Enhancing Consular Services for Indonesian Citizens Abroad: A Framework for AI-Powered Question Answering

I. Analysis and Refinement of the Research Idea: "Fine-Tuning BERT for Automated Consular Question Answering for Indonesian Citizens Abroad"

The initial research concept, "Fine-Tuning BERT for Automated Consular Question Answering for Indonesian Citizens Abroad," presents a timely and relevant area of investigation. However, a critical evaluation is necessary to refine it into a robust Master's thesis topic, considering the current advancements in Natural Language Processing (NLP) and the specific needs of Indonesian citizens.

A. Critical Evaluation of the Initial Concept

The proposal to develop an automated question-answering (QA) system for Indonesian consular services using fine-tuned Bidirectional Encoder Representations from Transformers (BERT) models holds several strengths. It directly addresses a significant real-world challenge: providing accessible and timely information to Indonesian citizens residing overseas.[1] The use of BERT, a foundational transformer model, is logical given its well-documented success in various NLP tasks through fine-tuning.[3] Furthermore, the focus on a specific case study—Indonesian citizens—allows for a depth of analysis that can yield impactful, contextually relevant solutions.

However, the initial concept also presents weaknesses and challenges that must be addressed. The primary concern is the specificity of "BERT." While BERT and its Indonesian variants like IndoBERT [3] have been instrumental, the NLP field is advancing rapidly. Newer, often larger, and more capable architectures such as GPT-4 [7], Llama3 [8], and Gemma2 [9] are demonstrating superior performance on many benchmarks. Notably, the SahabatAI models, developed specifically for Indonesian and its regional dialects, are based on Gemma2 and Llama3 architectures.[8] A Master's thesis aiming for contemporary relevance should carefully consider whether "BERT" is a deliberate choice due to specific constraints (e.g., computational resources, focus on existing Indonesian BERT variants) or if it serves as a general term for transformer-based models. Sticking rigidly to an older interpretation of BERT might limit the thesis's engagement with state-of-the-art approaches and could result in comparisons against less current baselines. The decision here will significantly influence the novelty and potential performance benchmarks of the proposed system.

Another challenge is the broad scope of "consular question answering." Consular services encompass a vast range of queries, from simple factual questions to complex, sensitive issues.[10] The thesis must clearly define the types of questions the system will address. Additionally, the Indonesian language, with its numerous regional dialects, poses significant low-resource challenges.[3] While Bahasa Indonesia is the official language, many citizens, particularly certain demographics like migrant workers, may be more comfortable or only fluent in their regional dialects.[16] Finally, the availability of suitable data for fine-tuning a model on specific consular topics in Indonesian and its dialects is a critical hurdle that needs a well-thought-out strategy.

B. Potential Enhancements and Expansions

To elevate the research idea, several enhancements and expansions should be considered, particularly concerning multilingual support and the implications for global consular services.

1. The Imperative of Multilingual Support: Indonesian, English, and Regional Dialects

A crucial enhancement is the incorporation of multilingual capabilities. Indonesian citizens abroad exhibit varying levels of proficiency in Bahasa Indonesia and English, the latter often serving as a lingua franca in international contexts. More importantly, for inclusivity and effective communication with specific demographics, such as Indonesian Migrant Workers (PMI), support for regional dialects is paramount.[16] Users increasingly expect services in their native language; it is no longer a luxury but a "new normal".[18] Multilingual chatbots demonstrably improve accessibility, user satisfaction, and trust.[18]

The existing SARI (Sahabat Artificial Migran Indonesia) chatbot, developed for Indonesian migrant workers, already incorporates regional languages (specifically Javanese) precisely because many users are not fully comfortable with formal Bahasa Indonesia.[16] This government-backed initiative underscores the recognized need for dialectal support. Furthermore, the advanced SahabatAI models natively support Indonesian, Javanese, Sundanese, and English, providing a strong technological foundation for such a multilingual system.[8]

However, developing such a system is not without challenges. Data scarcity remains a significant issue for many Indonesian regional dialects.[3] The NLP complexities associated with these dialects, including unique grammatical structures and vocabulary, require specialized approaches.[14] Moreover, ensuring cultural sensitivity and avoiding misinterpretations or offensive content due to translation errors is

critical, especially in the context of official government communication.[13]

2. Implications of Multilingual versus Single-Language Systems for Global Consular Services

The choice between a multilingual and a single-language system has profound implications for the delivery of global consular services.

- **Accessibility and Inclusivity:** Multilingual systems dramatically broaden the reach of consular services, ensuring that citizens are not excluded or disadvantaged due to language barriers.[18] This is fundamental for public services that aim for equitable access to information and assistance.
- **User Trust and Satisfaction:** Interacting with citizens in their native tongue, while respecting cultural nuances, fosters a sense of understanding and value, thereby building trust in the services provided.[13] This is particularly vital in consular interactions, which can often occur during times of stress or uncertainty for the citizen.
- **Operational Efficiency and Cost:** While the initial development of a multilingual system is more complex and potentially costly, the long-term benefits include enhanced operational efficiency. Automation in multiple languages can reduce the reliance on a large cadre of multilingual human agents, offering scalability and 24/7 availability of support.[13]
- **Data Management Complexity:** A significant operational challenge for multilingual systems is the maintenance of content. Consular information, such as visa regulations or emergency procedures, is frequently updated. Ensuring consistency, accuracy, and cultural appropriateness across all supported language versions requires robust content management strategies and workflows.[18]

The SARI chatbot serves as an important reference point here. Its design, which explicitly includes regional languages and an empathetic, "humanist AI" approach, was driven by the identified needs of Indonesian citizens abroad, particularly vulnerable migrant workers.[16] This indicates that a purely informational QA system might be insufficient for comprehensive consular support; an element of empathetic and culturally aware interaction is highly desirable. A thesis project can build upon SARI's precedent by potentially expanding the range of supported regional dialects, applying similar empathetic design principles to a broader array of consular queries, or investigating more advanced NLP techniques to achieve these nuanced interaction goals.

C. Strategies for Narrowing Scope for a Master's Thesis

Given the breadth of consular services and the complexities of multilingual NLP, narrowing the scope is essential for a feasible Master's thesis. Several strategies can be employed:

- **Focus on Specific Consular Services:** Instead of attempting to cover all consular functions, the thesis could concentrate on a well-defined set of services. Examples include passport renewal procedures, visa application guidance (either for family members of Indonesian citizens or for Indonesian citizens applying for visas to other countries), assistance with the *Lapor Diri* (self-registration for citizens abroad) process, guidance on legal document attestation, or protocols for emergency assistance.[1]
- **Target Specific User Groups:** The needs and linguistic patterns of different user groups can vary significantly. The thesis could focus on Indonesian migrant workers (aligning with the SARI chatbot's target audience), students studying abroad, or general expatriates and their families.
- **Limit Regional Dialects:** While broad dialectal support is ideal, for a Master's thesis, it may be more practical to focus on Bahasa Indonesia and perhaps one or two major regional dialects where data resources are relatively more available or where models like SahabatAI offer strong existing support (e.g., Javanese, Sundanese).[8]
- **Concentrate on a Specific AI Technique or Research Question:** Rather than implementing a comprehensive system with all possible AI features, the thesis could conduct a deep dive into a particular technique. For instance, it could focus on the fine-tuning intricacies of a specific model like SahabatAI for consular tasks, or the development and evaluation of a robust Retrieval Augmented Generation (RAG) pipeline for answering frequently asked consular questions.

D. Proposed Refined Research Topic and Questions (Example)

Based on the analysis, here are examples of refined research topics:

- **Topic Example 1 (Leveraging SahabatAI for Multilingual and Empathetic QA):** "Developing a Multilingual and Empathetic Question Answering System using SahabatAI for Indonesian Consular Services, with a Focus on Javanese and Sundanese Dialect Support."
- **Topic Example 2 (RAG-focused for Specific Services):** "A Retrieval-Augmented Generation Approach for Answering Frequently Asked Questions on Indonesian Passport and Visa Services in Bahasa Indonesia and English."

Corresponding research questions could include:

1. How effectively can advanced pre-trained language models, such as IndoBERT and SahabatAI, be fine-tuned to address specific question-answering tasks within the Indonesian consular domain, covering both Bahasa Indonesia and selected regional dialects?
2. What is the tangible impact of incorporating regional dialects (e.g., Javanese, Sundanese) on the accessibility, perceived usefulness, and user satisfaction of an automated consular QA system for Indonesian citizens abroad?
3. Can a Retrieval Augmented Generation (RAG) approach demonstrably improve the system's ability to provide accurate and up-to-date answers to questions requiring highly specific or dynamically changing consular information, as compared to fine-tuning methodologies alone?
4. What are the primary challenges and effective strategies for curating, creating, or augmenting a suitable dataset for training and rigorously evaluating a consular QA system tailored for Indonesian citizens, considering the linguistic diversity and data scarcity?
5. How can empathetic response generation be effectively incorporated into a consular QA system, and what evaluation methodologies can be developed to assess its impact on user experience and trust, particularly for sensitive queries?

Table 1: Comparison of Language System Approaches for Indonesian Consular Chatbots

Approach	Pros	Cons	Data Requirements & Availability	Technical Complexity	Potential Impact for Indonesian Citizens Abroad
Monolingual - Bahasa Indonesia	- Simpler to develop and maintain. - More readily available NLP resources for standard Indonesian.[6]	- Excludes citizens not fluent in formal Bahasa Indonesia. - May not address needs of all user groups (e.g., some PMIs).[16]	- Relatively better data availability (e.g., IndoQA, translated SQuAD).[31] - General Indonesian corpora are larger.	Moderate.	- Reduced accessibility for non-fluent speakers. - Potentially lower trust if language is a barrier.

Multilingual - Bahasa Indonesia + English	- Caters to a wider audience, including those more comfortable with English. - English is a common language for international communication.	- Increases complexity of development and content management. - Requires robust translation or bilingual models.	- Good availability for English NLP resources. - Parallel corpora for Indonesian-English exist but may lack domain specificity.	Moderate to High.	- Improved accessibility for English speakers. - May still exclude those reliant on regional dialects.
Multilingual - Bahasa Indonesia + English + Key Regional Dialects (e.g., Javanese, Sundanese)	- Maximizes inclusivity and accessibility, especially for specific demographics like PMIs.[16] - Builds higher user trust and satisfaction by communicating in preferred dialects.[18] - SahabatAI offers a strong foundation.[8]	- Significant data scarcity for most regional dialects.[3] - High NLP complexity for dialects.[14] - Potential for cultural missteps if not handled carefully. - Increased maintenance overhead.	- SahabatAI provides pre-trained models and data for Javanese & Sundanese.[8] - NusaX dataset for some regional languages, but primarily for sentiment.[3] - Other dialects remain very low-resource.	High.	- Highest potential for accessibility, trust, and inclusivity across diverse Indonesian citizen groups. - Directly addresses needs identified by SARI chatbot initiative.[16]

This comparative analysis underscores that while supporting regional dialects presents the highest complexity, it also offers the greatest potential impact for inclusivity and user trust, aligning closely with the goals of public service delivery for a diverse citizenry. The availability of models like SahabatAI mitigates some of the challenges for specific major dialects.

