14] A. W. Kushniruk, M. M. Triola, E. M. Borycki, B. Stein, and J. L. Kannry, "Technology induced error and usability: The relationship between usability problems and prescription errors when using a handheld application," *Int. J. Med. Inform.*, vol. 74, no. 7–8, pp. 519–526, Aug. 2005, doi: 10.1016/J.IJMEDINF.2005.01.003.
- [15] X. Chen, H. Du, P. Zheng, and J. Mou, "Understanding information sensitivity in the digital era: A literature analysis and future research agenda," *Electron. Commer. Res. Appl.*, vol. 74, p. 101550, Nov. 2025, doi: 10.1016/J.ELERAP.2025.101550.
- [16] A. Osonuga, A. A. Osonuga, S. C. Fidelis, G. C. Osonuga, J. Juckes, and D. B. Olawade, "Bridging the digital divide: artificial intelligence as a catalyst for health equity in primary care settings," *Int. J. Med. Inform.*, vol. 204, p. 106051, Dec. 2025, doi: 10.1016/J.IJMEDINF.2025.106051.
- [17] A. Farea and F. Emmert-Streib, "Understanding question-answering systems: Evolution, applications, trends, and challenges," *Eng. Appl. Artif. Intell.*, vol. 156, p. 110997, Sep. 2025, doi: 10.1016/J.ENGAPPAI.2025.110997.
- [18] X. Wang, Z. Sun, P. Wang, and B. Wei, "MedicalGLM: A Pediatric Medical Question Answering Model with a quality evaluation mechanism," *J. Biomed. Inform.*, vol. 165, p. 104793, May 2025, doi: 10.1016/J.JBI.2025.104793.
- [19] V. Naidoo and K. K. Chadha, "Culturally responsive AI chatbots: From framework to field evidence," *Computers in Human Behavior: Artificial Humans*, vol. 6, p. 100224, Dec. 2025, doi: 10.1016/J.CHBAH.2025.100224.
- [20] S. Xiang, B. Liu, X. Jiang, Z. Huang, and Y. Ma, "Knowledge precedence networks: Mining progression patterns of scientific discoveries beyond prerequisites," *Inf. Process. Manag.*, vol. 63, no. 2, p. 104424, Mar. 2026, doi: 10.1016/J.IPM.2025.104424.
- [21] C. Ng and Y. Fung, "Educational Personalized Learning Path Planning with Large Language Models," Jul. 2024. [Online]. Available: <http://arxiv.org/abs/2407.11773>. [Accessed: Nov. 12, 2025]
- [22] E. M. Kivuti, D. Kaburu, and K. O. Ogada, "An Interactive Multimedia Model for Developing QOE-Enhanced E-Learning Platforms," *International Journal of Web-Based Learning and Teaching Technologies*, vol. 19, no. 1, Oct. 2024, doi: 10.4018/IJWLTT.347662.
- [23] J. Linardon *et al.*, "Does the Method of Content Delivery Matter? Randomized Controlled Comparison of an Internet-Based Intervention for Eating Disorder Symptoms With and Without Interactive Functionality," *Behav. Ther.*, vol. 53, no. 3, pp. 508–520, May 2022, doi: 10.1016/J.BETH.2021.12.001.
- [24] L. Lim *et al.*, "Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning," *Comput. Human Behav.*, vol. 139, p. 107547, Feb. 2023, doi: 10.1016/J.CHB.2022.107547.
- [25] A. Gril *et al.*, "The impact of different types of self-regulation scaffolds on learning science with hypermedia," *Comput. Human Behav.*, vol. 169, p. 108670, Aug. 2025, doi: 10.1016/J.CHB.2025.108670.
- [26] G. Kacmaz and A. K. Dubé, "Measuring teacher scaffolding in game-based learning: Emotional and responsibility scaffolds lead while resources and previous game experience shape practices," *Comput. Educ.*, vol. 237, p. 105393, Nov. 2025, doi: 10.1016/J.COMPEDU.2025.105393.

