I. INTRODUCTION

Early marriage continues to pose a significant societal concern in Indonesia, particularly in Lombok, West Nusa Tenggara (NTB), where socio-cultural norms exert strong influence. The consequences include intergenerational poverty, limited educational attainment, and heightened exposure to physical and psychological health risks. Despite ongoing efforts by governmental and non-governmental bodies, early marriage rates remain high [1]. Recent statistics from the Central Bureau of Statistics (2022) identify NTB as having the highest national percentage 16.23% of women aged 20–24 who entered marriage or cohabitation before age 18 [2].

Advancements in digital technology, especially in Natural Language Processing (NLP), have opened new avenues for social assistance. One such innovation is virtual chatbot systems [3], [4], which mimic human conversation and serve in emotional support, education, and health services. The constant availability and anonymous nature of chatbots make them effective for addressing sensitive issues like early marriage.

This study introduces an NLP-driven counseling chatbot that combines emotion recognition and reason identification in its dialogue system. IndoBERT, a pre-trained transformer model for Bahasa Indonesia, acts as the main classification engine and the embedding model for detecting reasons based on semantic similarity. Initial training used 2,000 manually labeled synthetic dialogues created by a Large Language Model (LLM), designed to capture typical interactions between counselors and youth. To improve contextual flexibility and enhance emotion inference, an additional 10,000 unlabeled samples were added during a second training phase.

A central contribution is the introduction of function-oriented emotional categories such as ‘Enthusiastic’, ‘Gentle’, ‘Analytical’, ‘Inspirational’, and ‘Cautionary’. These categories better capture the purpose of counseling conversations than traditional emotion labels. There is also a reason identification module based on semantic similarity that classifies user inputs into the categories of Education, Economy, Religion, or Culture. Unlike standard keyword-matching methods, this module uses IndoBERT base and IndoBERT-based sentence embeddings to improve generalization and context awareness. Together, these two components provide a deeper understanding of user context. This allows the chatbot to give responses that are both emotionally sensitive and informed by the context. Additionally, the system introduces a gesture-mapping framework. This framework links each emotional category to specific nonverbal expressions, strengthening the connection between verbal and physical communication. By using IndoBERT’s modeling abilities, the framework detects subtle emotional nuances and identifies relationships between user inputs and predefined reason categories. This creates a more empathetic and context-aware virtual counseling agent aimed at preventing early marriage [5], [6], [7].

Following the growing effectiveness of Transformer-based architectures, recent research has confirmed the use of BERT and its derivatives for emotion and text classification tasks. In multi-class classification settings, combining BERT with active learning techniques has achieved an F1-score of 0.83. At the same time, it has reduced annotation costs by up to 85%. This highlights its efficiency in resource-limited environments [8]. In multilingual emotion recognition, architectures such as RoBERTa-MA and XLNet-MA have demonstrated accuracy rates of 62.4% and 60.5%, respectively, reflecting the contributions of multi-attention mechanisms to emotional inference from text [9].

For Bahasa Indonesia, IndoBERT pretrained on domain-relevant corpora has emerged as a robust baseline. When combined with deep learning structures such as BiLSTM, BiGRU, and attention layers, IndoBERT-based models have achieved up to 91% accuracy and 78% on benchmarks such as IndoNLU, confirming their capacity to detect nuanced emotional content in Indonesian text [10]. Parallel approaches employing CNN and BiGRU hybrids have reported F1-scores exceeding 80% on general emotion datasets like ISEAR and WASSA, demonstrating the continued relevance of deep learning under tailored configurations [11]. Additionally, emerging methods that incorporate large language models with pseudo-labeling, such as ChatGPT-supported few-shot annotation using CamemBERT have achieved promising F1-scores up to 0.6662 in multilingual emotion classification tasks [12].

While prior studies have made considerable progress in emotion classification and facial expression synthesis, the translation of functional text-based emotions into embodied gestures remains largely underexplored particularly in culturally sensitive and low-resource environments. Addressing this gap, the present study proposes a novel framework that utilizes IndoBERT-derived emotional outputs to drive synchronized body and hand gestures, grounded in principles of nonverbal communication and attuned to local socio-cultural dynamics.

This study aims to investigate whether a multi-phase fine-tuning strategy combined with class balancing techniques can improve emotion recognition performance, particularly for minority emotion categories critical to early marriage prevention. By aligning textual emotion recognition with expressive gestural responses, this approach introduces a multimodal layer of interaction that enhances the emotional authenticity of virtual counseling agents. Rather than treating gesture generation as a cosmetic addition, the framework positions it as a core communicative feature. In doing so, it improves the design of emotionally intelligent systems that not only understand emotion but also express it. This is crucial for building empathy, trust, and connection in sensitive counseling dialogues.