II. The Landscape of Indonesian Consular Services and Current AI Interventions

Understanding the existing consular services, the demographics and needs of Indonesian citizens abroad, and current AI initiatives like the SARI chatbot is crucial for contextualizing the proposed thesis. This landscape informs the specific problems the AI system aims to solve and where it can offer the most value.

A. Common Consular Services, User Demographics, and Information Needs of Indonesian Citizens Abroad

Indonesian diplomatic missions (KBRI - Embassies, KJRI - Consulates General) provide a wide array of services to citizens residing or traveling overseas. Key services include:

- **Passport Services:** Issuance of new passports and renewal of expiring ones are among the most frequent requests.[27]
- **Visa Applications:** While primarily for foreigners visiting Indonesia, consular offices also assist Indonesian citizens with information or processes related to their visas for staying in the host country or traveling to third countries, especially concerning family members.[24]
- **Document Legalization and Attestation:** Legalization of various documents (e.g., birth certificates, marriage certificates, academic transcripts, business documents) for use in Indonesia or the host country is a common need.[10]
- **Lapor Diri (Self-Reporting):** Indonesian citizens residing abroad are encouraged or required to register themselves with the nearest Indonesian mission. This facilitates communication, assistance, and data management.[1]
- **Assistance for Indonesian Migrant Workers (PMI):** This is a critical area, involving support for issues such as unpaid wages, exploitative working conditions, contract disputes, human trafficking, and provision of legal aid and shelter.[12] The Indonesian Migrant Workers Protection Agency (BP2MI) plays a significant role in this.[12]
- **Citizenship Matters:** Services related to confirmation of Indonesian citizenship (e.g., *Surat Keterangan Status Kewarganegaraan Republik Indonesia* - SKSK RI), renunciation of citizenship (*pelepasan kewarganegaraan*), and issues concerning dual citizenship for children.[28]
- **Emergency Assistance:** Support for citizens in distress due to accidents, becoming victims of crime, death of a family member, international parental child abduction, or requiring evacuation during crises.[11]
- **Customs Information:** Guidance for returning citizens on regulations concerning the import of personal belongings (*barang pindahan*).[26]

The **user demographics** are diverse, encompassing students, professionals, tourists, families, and a very significant population of Indonesian Migrant Workers (PMIs).[12] Many PMIs originate from specific regions in Indonesia, making familiarity with regional dialects particularly important for effective communication.[16]

Their **information needs and challenges** are varied:

- Access to clear, accurate, easily understandable, and up-to-date information regarding procedures, documentary requirements, official fees, and processing times for various services.
- Difficulties arising from incomplete or missing documentation, which can complicate applications for services like citizenship confirmation.[44]
- Urgent need for assistance and guidance in crisis situations or when facing legal, employment, or personal safety issues.[11]
- Navigating complex legal and administrative frameworks, both Indonesian and those of the host country.[44]

Existing platforms like the **Peduli WNI portal** [1] and **BP2MI resources** [12] serve as primary channels for Indonesian citizens to access information, report themselves, and lodge complaints or requests for assistance. An AI-powered QA system would aim to complement and enhance these existing services by providing instant, accessible responses to a wide range of queries.

The nature of these services reveals a dual requirement for any automated assistance system. On one hand, there is a substantial need for providing standardized, factual information for routine procedures like passport applications or Lapor Diri, which can be effectively handled by a traditional QA system drawing from FAQs and official documents.[2] On the other hand, many consular interactions, especially those involving PMIs facing exploitation [12] or citizens in emergencies [11], are highly sensitive and require not just information but also empathetic handling and clear pathways to human intervention. This duality suggests that a single, monolithic AI approach might not be optimal. A more sophisticated system might triage queries, offering direct answers for informational requests while employing empathetic response strategies and immediate escalation protocols for crisis situations, drawing inspiration from the design philosophy of the SARI chatbot.[16] The novelty for a thesis could lie in developing an AI capable of discerning query types and modulating its response style (informational versus empathetic) and recommended action (provide answer versus escalate to human agent).

B. The SARI Chatbot: A Case Study in AI for Indonesian Migrant Worker Support

The SARI (Sahabat Artificial Migran Indonesia) chatbot, launched in April 2025, represents a significant advancement in the use of AI for Indonesian consular services.[16]

- **Collaboration:** It is a joint initiative of the Indonesian Ministry of Foreign Affairs (Kemlu) and UN Women.[16]
- **Purpose:** Primarily designed to protect Indonesian female migrant workers from violence and exploitation by providing accessible information and initial support services.[17]
- **Features and Technology:**
  - The chatbot is AI-powered and integrated into the existing "Safe Travel" mobile application, which provides practical information for Indonesian citizens abroad.[16]
  - A key feature is its support for various **regional languages**, including Javanese. This was a deliberate design choice, acknowledging that many migrant workers are more comfortable or fluent in their local dialects than in formal Bahasa Indonesia.[16]
  - SARI is conceptualized as a "**humanist AI**," developed to provide **empathetic responses** and understand human feelings, particularly crucial when addressing sensitive issues like violence faced by workers.[16] It was trained to provide empathy and respond to needs when problems arise.
  - The development process was **human-centered and participatory**, involving direct consultations with female migrant workers, service providers, civil society organizations, and youth representatives over a six-month period to ensure its relevance and effectiveness.[16]
  - It aims to provide initial support that complements the in-person services offered by Indonesian missions, particularly in handling cases of violence against female migrant workers.[17]

The SARI initiative is significant as it demonstrates the Indonesian government's commitment to leveraging AI for citizen protection and service delivery. It explicitly acknowledges the linguistic diversity of its citizens abroad and the necessity of empathy in digital interactions, especially for vulnerable groups. For the proposed Master's thesis, SARI serves as both a validation of the need for such AI tools and a benchmark.

The existence of SARI has direct implications for defining the novelty of a new thesis project. SARI is a recently launched, sophisticated AI chatbot targeting a core group of Indonesian citizens abroad (migrant workers) and already incorporates advanced features like regional language support and empathetic design.[16] This means a new

thesis must clearly differentiate its contributions. Simply replicating SARI's functionality for a slightly different user group or with a marginally different model might not offer sufficient novelty.[47] However, this also presents opportunities. The thesis could aim to:

1. Address a broader range of consular services not currently covered by SARI's primary focus on migrant worker protection.
2. Support different or additional regional Indonesian dialects beyond those currently implemented in SARI (details of SARI's full dialectal support are not extensively public).
3. Explore and evaluate more advanced or alternative NLP techniques for achieving robust QA and empathetic interaction than what SARI might be using (the underlying technology of SARI is described generally as "AI" [16]).
4. Conduct a rigorous comparative analysis of different AI approaches (e.g., fine-tuned SahabatAI versus a RAG architecture using SahabatAI) for tasks similar to those SARI addresses, thereby providing deeper technical insights and performance benchmarks. Given that detailed technical specifications of SARI's AI engine are not widely available, a thesis can make a valuable contribution by exploring, implementing, and rigorously evaluating specific state-of-the-art models and techniques to achieve similar or expanded goals in the Indonesian consular domain.

C. Other Relevant AI Applications in Government and Consular Services (Global Benchmarks)

Globally, governments are increasingly adopting AI to enhance public service delivery. Several international examples offer benchmarks and insights:

- **Singapore's GovTech Chatbots:** The Singaporean government has deployed multiple AI-powered chatbots, such as "Ask Jamie" (a virtual assistant across over 70 government websites) and "HealthBuddy" (for healthcare inquiries). These systems are multilingual (supporting English, Mandarin, and Malay), handle a high volume of diverse citizen inquiries, and have reportedly led to significant reductions in call center workload and faster response times.[48] This demonstrates the feasibility and benefits of large-scale government chatbot deployment.
- **VFS Global AI Chatbot for UK Visas:** VFS Global, a major visa application service provider, launched a Generative AI-powered chatbot to assist UK visa applicants across 141 countries. It offers 24/7 support, provides country-specific information, operates within a secure AI framework, and incorporates an ethical AI design with bias mitigation strategies.[35] This is highly relevant for any thesis component dealing with visa-related QA.

- **U.S. Citizenship and Immigration Services (USCIS):** USCIS leverages AI and Machine Learning (ML) to refine processes such as asylum seeker application processing, including data quality assessment and document verification.[49] This indicates AI's utility in complex administrative and document-intensive consular tasks.
- **General AI in Public Services:** AI is widely used to automate routine tasks, improve citizen interactions, and enhance operational efficiency.[49] AI-driven chatbots for citizen helpdesks are becoming increasingly common across various government functions.[48] Natural Language Processing (NLP) is also specifically cited for its role in language translation to improve accessibility for foreign speakers within a nation.[50]

While these global examples showcase mature and effective AI deployments, it is important to consider the specific context of Indonesia. Indonesia's public sector faces challenges in AI adoption, including lagging in global AI readiness indices, and gaps in digital policy, data infrastructure, and technical capacity. Siloed data systems and concerns around public trust and ethics also need to be addressed.[54] These local challenges mean that directly replicating a system like Singapore's "Ask Jamie" might face unique obstacles or require context-specific adaptations.

Nevertheless, the global best practices—emphasizing multilingual support, 24/7 availability, domain-specificity, and ethical considerations—provide a high bar and valuable lessons. The novelty of an Indonesian-focused thesis may lie in effectively adapting advanced AI techniques to the unique Indonesian context, considering its specific linguistic landscape, data limitations, and existing infrastructure. The development of local initiatives like the SARI chatbot and advanced Indonesian language models such as SahabatAI [8] are positive indicators of progress and a national commitment to leveraging AI. A thesis can contribute by proposing solutions that are contextually appropriate for Indonesia, perhaps by focusing on specific high-impact consular services or by rigorously evaluating the application of locally developed resources like SahabatAI, while drawing inspiration from global standards in user experience and service delivery goals.

III. Advanced AI Technologies for Consular Question Answering

The choice of AI technology is pivotal for developing an effective consular QA system. While the initial proposal focuses on fine-tuning BERT, the current NLP landscape offers a broader spectrum of powerful models and techniques, including those specifically developed for Indonesian and its dialects, as well as innovative architectures like Retrieval Augmented Generation (RAG) and Knowledge Graphs

(KGs).

A. BERT and Fine-Tuning for Indonesian Language Nuances

BERT and its variants have been foundational in advancing NLP capabilities for numerous languages, including Indonesian.