- [27] M. Saud and I. Wahyudi, "Shaping health behavior and social connectedness during COVID-19 through media adaptation strategies in Indonesia," *Vacunas (English Edition)*, vol. 26, no. 2, p. 500387, Apr. 2025, doi: 10.1016/J.VACUNE.2025.500387.
- [28] P. L. Liu, "COVID-19 Information Seeking on Digital Media and Preventive Behaviors: The Mediation Role of Worry," *Cyberpsychol. Behav. Soc. Netw.*, vol. 23, no. 10, pp. 677–682, Oct. 2020, doi: 10.1089/CYBER.2020.0250.
- [29] L. Naamati-Schneider and D. Alt, "Developing AI literacy in medical and health professions education: Exploring ChatGPT, ChatBots development, and simulation-based learning," *Nurse Educ. Pract.*, vol. 88, p. 104572, Oct. 2025, doi: 10.1016/J.NEPR.2025.104572.
- [30] C. R. Dias, M. R. Pereira, and A. P. Freire, "Qualitative review of usability problems in health information systems for radiology," *J. Biomed. Inform.*, vol. 76, pp. 19–33, Dec. 2017, doi: 10.1016/J.JBI.2017.10.004.
- [31] B. McBride, L. T. Nguyen, D. Wiljer, N. C. Vu, C. K. Nguyen, and J. O'Neil, "Development of a Maternal, Newborn and Child mHealth Intervention in Thai Nguyen Province, Vietnam: Protocol for the mMom Project," *JMIR Res. Protoc.*, vol. 7, no. 1, Jan. 2018, doi: 10.2196/RESPROT.7912.
- [32] D. N. Aisyah *et al.*, "Using WhatsApp for Nutrition Surveillance Among Children Under 5 Years in West Java, Indonesia: Cross-Sectional Survey and Feasibility Study," *JMIR Pediatr. Parent.*, vol. 8, Jan. 2025, doi: 10.2196/58752.
- [33] G. Gilano, A. Dekker, and R. Fijten, "Effect of mobile phone messaging on uptake of maternal and child health service in southern Ethiopia: Protocol for cluster randomized controlled trial," *Clinical Nutrition Open Science*, vol. 54, pp. 89–99, Apr. 2024, doi: 10.1016/J.NUTOS.2024.02.002.
- [34] S. Burrola-Méndez *et al.*, "Development, Implementation, and Usability Evaluation of the CANMI App to Monitor the Quality of Maternal and Child Nutrition Care in Primary Health Units: Mixed Methods Pilot Study," *JMIR Form. Res.*, vol. 9, Jan. 2025, doi: 10.2196/77539.
- [35] S. Till *et al.*, "Digital Health Technologies for Maternal and Child Health in Africa and Other Low- and Middle-Income Countries: Cross-disciplinary Scoping Review With Stakeholder Consultation," *J. Med. Internet Res.*, vol. 25, Jan. 2023, doi: 10.2196/42161.
- [36] A. Mimoudi, "Generative AI to bridge the educational divide: Personalized learning and challenges," *Social Sciences & Humanities Open*, vol. 12, p. 102140, Jan. 2025, doi: 10.1016/J.SSAHO.2025.102140.
- [37] Y. Liu *et al.*, "Toward a Large Language Model-Driven Medical Knowledge Retrieval and QA System: Framework Design and Evaluation," *Engineering*, vol. 50, pp. 270–282, Jul. 2025, doi: 10.1016/J.ENG.2025.02.010.
- [38] A. Tarifa and C. Chabir, "How scenarios can impact small models for QA task in education?," *Procedia Comput. Sci.*, vol. 270, pp. 4003–4013, Jan. 2025, doi: 10.1016/J.PROCS.2025.09.525.
- [39] M. A. Borroto and F. Ricca, "SPARQL-QA-v2 system for Knowledge Base Question Answering," *Expert Syst. Appl.*, vol. 229, p. 120383, Nov. 2023, doi: 10.1016/J.ESWA.2023.120383.
- [40] S. Jin, X. Lian, H. Jung, J. Park, and J. Suh, "Building a deep learning-based QA system from a CQA dataset," *Decis. Support Syst.*, vol. 175, p. 114038, Dec. 2023, doi: 10.1016/J.DSS.2023.114038.
- [41] S. Palmer *et al.*, "Policy, Systems, and Environmental Changes in Child Nutrition Programs: A Systematic Literature Review," *Advances in Nutrition*, vol. 15, no. 1, p. 100116, Jan. 2024, doi: 10.1016/J.ADVNUT.2023.09.004.
- [42] S. Chen, E. Turanoglu Bekar, J. Bokrantz, and A. Skoogh, "AI-enhanced digital twins in maintenance: Systematic review, industrial challenges, and bridging research–practice gaps," *J. Manuf. Syst.*, vol. 82, pp. 678–699, Oct. 2025, doi: 10.1016/J.JMSY.2025.07.006.
- [43] J. Dobrova *et al.*, "RAGCare-QA: A benchmark dataset for evaluating retrieval-augmented generation pipelines in theoretical medical knowledge," *Data Brief*, vol. 63, p. 112146, Dec. 2025, doi: 10.1016/J.DIB.2025.112146.
- [44] A. Simpson, L. Reynolds, I. Light, and J. Attenborough, "Talking with the experts: Evaluation of an online discussion forum involving mental health service users in the education of mental health nursing students," *Nurse Educ. Today*, vol. 28, no. 5, pp. 633–640, Jul. 2008, doi: 10.1016/J.NEDT.2007.09.009.
- [45] B. D. Unis *et al.*, "Nursing curricula in Indonesia and Scandinavia: What can we learn from each other to strengthen nursing education," *Nurse Educ. Today*, vol. 149, p. 106657, Jun. 2025, doi: 10.1016/J.NEDT.2025.106657.
- [46] R. D. Guntur, M. Lobo, D. M. Sihotang, Y. P. Bria, and D. Kusumaningrum, "Health Education Campaign to Improve Malaria Knowledge, Prevention, and Treatment Behaviors in Rural East Nusa Tenggara Province, Indonesia: Protocol for a Cluster-Assigned Quasi-Experimental Study," *JMIR Res. Protoc.*, vol. 14, Jan. 2025, doi: 10.2196/66982.
- [47] A. Bahattab, M. Hanna, G. T. Voicescu, I. Hubloue, F. Della Corte, and L. Ragazzoni, "e-Learning Evaluation Framework and Tools for Global Health and Public Health Education: Protocol for a Scoping Review," *JMIR Res. Protoc.*, vol. 12, no. 1, Jan. 2023, doi: 10.2196/49955.
- [48] D. Corral and S. K. Carpenter, "Effects of retrieval practice on retention and application of complex educational concepts," *Learn. Instr.*, vol. 100, p. 102219, Dec. 2025, doi: 10.1016/J.LEARNINSTRUC.2025.102219.
- [49] X. Zhang, Y. Lv, J. Zhang, W. Admiraal, and Y. Liu, "How Teachers' Self-efficacy and self-directed learning beliefs shape their teaching quality: A network analysis," *Teach. Teach. Educ.*, vol. 170, p. 105295, Feb. 2026, doi: 10.1016/J.TATE.2025.105295.
- [50] C. P. Thompson and M. A. Hughes, "The Effectiveness of Spaced Learning, Interleaving, and Retrieval Practice in Radiology Education: A Systematic Review," *Journal of the American College of Radiology*, vol. 20, no. 11, pp. 1092–1101, Nov. 2023, doi: 10.1016/J.JACR.2023.08.028.
- [51] A. Bangor, P. Kortum, and J. Miller, "Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale," 2009.
- [52] A. Granić and N. Marangunić, "Technology acceptance model in educational context: A systematic literature review," *Br. J. Educ. Technol.*, vol. 50, no. 5, pp. 2572–2593, May 2019, doi: 10.1111/bjet.12864.
- [53] F. Rangraz Jeddi, E. Nabovati, R. Bigham, and R. Farrahi, "Usability evaluation of a comprehensive national health information system: A heuristic evaluation," *Inform. Med. Unlocked*, vol. 19, p. 100332, Jan. 2020, doi: 10.1016/J.IMU.2020.100332.
- [54] B. Maqbool and S. Herold, "Potential effectiveness and efficiency issues in usability evaluation within digital health: A systematic literature review," *Journal of Systems and Software*, vol. 208, p. 111881, Feb. 2024, doi: 10.1016/J.JSS.2023.111881.