- **Leveraging IndoBERT, NusaBERT, and other Pre-trained Models:**
  - Several BERT-based models have been specifically pre-trained or fine-tuned for Indonesian. **IndoBERT**, developed as part of the IndoNLU and IndoLEM benchmark initiatives, has demonstrated strong performance on various Indonesian Natural Language Understanding tasks.[5] For instance, IndoBERT-QA, fine-tuned on a translated version of the SQuAD dataset, is designed for question answering in Indonesian.[5]
  - **NusaBERT**, another model trained on large Indonesian text data, has also shown effectiveness, particularly in tasks like sentiment analysis across Indonesian regional languages.[3] Studies comparing these models indicate that those pre-trained specifically on Indonesian data, such as IndoBERT (IndoNLU) and NusaBERT, generally outperform multilingual BERT models for Indonesian-specific tasks.[3] IndoBERT-large (IndoNLU) achieved high F1-scores for sentiment analysis in Javanese, Minangkabau, and Banjar.[3]
  - The **IndoNLU** and **IndoLEM** projects provide not only pre-trained models but also crucial benchmark datasets for evaluating Indonesian NLU capabilities.[6]
- **Addressing Challenges in Low-Resource Indonesian Languages and Dialects:**
  - A significant challenge is that most of Indonesia's over 700 regional languages are low-resource, meaning they have very limited digital data available for training NLP models.[3] This scarcity impacts the performance of models on these languages.
  - Performance of even Indonesian-specific models can vary considerably across different regional languages. For example, one study found Batak Toba to be a particularly challenging language for all evaluated BERT-based models.[3]
  - However, research also shows that fine-tuning on relatively small, task-specific datasets can lead to substantial performance improvements for low-resource languages. A notable example is the sharp improvement in Javanese-to-Indonesian machine translation quality after fine-tuning models on just 500 paired sentences.[4] This finding is highly relevant for a Master's thesis methodology, suggesting that targeted fine-tuning can be effective even with limited consular-specific data in regional dialects.

B. The Potential of SahabatAI Models for Enhanced Consular QA

The recent emergence of SahabatAI models, a collaborative effort by Indonesian tech companies and research institutions, presents a significant opportunity for developing advanced consular QA systems.[8]

- **Overview:** SahabatAI is a collection of Large Language Models (LLMs), including versions based on Gemma2 (e.g., Gemma2 9B CPT Sahabat-AI v1 base) and Llama3 (e.g., Llama3 8B CPT Sahabat-AI v1 Instruct). These models have been extensively pre-trained and, in some cases, instruct-tuned specifically for Bahasa Indonesia, English, and key Indonesian regional dialects such as Javanese and Sundanese.[8]
- **Capabilities:** The SahabatAI models have been evaluated on a range of NLP tasks relevant to QA, including question answering itself, sentiment analysis, translation, and abstractive summarization.[9] The instruct-tuned versions, like Llama3 8B CPT Sahabat-AI v1 Instruct, have been further fine-tuned using a substantial number of instruction-completion pairs in Indonesian, Javanese, Sundanese, and English, making them adept at following instructions and generating coherent responses in these languages.[8]
- **Relevance to Consular QA:** The native support for Bahasa Indonesia and major regional dialects like Javanese and Sundanese makes SahabatAI models prime candidates for the proposed multilingual consular chatbot. Their development involved training on vast amounts of Indonesian-centric data, potentially capturing linguistic nuances and cultural contexts more effectively than generic multilingual models.[9]
- **Limitations and Considerations:** It is important to note that the base versions of SahabatAI models are typically released without extensive safety alignment. The developers explicitly state that users should perform their own safety fine-tuning and implement security measures.[8] This is a critical ethical consideration that any thesis using these models must address thoroughly, especially given the sensitive nature of consular interactions.

The availability of SahabatAI models can be seen as a potential game-changer for supporting Indonesian regional dialects within a consular QA system. One of the most significant hurdles for building multilingual NLP systems in Indonesia is the scarcity of data for its numerous regional languages.[3] SahabatAI's focused pre-training and instruction-tuning on large datasets including Javanese and Sundanese (e.g., the Gemma2 base model used ~50B tokens, with specific data piles for Javanese and Sundanese, and the Llama3 instruct model used hundreds of thousands of instruction pairs in these dialects [8]) directly addresses this data scarcity for these specific major

dialects. This provides a powerful, high-quality foundation upon which a consular QA system can be built. Consequently, the novelty for a thesis leveraging SahabatAI could shift from the foundational task of building language capabilities from scratch for these dialects towards more specialized objectives such as domain adaptation to consular services, developing sophisticated empathetic response generation strategies, or conducting rigorous comparative studies of SahabatAI against other approaches in this specific application context.

C. Exploring Hybrid and Alternative Approaches

Beyond direct fine-tuning of transformer models, hybrid architectures and alternative AI techniques offer compelling advantages for the complexities of consular QA.

1. Retrieval Augmented Generation (RAG) for Dynamic and Factual Consular Information

RAG is an increasingly popular technique that enhances the capabilities of LLMs by connecting them to external, up-to-date knowledge sources at the time of query.[56] Instead of relying solely on the knowledge encoded within the LLM's parameters (which can become outdated), RAG first retrieves relevant information from a specified corpus (e.g., official government documents, FAQs) and then provides this information as context to the LLM to generate an answer.

- **Benefits for Consular Services:** Consular information, such as visa requirements, travel advisories, embassy operating hours, and emergency contact details, is often dynamic and subject to frequent changes. RAG is well-suited for such scenarios because it can draw upon the latest verified agency data, thereby improving the accuracy of responses and reducing the risk of providing outdated or incorrect information (hallucinations).[56] Furthermore, RAG systems can often cite the source of the information used to generate an answer, which can significantly increase user trust and transparency—qualities that are paramount for government services.[58] Updating the knowledge base for a RAG system is typically more straightforward and cost-effective than retraining an entire LLM.[58]
- **Application in Government and Public Services:** RAG is being actively explored and adopted by government agencies globally to provide more accurate and specific answers to citizen queries.[56] For instance, the Singaporean government has even developed a "RAG Playbook" to guide the development of RAG systems for Whole-of-Government projects.[61]
- **Indonesian Context:** Research into RAG applications for the Indonesian language is emerging and shows promise. Studies have explored RAG for tasks

such as enhancing Quran translation with contextual information and developing legal QA systems for Indonesian government regulations.[62]

- **Multilingual RAG:** RAG architectures can be adapted for multilingual scenarios by enabling retrieval from language-specific knowledge bases or by integrating real-time machine translation capabilities within the pipeline.[20]

2. Knowledge Graphs (KGs) for Structured Consular Knowledge Representation

Knowledge Graphs offer another powerful approach, particularly for domains characterized by structured information and complex relationships, such as the legal and administrative rules underpinning many consular services. KGs represent information as a network of entities (e.g., "passport," "visa type," "consular office") and the relationships between them (e.g., "requires document," "is processed at," "has fee").[67]

- **Benefits for Consular Services:** KGs can integrate diverse data sources (databases, documents, websites) into a unified, semantically rich structure. This allows for more precise querying and reasoning over complex consular rules and procedures. For example, a KG could model the dependencies between different application steps or the eligibility criteria for various services.
- **Application in Public Administration and Legal Domains:** KGs have been proposed for modeling event-centric knowledge in public administration to capture the dynamics of citizen interactions with government services.[67] In the Indonesian context, the **LexID** project developed a knowledge graph for Indonesian legal documents, demonstrating the feasibility of applying KG technology to formal Indonesian texts.[70] The **AC-IQuAD** dataset also links Indonesian natural language questions to SPARQL queries over Wikidata, facilitating research in KG-based Question Answering (KGQA).[71]
- **KGQA and GraphRAG:** KGQA involves answering natural language questions by translating them into formal queries (like SPARQL) that can be executed against the KG.[68] A more recent development is **GraphRAG**, which combines the strengths of KGs and RAG. In this approach, the KG serves as a structured and reliable knowledge source from which relevant contextual information is retrieved to augment the LLM's response generation process.[69]

The potential of RAG and KG-based approaches for Indonesian consular services remains largely untapped. While fine-tuning is a valid starting point, a Master's thesis could achieve significant novelty by being among the first to rigorously design, implement, and evaluate a RAG or GraphRAG system specifically for Indonesian consular QA. This would directly address the limitation of "static knowledge" inherent

in purely fine-tuned models, which is a critical concern for dynamic consular information. Such an approach, potentially using SahabatAI as the core generative LLM, would align with international trends in government AI adoption (e.g., Singapore's RAG Playbook [61]) and address the practical need for current, verifiable, and context-aware information delivery in consular services.

D. Comparative Analysis: Fine-Tuning vs. RAG vs. Hybrid Models for the Consular Domain

Choosing the right AI architecture involves understanding the trade-offs between fine-tuning, RAG, and hybrid models:

- **Fine-Tuning:**
  - *Pros:* Can learn domain-specific language styles, jargon, and nuanced understanding if sufficient high-quality, in-domain training data is available. Often results in lower latency at inference time compared to RAG, as no external retrieval step is needed.[59]
  - *Cons:* The knowledge encoded in the model becomes static after training and can quickly become outdated unless the model is frequently retrained, which is costly and resource-intensive. There's a risk of "catastrophic forgetting," where fine-tuning on a specific domain erodes the model's general knowledge. Fine-tuned models may also be more prone to hallucination if queried on topics outside their specific training distribution.[58]
- **Retrieval Augmented Generation (RAG):**
  - *Pros:* Can access and incorporate the most up-to-date information from external knowledge sources, making it ideal for domains with dynamic content like consular regulations. Can cite sources, enhancing transparency and user trust. Generally less prone to factual hallucination for queries covered by the knowledge base. More cost-effective to update the knowledge base than to retrain an entire LLM.[56]
  - *Cons:* The performance is heavily dependent on the quality and comprehensiveness of the knowledge base and the effectiveness of the retrieval component. May have higher inference latency due to the added retrieval step. Might not capture stylistic nuances or highly specific domain language as effectively as a well-fine-tuned model unless the generator LLM itself is also fine-tuned.[58]
- **Hybrid Models (Fine-tuning + RAG):**
  - *Pros:* Aims to combine the best of both worlds. The LLM can be fine-tuned for domain-specific language, style, and foundational knowledge, while RAG provides access to dynamic, factual information at query time.[58]

Parameter-Efficient Fine-Tuning (PEFT) techniques used during the fine-tuning stage might better preserve the LLM's general reasoning abilities, which are beneficial for the RAG component.[72] Research suggests that a smaller, fine-tuned LLM combined with RAG can achieve performance comparable to or even better than a larger, non-fine-tuned LLM with RAG.[72]

- ○ *Cons:* Can be more complex to design, implement, and maintain. Requires careful balancing of the roles of the fine-tuned knowledge versus the retrieved knowledge.

For the consular domain, especially in a low-resource language context and dealing with potentially low-frequency (niche) queries, RAG often demonstrates an advantage over fine-tuning alone, particularly for accessing the "long tail" of factual knowledge.[72] A hybrid approach, where a model like SahabatAI is first fine-tuned on general consular dialogue patterns and then integrated into a RAG architecture accessing specific, up-to-date consular documents, could offer a robust and effective solution.

Table 2: Overview of Key Indonesian NLP Models and Datasets for Consular QA

Resource Type	Name	Languages/ Dialects Covered	Key Features/Task Focus	Potential for Consular QA	Known Limitations/ Data Source
Model	IndoBERT (IndoNLU/IndoLEM variants)	Bahasa Indonesia	General NLU, Sentiment, QA (with fine-tuning)	Base for fine-tuning on consular tasks, especially for standard Indonesian.	May require significant domain-specific fine-tuning. Performance on dialects varies.[3]
Model	NusaBERT	Bahasa Indonesia, some regional languages (in pre-training)	Sentiment analysis, general NLU	Potential for fine-tuning, especially if regional language pre-training is beneficial.	Base version showed inconsistent performance in one study.[3]
Model	SahabatAI (Gemma2 9B	English, Indonesian,	General language	Strong base for	Base model not

	CPT Sahabat-AI v1 base)	Javanese, Sundanese	capabilities, pre-trained on diverse Indonesian data including dialects.	fine-tuning for multilingual consular QA, especially with Javanese/Sundanese. Can be used as generator in RAG.	safety-aligned; requires safety fine-tuning.[9]
Model	SahabatAI (Llama3 8B CPT Sahabat-AI v1 Instruct)	English, Indonesian, Javanese, Sundanese	Instruction-following, QA, translation, summarization. Fine-tuned on instruction pairs in supported languages.	Excellent candidate for fine-tuning or as a generator in RAG for multilingual, instruction-driven consular QA.	Not safety-aligned; requires safety fine-tuning. Potential for hallucination if not grounded.[8]
Dataset	IndoQA	Bahasa Indonesia	Monolingual extractive QA from Indonesian Wikipedia.	Small dataset for initial fine-tuning or evaluation for Indonesian QA.	Limited size (4,413 examples), general domain (Wikipedia).[3][2]
Dataset	AC-IQuAD	Bahasa Indonesia	Automatically constructed QA pairs from Wikidata, questions linked to SPARQL.	Can be used for KGQA or adapted if context is rich. Provides Indonesian questions.	Automatic construction may lead to quality issues. Focus on Wikidata entities.[71]
Dataset	Translated SQuAD v2.0	Bahasa Indonesia	Extractive QA.	Used to fine-tune	Quality dependent

	(Indonesian)	(translated from English)		IndoBERT-QA. Can serve as a larger, albeit translated, QA dataset.	on machine translation accuracy. May lack natural Indonesian phrasing.[5]
Dataset	TyDi QA (Indonesian subset)	Bahasa Indonesia (part of multilingual dataset)	Typologically diverse QA.	Provides Indonesian QA pairs designed for linguistic diversity.	Subset size for Indonesian may be limited. General domain.[5]
Dataset	IndoNLG (QA component)	Indonesian, Javanese, Sundanese	Question Answering (among other NLG tasks).	Provides QA data in target languages, including Javanese and Sundanese.	QA is one of several tasks; dataset size for QA specifically needs checking.[6]
Dataset	NusaX	Indonesian, English, 10 Indonesian local languages (Acehnese, Balinese, Banjarese, Buginese, Madurese, Minangkabau, Javanese, Ngaju, Sundanese, Toba Batak)	Parallel sentiment dataset.	While not directly QA, provides parallel text in regional languages that could be invaluable for understanding dialectal nuances or for pre-training/ fine-tuning multilingual embeddings if adapted.	Focus is on sentiment, not QA. Quality and size for each dialect vary.[3]

Dataset	Official Consular Information (FAQs from KBRI/KJRI, Kemlu, BP2MI)	Primarily Bahasa Indonesia, some English	Consular procedures, regulations, advice.	Primary source for creating a domain-specific consular QA dataset or knowledge base for RAG.	Unstructured or semi-structured; requires significant effort to curate into QA pairs or structured knowledge.[2]

This table highlights the available resources but also underscores a critical gap: the lack of a dedicated, high-quality, publicly available dataset specifically for Indonesian consular question answering, particularly one that includes regional dialects and queries requiring empathetic responses. This data gap represents both a significant challenge and a prime opportunity for a novel contribution within a Master's thesis. Strategies such as scraping official FAQs, leveraging LLMs for initial QA pair generation from consular documents followed by human validation, or focusing on advanced fine-tuning techniques that are effective with limited labeled data will be essential. The creation of even a modest, well-curated, and publicly shared Indonesian consular QA dataset would be a valuable artifact for the research community.

IV. Methodological Framework for a Master's Thesis

Developing a robust methodological framework is essential for a Master's thesis. This involves a clear strategy for dataset acquisition and preparation, model development and integration, and a rigorous experimental design and evaluation protocol.

A. Dataset Strategy: Acquisition, Curation, and Augmentation for Consular QA in Indonesian (and selected dialects)

The success of any NLP model, particularly for a specialized domain like consular QA, hinges on the availability and quality of relevant data. Given the current landscape, a multi-pronged dataset strategy will be necessary.

- **Utilizing Existing Indonesian QA Datasets:**
  - Several general-purpose Indonesian QA datasets can serve as starting points for pre-training, initial fine-tuning, or as sources for transfer learning techniques. These include:
    - **IndoQA:** A monolingual Indonesian QA dataset comprising 4,413 examples from Indonesian Wikipedia.[32] While small, it offers natively Indonesian QA

pairs.

- **AC-IQuAD:** This dataset was automatically constructed by leveraging Wikidata, providing Indonesian questions linked to SPARQL queries. It could potentially be adapted if the associated context paragraphs are sufficiently rich, or used for exploring KGQA approaches.[71]
- **Translated SQuAD v2.0:** Versions of the Stanford Question Answering Dataset translated into Indonesian have been used to fine-tune models like IndoBERT-QA.[5] However, the quality and naturalness of machine-translated datasets are always a concern and must be carefully evaluated.
- **TyDi QA (Indonesian subset):** This multilingual QA dataset includes an Indonesian portion and is designed to cover typologically diverse questions, offering more linguistic variety.[5]
- **IndoNLG Benchmark:** This benchmark for Indonesian Natural Language Generation also includes a question-answering task, potentially providing data in Indonesian, Javanese, and Sundanese.[6]

○ For low-resource language scenarios, a common strategy is to use **machine translation** to translate high-quality English QA datasets (like SQuAD) into the target language (e.g., Indonesian or its dialects). This creates "weakly supervised" training data.[73] While cost-effective, this approach requires careful post-processing or validation to mitigate errors introduced by the translation process.

- **Strategies for Creating a Domain-Specific Corpus (Consular Information):**
  ○ **Source Materials:** The primary sources for consular information will be official government websites and documents. These include:
    - FAQs and informational pages from the websites of Indonesian Embassies (KBRI) and Consulates General (KJRI) worldwide (e.g.[2]).
    - The official Peduli WNI portal.[2]
    - Documents and publications from the Indonesian Ministry of Foreign Affairs (Kementerian Luar Negeri - Kemlu).
    - Resources from the Indonesian Migrant Workers Protection Agency (BP2MI).[12]
    - Relevant Indonesian laws, regulations, and ministerial decrees pertaining to consular affairs and citizen services.
  ○ **Data Collection:** This will likely involve web scraping of official websites to gather text and documents, followed by manual or semi-automated extraction of potential question-answer pairs or relevant passages for a RAG system.
  ○ **Annotation:** If new QA pairs are to be created for fine-tuning, this is the most labor-intensive step. It requires careful annotation of questions, context

passages, and answers. For a Master's thesis, generating a very large, high-quality, manually annotated dataset from scratch is likely too ambitious. The focus might be on:

- Curating and cleaning existing FAQs.
- A smaller-scale, targeted annotation effort for specific types of consular questions or for evaluating empathetic responses.
- Leveraging LLMs to generate synthetic QA pairs from consular documents, followed by human review and validation to ensure quality and relevance.
  - **Data Augmentation:** For low-resource scenarios, particularly for regional dialects or specific types of queries where data is scarce, data augmentation techniques can be explored. These might include back-translation (translating to English and back to Indonesian/dialect), using LLMs for paraphrasing existing questions and answers, or more advanced synthetic data generation methods.[72]

A critical aspect to acknowledge is the current lack of a dedicated, comprehensive, high-quality public dataset for Indonesian consular question answering, especially one that incorporates regional dialects or examples of empathetic interactions. While existing general Indonesian QA datasets [31] provide some foundation, they do not capture the specific language, nuanced topics, and interaction styles prevalent in consular services. The data used to train the SARI chatbot, for instance, is not publicly detailed. This data gap presents a significant challenge but also a substantial opportunity for contribution. A Master's thesis could make a novel contribution by:

1. Systematically curating and releasing even a small but high-quality Indonesian consular QA dataset (perhaps focused on a specific service like passport renewal or Lapor Diri, and potentially including examples in one major regional dialect).
2. Developing and evaluating robust methods for adapting existing general QA datasets to the consular domain using techniques like few-shot learning, domain adaptation fine-tuning, or sophisticated synthetic data generation and filtering. The development of such a dataset, or effective methods for working with limited domain-specific data, would itself be a valuable artifact and contribution to the Indonesian NLP community.

B. Model Development and Fine-Tuning/Integration Strategy

The choice of base model and the strategy for adapting it to the consular domain are central to the methodology.