Authors' Profiles



Wiwin Suwarningsih received her bachelor's degree in informatics from Kebangsaan University of the Republic of Indonesia, Bandung, in 1996. She earned her master's degree in Informatics in 2000 and her doctoral degree in 2017 from the School of Electrical and Informatics Engineering, Bandung Institute of Technology. Currently, she works at the Research Center for Data and Information Sciences, National Agency of Research and Innovation (BRIN, Indonesia). Her research interests are artificial intelligence, computational linguistics, Indonesian natural language processing and text mining, information retrieval and question answering systems. Since 2017, she has been working on a project regarding Indonesian medical information retrieval using deep learning to support high-quality medical information.



Nuryani received her bachelor's degree in computer science from Gadjah Mada University (UGM) in 2003. She earned her master's degree in 2013 and her doctoral degree in 2025 from the School of Electrical and Informatics Engineering, Bandung Institute of Technology. Currently, she works at the Research Center for Data and Information Sciences, National Agency of Research and Innovation (BRIN, Indonesia). Her research interests include computer networks, deep learning, and natural language processing, especially in sentiment analysis.



Endang Suryawati received her bachelor's degree in computer science from ST. INTEN, Bandung, in 2006. She earned her master's degree in 2016 from the School of Electrical and Informatics Engineering, Bandung Institute of Technology. Currently, she works at the Research Center for Artificial Intelligence and Cyber Security, National Agency of Research and Innovation (BRIN, Indonesia). Her research interests include artificial intelligence, deep learning, transformers, and multimodal large language models, especially in image-to-text captioning.

How to cite this paper

Wiwin Suwarningsih, Nuryani, Endang Suryawati, "Integrating Question Answering Technology into E-Health Education: A Computer-Supported Learning Approach", International Journal of Modern Education and Computer Science(IJMECS), Vol.18, No.4, pp. 182-200, 2026. DOI:10.5815/ijmecs.2026.04.11