- **Choice of Base Model:**

- **IndoBERT/NusaBERT:** These models are viable options, particularly if the research questions specifically involve evaluating or extending these established Indonesian BERT variants.[3] They provide solid baselines for standard Indonesian.
- **SahabatAI (Gemma2 or Llama3 based):** Given their state-of-the-art performance, pre-training on Indonesian-centric data, and native support for key regional dialects (Javanese, Sundanese), SahabatAI models are highly recommended candidates.[8] Fine-tuning or using these models within a RAG framework for consular QA would represent a significant and contemporary research effort.
- **Fine-Tuning Approach (if pursued):**
  - The decision between **full fine-tuning** versus **Parameter-Efficient Fine-Tuning (PEFT)** techniques like Low-Rank Adaptation (LoRA) is important.[58] For larger models like SahabatAI, PEFT methods are often preferred as they are computationally less expensive, require less memory, and can sometimes lead to better generalization by preserving more of the original model's capabilities.
  - Fine-tuning will be task-specific, using the curated consular QA dataset (question-context-answer triplets for extractive QA, or question-answer pairs for generative QA).
- **RAG Implementation (if pursued):**
  - **Retriever:** Start with a strong baseline like BM25. Then explore neural retrievers such as Dense Passage Retriever (DPR) or ColBERT.[73] The choice of embedding model for representing queries and documents is crucial; options include Universal Sentence Encoder (USE), Sentence-BERT (SBERT) variants, or potentially Indonesian-specific embedding models if available and effective.[75]
  - **Knowledge Source:** Consular documents, official FAQs, and relevant regulations will need to be processed, chunked, embedded, and stored in a vector database (e.g., FAISS, ChromaDB).
  - **Generator:** A fine-tuned SahabatAI model, a fine-tuned IndoBERT, or other capable LLMs can serve as the generator component, taking the retrieved context and the original query to synthesize an answer.[75]
  - Consider incorporating **advanced RAG techniques** such as iterative retrieval, query rewriting/refinement for ambiguous questions, or a re-ranking step to improve the relevance of retrieved passages before they are fed to the generator.[64]
- **KG-based QA (if chosen as a focus):**
  - This would involve constructing a knowledge graph representing key consular

entities (e.g., visa types, document requirements, office locations) and their relationships, potentially inspired by the LexID project for Indonesian legal documents.[70]
- The QA mechanism could involve translating natural language questions into SPARQL queries or using the KG to provide highly structured context within a RAG pipeline (GraphRAG).

C. Experimental Design and Evaluation Protocol

A rigorous experimental design is necessary to validate the thesis's contributions.

- **Baselines for Comparison:** The performance of the developed system should be compared against several baselines:
  - A simple **rule-based system** if the focus is on very common, static FAQs.
  - A standard **Information Retrieval (IR) model** (e.g., BM25 using Elasticsearch or Whoosh) applied directly to the corpus of consular documents.
  - The **zero-shot or few-shot performance** of the base LLMs (e.g., SahabatAI before any consular-specific fine-tuning) on the QA task.
  - If comparing models: Fine-tuned IndoBERT vs. Fine-tuned SahabatAI.
  - If comparing architectures: Fine-tuning alone vs. RAG alone vs. a Hybrid model (Fine-tuning + RAG).[58]
- **Evaluation Metrics:** A comprehensive set of metrics is needed to assess different facets of the QA system's performance:
  - **Accuracy-based Metrics (primarily for extractive QA):**
    - **Exact Match (EM):** The percentage of predictions that match the ground truth answer exactly.[31]
    - **F1-Score:** The harmonic mean of precision and recall, considering token overlap between prediction and ground truth. More lenient than EM.[31]
  - **Text Generation Quality Metrics (for abstractive QA, RAG-generated answers):**
    - **ROUGE (Recall-Oriented Understudy for Gisting Evaluation):** Measures n-gram overlap, particularly ROUGE-L (longest common subsequence) is common.[76]
    - **BLEU (Bilingual Evaluation Understudy):** Typically used for machine translation, but can be adapted if QA is framed as generating a target answer string, or for evaluating translated components.[4]
    - **METEOR (Metric for Evaluation of Translation with Explicit Ordering):** Another translation metric considering synonymy and stemming.[76]
    - **BERTScore:** Uses contextual embeddings from BERT to compare the semantic similarity between generated and reference answers, often

correlating better with human judgment than n-gram based metrics.[4]

- **Semantic Similarity:** Cosine similarity between the embeddings of generated answers and reference answers, using a robust multilingual embedding model.[64]
- **Human Evaluation:** This is indispensable, especially for a consular QA system where nuance, cultural appropriateness, and empathy are important. Human evaluators should assess:
  - **Correctness/Accuracy:** Factual accuracy of the provided information.
  - **Fluency:** Grammatical correctness and naturalness of the language.
  - **Coherence:** Logical flow and understandability of the response.
  - **Relevance:** How well the answer addresses the user's question.
  - **Completeness:** Whether the answer provides all necessary information.
  - **Empathy and Cultural Appropriateness:** Particularly for sensitive queries or when interacting in regional dialects. This requires carefully designed rubrics and annotators familiar with Indonesian culture and the specific dialects.[64]
  - **User Satisfaction:** Surveys or feedback mechanisms to gauge overall user experience.[53]
- **For RAG Systems:**
  - **Retrieval Metrics:** Mean Reciprocal Rank (MRR), Precision@k, Recall@k to evaluate the effectiveness of the retriever component in finding relevant documents/passages.[73]
  - **Faithfulness:** Assessing whether the generated answer is factually consistent with and grounded in the retrieved context.[78]

Standard QA metrics like EM and F1 are well-suited for factual, extractive questions, but they fall short in evaluating the more nuanced aspects required for a consular chatbot, such as the ability to generate empathetic or culturally appropriate responses. This is particularly relevant given the inspiration from the SARI chatbot's "humanist AI" approach.[16] Therefore, a significant component of the thesis's evaluation methodology must involve human assessment. Developing or adapting a specific framework and rubrics for evaluating these qualitative dimensions in the Indonesian consular context could itself be a novel contribution. This might involve defining clear criteria for what constitutes an "empathetic" or "culturally sensitive" response within this domain, and training human evaluators to apply these criteria consistently.[64]

D. Technical Implementation Details and Tools

The practical implementation of the thesis will rely on a range of standard NLP tools

and resources:

- **Programming Languages:** Python will be the primary language.
- **Core NLP/ML Frameworks:**
  - **Hugging Face Transformers library:** Provides access to a vast number of pre-trained models (including IndoBERT, SahabatAI variants, multilingual models), tokenizers, and fine-tuning scripts.[7]
  - **PyTorch or TensorFlow:** As the underlying deep learning frameworks for model training and inference.
- **Libraries for RAG/KG (if applicable):**
  - **LangChain or LlamaIndex:** Frameworks for building applications with LLMs, including RAG pipelines.
  - **Vector Databases:** FAISS, Milvus, ChromaDB, or Pinecone for storing and searching document embeddings in RAG systems.
  - **Graph Databases (for KG):** Neo4j [69] or GraphDB [82] if a knowledge graph component is developed.
  - **spaCy:** For efficient text processing tasks like named entity recognition or dependency parsing, if needed.[7]
- **Computational Resources:** Fine-tuning large language models, even with PEFT, can be computationally intensive and require access to GPUs. Resources like Google Colab Pro, university high-performance computing clusters, or cloud computing platforms (AWS, GCP, Azure) will likely be necessary.[8]

Table 3: Proposed Evaluation Metrics for the Consular QA System

Metric Category	Specific Metric	Description	Relevance to Consular QA Success
Automated – Accuracy (Extractive QA)	Exact Match (EM)	Percentage of predictions that exactly match the ground truth answer.[76]	Measures precision for factual, short-answer questions (e.g., "What is the fee?").
	F1-Score	Harmonic mean of precision and recall based on token overlap.[76]	More robust measure of factual accuracy, allowing partial matches.

Automated - Generation Quality (Abstractive/RAG QA)	ROUGE (esp. ROUGE-L)	Measures n-gram overlap, focusing on recall and longest common subsequence.[76]	Assesses how much of the reference answer's content is captured in generated responses.
	BLEU	Measures n-gram precision against reference translations/generations.[4]	Useful if comparing against human-written ideal responses or for translated elements.
	BERTScore	Semantic similarity using contextual embeddings from BERT.[4]	Captures semantic correctness even if wording differs; correlates better with human judgment.
	METEOR	Considers synonymy, stemming, and word order.[76]	Provides a more nuanced assessment of generation quality than simple n-gram overlap.
Automated - Retrieval (for RAG systems)	Mean Reciprocal Rank (MRR@k)	Average of the reciprocal ranks of the first correct retrieved document.[73]	Evaluates how quickly the retriever finds a relevant document.
	Recall@k	Proportion of questions for which a relevant document is found within the top k retrieved documents.[73]	Measures the retriever's ability to find relevant information.
Human Evaluation	Correctness	Factual accuracy of the information provided in the response.	Fundamental for trustworthy consular information.
	Fluency	Grammaticality and naturalness of the	Ensures responses are easy to

		generated language.	understand.
	Coherence	Logical consistency and flow of the response.	Important for complex explanations or multi-turn dialogues.
	Relevance	How well the response addresses the specific user query.	Ensures the system is helpful and on-topic.
	Completeness	Whether the response provides all necessary information for the query.	Critical for users seeking comprehensive guidance.
	Empathy Score	Assesses the perceived empathy, politeness, and supportive tone of the response (using a defined rubric).[16]	Crucial for sensitive queries and building user trust, especially for vulnerable users.
	Cultural Appropriateness	Evaluates if the language and tone are culturally suitable for Indonesian users, including dialect speakers.	Avoids misunderstandings and ensures respectful interaction.
	User Satisfaction	Overall satisfaction with the interaction and the helpfulness of the system (e.g., via Likert scales).[53]	Holistic measure of the system's real-world utility.
RAG-Specific (Human or Automated)	Faithfulness/ Groundedness	Whether the generated answer is factually consistent with the retrieved documents.[78]	Crucial for reducing hallucinations and ensuring reliability in RAG systems.

This comprehensive evaluation plan, combining automated metrics with essential human judgment, will allow for a nuanced assessment of the consular QA system's performance, particularly its ability to deliver accurate, relevant, and empathetically appropriate responses.

V. Novelty, Contribution, and Ethical Imperatives

A successful Master's thesis must not only demonstrate technical proficiency but also make a clear novel contribution to its field and address the ethical implications of the research. For a project focused on AI for Indonesian consular services, these aspects are particularly salient.

A. Identifying and Articulating the Novelty and Contribution to NLP and Consular Services

The novelty of a Master's thesis in a well-established field like NLP often arises from applying or adapting existing techniques to new, challenging domains or languages, or by introducing new dimensions to the problem. Research contributions can generally be categorized as either **artifacts** (new methods, models, datasets, or tasks) or **knowledge** (new insights about methods, datasets, tasks, language, or people).[47]

For this thesis, potential areas of novelty and contribution include:

1. **Application of State-of-the-Art Indonesian LLMs to the Consular Domain:** A comprehensive study fine-tuning and evaluating models like **SahabatAI** for the specific nuances of Indonesian consular question answering would be a novel application. While SahabatAI is designed for Indonesian and its dialects [8], its application to the breadth of consular services is unexplored. (Contribution: Artifact - new method/application; Knowledge - about method performance).
2. **Robust Multilingual System with Regional Dialect Support:** Moving beyond standard Bahasa Indonesia to effectively support major regional dialects (e.g., Javanese, Sundanese) for a range of consular queries. This would involve leveraging SahabatAI's capabilities or developing novel fine-tuning strategies for dialectal adaptation in this domain, potentially extending beyond SARI's current scope or employing more advanced techniques. (Contribution: Artifact - new method/enhanced model; Knowledge - about language processing for dialects).
3. **Tailored Hybrid RAG/Fine-tuning Architecture for Indonesian Consular QA:** Designing, implementing, and rigorously evaluating a hybrid architecture that combines the strengths of fine-tuning (for style and basic domain knowledge) with RAG (for dynamic, factual information from official Kemlu/KBRI sources)

specifically for the Indonesian consular context. (Contribution: Artifact - new method/architecture; Knowledge - comparative performance).

4. **Development of Empathetic and Culturally Aware QA Capabilities:** Researching and implementing techniques to enable the QA system to generate responses that are not only factually correct but also empathetic, polite, and culturally sensitive, particularly for distressing or complex consular situations. This could involve novel prompting strategies, fine-tuning on appropriately styled data, or developing new evaluation rubrics for these qualitative aspects. (Contribution: Artifact - new method/task definition; Knowledge - about task/human-AI interaction).

5. **Creation of a New Consular QA Dataset for Indonesian:** Addressing the critical data gap by curating, annotating, and potentially releasing a new dataset for Indonesian consular question answering. Even a moderately sized, high-quality dataset focused on specific services or incorporating examples of empathetic interactions or regional dialect queries would be a significant contribution to the Indonesian NLP community.[83] (Contribution: Artifact - new dataset).

6. **Rigorous Comparative Analysis:** Conducting a thorough comparative study of different approaches (e.g., IndoBERT vs. SahabatAI, fine-tuning vs. RAG vs. hybrid) on the specific, low-resource task of Indonesian consular QA, providing valuable benchmarks and insights for future research and development. (Contribution: Knowledge - about methods).

The novelty here stems from the **specialization and contextualization** of AI techniques. While QA is a general NLP task, applying these methods to the unique requirements of Indonesian consular services—with its specific linguistic diversity (Indonesian and its dialects being relatively low-resource for specialized domains [3]), the sensitive nature of many queries (requiring empathy, as highlighted by SARI [16]), and the need for culturally appropriate interaction—constitutes a novel research endeavor. The thesis can claim originality not merely by using a new model, but by demonstrating how that model is customized, integrated, and evaluated to meet these specific and challenging contextual demands.

B. Ethical Considerations in Deploying AI for Consular Services

The deployment of AI in public services, particularly those involving direct citizen interaction and sensitive personal information like consular affairs, carries significant ethical responsibilities.[49] These must be integral to the research and system design, not an afterthought.

Key ethical principles, largely drawn from guidelines for AI in government, include [18]:

- **Transparency:** Citizens must be clearly informed when they are interacting with an AI system versus a human agent. If the system provides information or makes recommendations, the sources or reasoning (where possible) should be explainable. RAG systems that can cite sources offer an advantage here.[58]
- **Accountability:** Clear mechanisms for oversight, responsibility for the AI's actions, and avenues for redress in case of errors, bias, or harm are essential. Human oversight, especially for critical decisions or complex cases, must be maintained.[18]
- **Fairness and Non-Discrimination:** The AI system must be designed and trained to avoid perpetuating or amplifying existing societal biases related to language, ethnicity, gender, origin, or other protected characteristics. This is crucial for a diverse user base like Indonesian citizens abroad. It requires careful dataset curation, bias detection in training data and model outputs, and mitigation strategies.[18] The SARI chatbot's reported use of gender bias-free data is a relevant precedent.[17]
- **Privacy and Data Security:** Consular interactions frequently involve the exchange of highly sensitive personal information (e.g., passport details, visa applications, personal circumstances in emergency cases). Robust data governance frameworks, strong encryption, secure data handling practices, and strict compliance with relevant data protection regulations (such as Indonesia's Personal Data Protection Law and potentially GDPR if EU citizens' data is involved) are paramount.[18]

Specific ethical challenges for a consular chatbot include:

- **Accuracy and Reliability of Information:** Providing incorrect or misleading consular advice can have severe real-world consequences for citizens (e.g., visa denials, legal issues, missed opportunities).[18] The system must be designed for high factual accuracy.
- **Cultural Sensitivity and Linguistic Appropriateness:** Especially in a multilingual system supporting diverse dialects, the language used must be culturally appropriate and respectful, avoiding offense or misinterpretation.[13]
- **Handling of Sensitive and Crisis Situations:** The system must have clear protocols for identifying and handling highly sensitive queries or users in distress. This includes providing empathetic initial responses (if designed for this) and, most importantly, seamlessly and quickly escalating such cases to human consular staff.[18]
- **Model Safety and Alignment:** If using powerful base LLMs like SahabatAI, which

are released without full safety alignment, the thesis must explicitly address how safety measures (e.g., safety-specific fine-tuning, robust input/output filtering, guardrails) will be implemented and evaluated to prevent harmful, biased, or inappropriate outputs.[8]

Mitigation Strategies to be incorporated into the thesis methodology:

- **Dataset Curation:** Scrutinize training data for biases. If creating new data, use diverse annotators and guidelines that promote fairness and cultural sensitivity.
- **Model Training and Evaluation:** Regularly test for biases in model outputs across different demographic subgroups (if identifiable). Implement fairness-aware training techniques if appropriate.
- **Human-in-the-Loop (HITL):** Design clear pathways for escalating complex, sensitive, or ambiguous queries to human consular officers. The AI should augment, not replace, human expertise in critical situations.
- **Disclaimers and User Education:** Clearly communicate the AI's capabilities and limitations to users.
- **Adherence to Ethical AI Frameworks:** Align the project with established ethical AI principles, such as those outlined by UNESCO's Recommendation on the Ethics of Artificial Intelligence (RAM AI).[85]

Ethical design should be a core component of the research contribution. A thesis in this domain must deeply integrate these ethical considerations from the project's inception. For example, investigating methods for detecting and mitigating bias in Indonesian language data used for fine-tuning, or proposing system architectures that enhance transparency (like RAG citing its sources), would be valuable contributions. Ensuring the system does not inadvertently disadvantage speakers of certain regional dialects due to imbalances in training data is a key ethical and technical challenge to address. A robust discussion of these ethical challenges and the proposed solutions within the thesis will significantly enhance its impact and relevance.

Table 4: Potential Areas of Novelty and Contribution for the Master's Thesis

Area of Focus	Specific Contribution	Potential Impact
Low-Resource Dialect Support for Consular Needs	Fine-tuning/adapting SahabatAI for robust consular QA in Javanese and/or Sundanese, beyond basic FAQ. Developing strategies for	Improved accessibility and service quality for millions of Indonesian citizens who are native speakers of these dialects. Enhanced inclusivity

	handling code-switching or dialectal variations in user queries.	of government digital services.
Hybrid Model Architecture (Fine-tune + RAG/KG) for Indonesian Consular QA	Designing and evaluating a novel hybrid model combining fine-tuned SahabatAI (for style/empathy/core knowledge) with a RAG system accessing dynamic Kemlu/KBRI FAQs and documents, or a KG of consular regulations.	More accurate, up-to-date, and contextually relevant answers for a wider range of consular queries. Reduced model hallucination. Increased user trust through source citation.
Empathetic & Culturally-Aware Response Generation	Developing and evaluating prompting techniques, fine-tuning strategies, or persona-based generation for empathetic and culturally appropriate responses in Bahasa Indonesia and selected dialects, particularly for sensitive consular topics (e.g., migrant worker issues, emergencies).	Higher user satisfaction and trust, especially for vulnerable users. More effective communication in stressful situations. Contribution to human-centered AI design for public services.
Domain-Specific Consular QA Dataset Curation for Indonesian	Systematically curating, annotating, and releasing a new, high-quality dataset for Indonesian consular QA (e.g., focusing on passport/visa FAQs, Lapor Diri, or common PMI issues), potentially including parallel data in a major dialect or examples of empathetic interactions.	Provision of a valuable public resource for future research and development in Indonesian NLP and AI for public services. Enabling more robust model training and evaluation.
Comparative Analysis of SOTA Indonesian LLMs on Consular Tasks	Conducting a rigorous benchmark comparison of leading Indonesian LLMs (e.g., different SahabatAI versions, IndoBERT, NusaBERT) on a standardized set of consular QA tasks, evaluating performance across various	Evidence-based guidance for model selection in future Indonesian government AI projects. Deeper understanding of the capabilities and limitations of current Indonesian LLMs for specialized domains.

	metrics including accuracy, fluency, and potentially dialectal robustness.	
Ethical AI Framework and Bias Mitigation for Consular Chatbots	Investigating and proposing specific methods for identifying and mitigating linguistic or demographic biases in Indonesian training data for consular AI. Developing a framework for ensuring fairness, transparency, and accountability in the context of the proposed system.	More equitable and trustworthy AI-driven consular services. Contribution to responsible AI development practices for low-resource languages and diverse populations.

These areas offer avenues for a Master's thesis to make significant and recognized contributions, aligning with the types of advancements valued in the NLP research community.[47]

VI. Conclusion and Strategic Recommendations for the Thesis

The research idea of developing an AI-powered question-answering system for Indonesian consular services presents a compelling and impactful avenue for a Master's thesis. The analysis indicates that such a system is not only viable but also aligns with current technological advancements and pressing societal needs. Key insights reveal the critical importance of multilingual support, including regional dialects, to ensure inclusivity for all Indonesian citizens abroad. The SARI chatbot and the SahabatAI models serve as powerful local precedents and enablers for such an endeavor. Furthermore, the dynamic and often sensitive nature of consular information suggests that hybrid AI architectures, particularly those combining the contextual awareness of fine-tuned models with the factual grounding of Retrieval Augmented Generation (RAG) or Knowledge Graphs (KGs), hold significant promise. However, the path to developing such a system is underscored by the crucial challenges of dataset availability for this specific domain and the paramount importance of embedding ethical considerations into every stage of research and development.

Based on this comprehensive analysis, the following strategic recommendations are offered for shaping the Master's thesis:

1. **Refine the Research Focus with Specificity:**

- **Option A (Dialect-Focused, SahabatAI-centric):** "Enhancing Consular Service Accessibility for Javanese and Sundanese Speaking Indonesian Citizens Abroad through Fine-Tuned SahabatAI Question Answering." This leverages SahabatAI's strengths and addresses a clear inclusivity gap.
      - **Option B (RAG for Dynamic Information):** "A Retrieval-Augmented Generation System using SahabatAI for Providing Accurate and Up-to-Date Information on Indonesian Passport and Lapor Diri Services." This tackles the challenge of dynamic consular information and emphasizes transparency.
      - **Option C (Empathetic QA for Sensitive Topics):** "Developing and Evaluating Empathetic Question Answering Capabilities for Indonesian Migrant Worker Support using SahabatAI, Focusing on Common Labor Issues." This builds on SARI's humanist AI concept and addresses a vulnerable group.
2. **Prioritize SahabatAI as the Core LLM:** Given its advanced architecture, pre-training on Indonesian and key regional dialects (Javanese, Sundanese), and active development, SahabatAI [8] should be the primary LLM for fine-tuning or as the generator in a RAG system. This positions the thesis at the forefront of Indonesian NLP research.
3. **Adopt a Pragmatic and Novel Dataset Strategy:**
   - Acknowledge the lack of a dedicated public Indonesian consular QA dataset as both a challenge and an opportunity for contribution.
   - Combine existing general Indonesian QA resources (e.g., IndoQA, translated SQuAD [31]) for initial model adaptation if suitable.
   - Focus primary data efforts on curating a high-quality, albeit potentially smaller, domain-specific corpus by:
     - Systematically scraping and processing FAQs and informational content from official KBRI/KJRI websites, Kemlu, and BP2MI.
     - Employing LLMs (like SahabatAI itself in a few-shot setting) to generate initial QA pairs from these documents, followed by rigorous human validation and refinement.
   - If pursuing dialectal support, leverage SahabatAI's inherent capabilities and consider small-scale, targeted data collection or augmentation for specific consular scenarios in Javanese or Sundanese.
   - The release of any curated dataset, however modest, would be a valuable contribution.
4. **Emphasize Hybrid Architectures and Rigorous Evaluation:**
   - Strongly consider a hybrid approach (e.g., PEFT fine-tuned SahabatAI + RAG) to balance nuanced understanding with access to dynamic, factual information.

- The evaluation protocol must be comprehensive, incorporating automated metrics (F1, ROUGE, BERTScore [76]) and, crucially, **human evaluation**. Human assessment is vital for judging correctness, fluency, relevance, and particularly the qualitative aspects of empathy and cultural appropriateness, especially if the thesis aims to address these dimensions.[64]
5. **Integrate Ethical Considerations as a Core Research Component:**
   - Proactively address potential biases in data and models, especially concerning language, dialect, and user demographics.
   - Incorporate strategies for ensuring fairness, transparency (e.g., source citation in RAG), and accountability.
   - Thoroughly discuss data privacy and security measures for handling potentially sensitive user queries.
   - If using SahabatAI base models, detail the steps taken for safety alignment and responsible deployment within the research context.[8]
   - A dedicated chapter or section on the ethical framework and impact assessment will strengthen the thesis.

Looking beyond the immediate scope of a Master's thesis, this research can pave the way for several future avenues. These include the expansion of the QA system to encompass a wider array of Indonesian regional dialects, deeper integration with backend consular databases and systems to enable transactional capabilities (e.g., initiating an application process), and longitudinal studies to assess real-world user adoption, trust, and the long-term impact on consular service efficiency. Furthermore, AI-driven analytics on query patterns could eventually enable proactive consular assistance, anticipating citizens' needs before they even ask.

By carefully defining its scope, leveraging state-of-the-art Indonesian NLP resources, adopting a robust methodology, and foregrounding ethical principles, the proposed Master's thesis can make a significant contribution to both the field of AI and the enhancement of public services for Indonesian citizens worldwide.

Karya yang dikutip

1. Tutorial Pengisian Lapor Diri di Website Peduli WNI - e-Consular Service KBRI WDC, diakses Mei 8, 2025, https://consular.embassyofindonesia.org/lapordiri/intro.html
2. Peduli WNI - Kemenlu, diakses Mei 8, 2025, https://peduliwni.kemlu.go.id/faq.html
3. pdfs.semanticscholar.org, diakses Mei 8, 2025, https://pdfs.semanticscholar.org/0e95/b4b6dcdfd9952da9cfa7da512b88c6600499.pdf
4. Fine-Tuning Small Language Models for Indonesian Languages, Trivan Menezes,

UG '25 (19115) - Princeton Media Central, diakses Mei 8, 2025,
https://mediacentral.princeton.edu/media/Fine-Tuning+Small+Language+Models+
for+Indonesian+Languages%2C+Trivan+Menezes%2C+UG+25+%2819115%29/1_n
il56g8g

5. Comparative Analysis of Encoder-Based Pretrained Models: Investigating the
Performance of BERT Variants in Indonesian Question-Answering -
ResearchGate, diakses Mei 8, 2025,
https://www.researchgate.net/publication/384786769_Comparative_Analysis_of_E
ncoder-Based_Pretrained_Models_Investigating_the_Performance_of_BERT_Varia
nts_in_Indonesian_Question-Answering

6. Projects - IndoNLP, diakses Mei 8, 2025, https://indonlp.github.io/projects.html

7. Natural language processing models in 2025 | Pre-trained NLP models | NLP
solutions for businesses | Lumenalta, diakses Mei 8, 2025,
https://lumenalta.com/insights/7-of-the-best-natural-language-processing-mod
els-in-2025

8. GoToCompany/llama3-8b-cpt-sahabatai-v1-instruct · Hugging Face, diakses Mei
8, 2025,
https://huggingface.co/GoToCompany/llama3-8b-cpt-sahabatai-v1-instruct

9. GoToCompany/gemma2-9b-cpt-sahabatai-v1-base · Hugging Face, diakses Mei
8, 2025,
https://huggingface.co/GoToCompany/gemma2-9b-cpt-sahabatai-v1-base

10. General Information - e-Consular Service KBRI WDC, diakses Mei 8, 2025,
https://consular.embassyofindonesia.org/page/generalinformation.html

11. Indonesia Travel Advice & Safety | Smartraveller, diakses Mei 8, 2025,
https://www.smartraveller.gov.au/destinations/asia/indonesia

12. Governance and the Protection of Indonesian Migrant Workers in Malaysia,
diakses Mei 8, 2025,
https://socialpolicyworldwide.org/post/governance_and_the_protection_of_indon
esian_migrant_workers_in_malaysia_41

13. Chatbots and Multilingual Support: Enhancing User … - SmythOS, diakses Mei 8,
2025,
https://smythos.com/ai-agents/chatbots/chatbots-and-multilingual-support/

14. Multilingual Chatbots: Benefits, Challenges, and Creation Guide - Machine
Translation, diakses Mei 8, 2025,
https://www.machinetranslation.com/blog/multilingual-chatbot

15. [2203.13357] One Country, 700+ Languages: NLP Challenges for
Underrepresented Languages and Dialects in Indonesia - arXiv, diakses Mei 8,
2025, https://arxiv.org/abs/2203.13357

16. Indonesian Ministry Of Foreign Affairs And UN Women Launch … - VOI, diakses
Mei 8, 2025, https://voi.id/en/news/477218

17. Indonesia, UN Women Launch AI Chatbot SARI to Protect Female …, diakses Mei
8, 2025,
https://www.jakartadaily.id/international/16214999370/indonesia-un-women-launc
h-ai-chatbot-sari-to-protect-female-migrant-workers

18. Multilingual Chatbot: Benefits, Challenges & How to Build | WotNot, diakses Mei 8,

2025, https://wotnot.io/blog/multilingual-chatbot

19. Multilingual Chatbots : Know All Benefits & Features (2025) - ProProfs Chat, diakses Mei 8, 2025, https://www.proprofschat.com/blog/multilingual-chatbots/

20. In what ways can chatbots support multilingual communication for global businesses?, diakses Mei 8, 2025, https://www.quora.com/In-what-ways-can-chatbots-support-multilingual-communication-for-global-businesses

21. What are the challenges faced in designing multilingual chatbots? - Quora, diakses Mei 8, 2025, https://www.quora.com/What-are-the-challenges-faced-in-designing-multilingual-chatbots

22. Services for U.S. and Local Citizens - U.S. Embassy & Consulates in Indonesia, diakses Mei 8, 2025, https://id.usembassy.gov/services/

23. 10 Commonly Asked Indonesia Immigration Visa Questions, diakses Mei 8, 2025, https://doortoid.com/immigration-visa-questions/

24. General Information & FAQ - The Official eVisa website for Indonesia, diakses Mei 8, 2025, https://evisa.imigrasi.go.id/front/faqs/university

25. Konsuler dan layanan lainnya untuk warga negara Australia, diakses Mei 8, 2025, https://indonesia.embassy.gov.au/jaktindonesian/consular.html

26. Ketentuan Barang Pindahan - Official Website Direktorat Jenderal Bea dan Cukai, diakses Mei 8, 2025, https://www.beacukai.go.id/faq/ketentuan-barang-pindahan.html

27. FAQ - Kedutaan Besar Republik Indonesia Den Haag, Belanda, diakses Mei 8, 2025, https://indonesia.nl/faq/

28. Singapore, diakses Mei 8, 2025, https://kemlu.go.id/singapore/faq

29. KBRI Den Haag - Peduli WNI - Kemenlu, diakses Mei 8, 2025, https://peduliwni.kemlu.go.id/informasi_pelayanan/app/detail_kbri/.html?perwakilan_id=NTI4MA==

30. Lapor diri – Indonesia.nl - Kedutaan Besar Republik Indonesia Den Haag, Belanda, diakses Mei 8, 2025, https://www.indonesia.nl/id/konsuler/lapor-diri

31. Indobert QA · Models - Dataloop AI, diakses Mei 8, 2025, https://dataloop.ai/library/model/rifky_indobert-qa/

32. SEACrowd/indoqa · Datasets at Hugging Face, diakses Mei 8, 2025, https://huggingface.co/datasets/SEACrowd/indoqa

33. Bye Indonesia - Renunciation of Indonesian Citizenship Guide 2024, diakses Mei 8, 2025, https://www.byeindonesia.com/

34. The Embassy of the Republic of Indonesia - Washington, diakses Mei 8, 2025, https://kemlu.go.id/washington/tentang-perwakilan/kontak-kami

35. VFS Global Unveils AI Chatbot for UK Visa Applicants in 141 Countries - travelobiz, diakses Mei 8, 2025, https://travelobiz.com/vfs-global-unveils-ai-chatbot-for-uk-visa-applicants-in-141-countries/

36. Frequently Asked Questions (FAQs) About Indonesian Renunciation Process - Bye Indonesia, diakses Mei 8, 2025, https://www.byeindonesia.com/faq/

37. Kerja di luar negeri? Pastikan menjadi Pekerja Migran Indonesia yang Prosedural.,

diakses Mei 8, 2025,
https://kanimbatam.kemenkumham.go.id/berita/2020/12/kerja-di-luar-negeri-pastikan-menjadi-pekerja-migran-indonesia-yang-prosedural

38. Strategi KBRI Kuala Lumpur Dalam Perlindungan Kewarganegaraan: Studi WNI Tidak Berdokumen di Malaysia Tahun 2022–2023 - Penerbit, diakses Mei 8, 2025, https://jayapanguspress.penerbit.org/index.php/ganaya/article/download/2984/1353/10075

39. (PDF) Strategies of the Indonesian Embassy in Kuala Lumpur in Resolving Education Problems for Undocumented Children of Indonesian Migrant Workers in Malaysia - ResearchGate, diakses Mei 8, 2025, https://www.researchgate.net/publication/379359140_Strategies_of_the_Indonesian_Embassy_in_Kuala_Lumpur_in_Resolving_Education_Problems_for_Undocumented_Children_of_Indonesian_Migrant_Workers_in_Malaysia

40. FAQ - Riyadh, diakses Mei 8, 2025, https://kemlu.go.id/riyadh/faq

41. tbinternet.ohchr.org, diakses Mei 8, 2025, https://tbinternet.ohchr.org/_layouts/15/TreatyBodyExternal/DownloadDraft.aspx?key=WfmIQ6P9+ijQJ7rayk9DNY6HQ7iVwh7ubumUsKxf8cYV4brlnqh6nyHkZNK+tepb

42. Informasi Pengaduan - BP2MI - Badan Perlindungan Kerja Migran Indonesia, diakses Mei 8, 2025, https://aim.bp2mi.go.id/pengaduan/info_adu

43. BP2MI | BADAN PELINDUNGAN PEKERJA MIGRAN INDONESIA, diakses Mei 8, 2025, https://bp2mi.go.id/

44. New Clarity for Indonesian Citizens Abroad: Understanding Ministry of Law Regulation 6/2025, diakses Mei 8, 2025, https://www.rajahtannasia.com/wp-content/uploads/2025/03/2025_03_19-New-Clarity-for-Indonesian-Citizens-Abroad.pdf

45. New Clarity for Indonesian Citizens Abroad: Understanding Ministry of Law Regulation 6/2025 - Rajah & Tann Asia, diakses Mei 8, 2025, https://www.rajahtannasia.com/viewpoints/new-clarity-for-indonesian-citizens-abroad-understanding-ministry-of-law-regulation-6-2025/

46. Peduli WNI - Kemenlu, diakses Mei 8, 2025, https://peduliwni.kemlu.go.id/

47. The Nature of NLP: Analyzing Contributions in NLP Papers - arXiv, diakses Mei 8, 2025, https://arxiv.org/html/2409.19505v1

48. How Governments are Using AI: 8 Real-World Case Studies, diakses Mei 8, 2025, https://blog.govnet.co.uk/technology/ai-in-government-case-studies

49. Ethical and Responsible AI Adoption in Government - REI Systems, diakses Mei 8, 2025, https://www.reisystems.com/roadmap-to-transformation-the-next-generation-of-government-operations-with-ethical-and-responsible-ai-adoption/

50. Public Sector: Streamlining Citizen Services with AI-Powered Automation - Intalio.com, diakses Mei 8, 2025, https://www.intalio.com/blogs/public-sector-streamlining-citizen-services-with-ai-powered-automation/

51. In pursuit of seamless service: how AI is revolutionising citizen-government interactions, diakses Mei 8, 2025,

https://www.globalgovernmentforum.com/in-pursuit-of-seamless-service-how-ai-is-revolutionising-citizen-government-interactions/

52. (PDF) Chatbots: A Tool to Improve Public Service Delivery and Create Public Value, diakses Mei 8, 2025, https://www.researchgate.net/publication/389061226_Chatbots_A_Tool_to_Improve_Public_Service_Delivery_and_Create_Public_Value

53. Achieving effectiveness of public service in AI-enabled service from ..., diakses Mei 8, 2025, https://www.emerald.com/insight/content/doi/10.1108/tg-01-2025-0004/full/html

54. AI in public sector: A call for digital transformation - Fakultas Ilmu Administrasi UI, diakses Mei 8, 2025, https://fia.ui.ac.id/en/ai-in-public-sector-a-call-for-digital-transformation/

55. Navigating AI and superapp adoption in Indonesia's public sector - GovInsider, diakses Mei 8, 2025, https://govinsider.asia/intl-en/article/navigating-ai-and-superapp-adoption-in-indonesias-public-sector

56. What Is Retrieval Augmented Generation, and How Are State and Local Agencies Using It?, diakses Mei 8, 2025, https://statetechmagazine.com/article/2025/02/what-is-rag-perfcon

57. Get to Know about Retrieval-Augmented Generation (RAG) – Revolutionizing Natural Language Processing - Sampoerna University, diakses Mei 8, 2025, https://www.sampoernauniversity.ac.id/news/get-to-know-about-retrieval-augmented-generation-rag-revolutionizing-natural-language-processing

58. Comparing Retrieval Augmented Generation and fine-tuning - AWS Prescriptive Guidance, diakses Mei 8, 2025, https://docs.aws.amazon.com/prescriptive-guidance/latest/retrieval-augmented-generation-options/rag-vs-fine-tuning.html

59. RAG vs. fine-tuning - Red Hat, diakses Mei 8, 2025, https://www.redhat.com/en/topics/ai/rag-vs-fine-tuning

60. LLM Fine-Tuning vs Retrieval-Augmented Generation (RAG) - F22 Labs, diakses Mei 8, 2025, https://www.f22labs.com/blogs/llm-fine-tuning-vs-retrieval-augmented-generation-rag/

61. Retrieval-Augmented Generation Playbook | Singapore Government Developer Portal, diakses Mei 8, 2025, https://www.developer.tech.gov.sg/guidelines/standards-and-best-practices/retrieval-augmented-generation-playbook

62. Granularity-aware legal question answering: a case study of Indonesian government regulations | Faisal | International Journal of Advances in Intelligent Informatics, diakses Mei 8, 2025, https://mail.ijain.org/index.php/IJAIN/article/view/1105/0

63. Mastering RAG Implementation: Covering all the Basics - Signity Solutions, diakses Mei 8, 2025, https://www.signitysolutions.com/blog/mastering-rag-implementation?hsLang=en

64. arxiv.org, diakses Mei 8, 2025, https://arxiv.org/html/2504.16121
65. UTILIZING RETRIEVAL-AUGMENTED GENERATION IN LARGE LANGUAGE MODELS TO ENHANCE INDONESIAN LANGUAGE NLP - EJournal Universitas Nusa Mandiri, diakses Mei 8, 2025, https://ejournal.nusamandiri.ac.id/index.php/jitk/article/download/5916/1292
66. URAG: Implementing a Unified Hybrid RAG for Precise Answers in University Admission Chatbots – A Case Study at HCMUT - arXiv, diakses Mei 8, 2025, https://arxiv.org/html/2501.16276v1
67. An Event-Centric Knowledge Graph Approach for Public Administration as an Enabler for Data Analytics - MDPI, diakses Mei 8, 2025, https://www.mdpi.com/2073-431X/13/1/17
68. Knowledge Graphs: Question Answering - Docusign, diakses Mei 8, 2025, https://www.docusign.com/blog/knowledge-graphs-question-answering
69. Building, Improving, and Deploying Knowledge Graph RAG Systems on Databricks, diakses Mei 8, 2025, https://www.databricks.com/blog/building-improving-and-deploying-knowledge-graph-rag-systems-databricks
70. LexID: The Metadata and Semantic Knowledge Graph Construction ..., diakses Mei 8, 2025, https://lexin.cs.ui.ac.id/lexid-the-metadata-and-semantic-knowledge-graph-construction-of-indonesian-legal-document/index.html
71. (PDF) AC-IQuAD: Automatically Constructed Indonesian Question Answering Dataset by Leveraging Wikidata - ResearchGate, diakses Mei 8, 2025, https://www.researchgate.net/publication/377113292_AC-IQuAD_Automatically_Constructed_Indonesian_Question_Answering_Dataset_by_Leveraging_Wikidata
72. Fine Tuning vs. Retrieval Augmented Generation for Less Popular Knowledge - arXiv, diakses Mei 8, 2025, https://arxiv.org/html/2403.01432v3
73. [Literature Review] Building Efficient and Effective OpenQA Systems for Low-Resource Languages - Moonlight, diakses Mei 8, 2025, https://www.themoonlight.io/review/building-efficient-and-effective-openqa-systems-for-low-resource-languages
74. Building Efficient and Effective OpenQA Systems for Low-Resource Languages - arXiv, diakses Mei 8, 2025, https://arxiv.org/abs/2401.03590
75. RAG based Question-Answering for Contextual Response Prediction System - arXiv, diakses Mei 8, 2025, https://arxiv.org/html/2409.03708v1
76. Enhancing QA System Evaluation: An In-Depth Analysis of Metrics and Model-Specific Behaviors, diakses Mei 8, 2025, https://accesson.kr/jistap/v.13/1/85/55253
77. Comparative Analysis of State-of-the-Art Q\&A Models: BERT, RoBERTa, DistilBERT, and ALBERT on SQuAD v2 Dataset - ResearchGate, diakses Mei 8, 2025, https://www.researchgate.net/publication/378327924_Comparative_Analysis_of_State-of-the-Art_QA_Models_BERT_RoBERTa_DistilBERT_and_ALBERT_on_SQuAD_v2_Dataset
78. Evaluating Correctness and Faithfulness of Instruction-Following Models for

Question Answering | Transactions of the Association for Computational Linguistics - MIT Press Direct, diakses Mei 8, 2025, https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00667/121196/Evaluating-Correctness-and-Faithfulness-of

79. Evaluation and monitoring metrics for generative AI - Azure AI Foundry | Microsoft Learn, diakses Mei 8, 2025, https://learn.microsoft.com/en-us/azure/ai-foundry/concepts/evaluation-metrics-built-in

80. 5 Key Applications of AI Virtual Assistants in Customer Service - Number Analytics, diakses Mei 8, 2025, https://www.numberanalytics.com/blog/ai-virtual-assistants-customer-service-applications

81. Telco-DPR: A Hybrid Dataset for Evaluating Retrieval Models of 3GPP Technical Specifications - arXiv, diakses Mei 8, 2025, https://arxiv.org/html/2410.19790v1

82. Building the Knowledge Graph for Zakat (KGZ) in Indonesian Language - Akademi Sains Malaysia, diakses Mei 8, 2025, https://www.akademisains.gov.my/asmsj/?mdocs-file=6430

83. Revealing Trends in Datasets from the 2022 ACL and EMNLP Conferences - arXiv, diakses Mei 8, 2025, https://arxiv.org/html/2404.08666v1

84. What Are The Ethical Considerations In The Deployment Of Artificial Intelligence In Public Services? - Consensus Academic Search Engine, diakses Mei 8, 2025, https://consensus.app/questions/what-ethical-considerations-deployment-artificial/

85. AI Governance and Ethics: Lessons from the U.S. Visa Revocation Policy, diakses Mei 8, 2025, https://moderndiplomacy.eu/2025/03/11/ai-governance-and-ethics-lessons-from-the-u-s-visa-revocation-policy/

86. Latest 12+ Interesting Natural Language Processing Thesis Topics - PHD Services -, diakses Mei 8, 2025, https://phdservices.org/natural-language-processing-thesis-topics/