

Revolutionizing Consular Services for Indonesian Citizens: An AI-Powered Approach with SARI and Retrieval-Augmented Generation

I. Executive Summary

This report outlines a strategic approach to enhancing consular services for Indonesian citizens (WNI) through the integration of Artificial Intelligence (AI), with a particular focus on the Sahabat Artifisila Migran Indonesia (SARI) initiative. The global public sector is increasingly leveraging AI to meet evolving citizen expectations for accessible, efficient, and inclusive services. The Indonesian Ministry of Foreign Affairs (MoFA) is aligning with this trend, aiming to augment its digital service capabilities.

The SARI initiative, a collaboration between MoFA and UN Women, alongside technology partners, represents a targeted application of AI to provide empathetic, non-discriminatory information and support, especially for vulnerable Indonesian female migrant workers. SARI, envisioned as an AI-powered chatbot integrated with existing platforms like the Safe Travel mobile application, aims to offer immediate, interactive assistance for sensitive issues, complementing traditional consular services and web-based portals like Portal Peduli WNI.

The successful implementation of such an AI system hinges on a robust Retrieval-Augmented Generation (RAG) architecture. This architecture involves ingesting a diverse range of consular knowledge—including MoFA regulations, WNI protection guidelines, FAQs, and specialized content for empathetic interaction—into a structured, machine-readable format. Key components include advanced document chunking strategies, suitable multilingual embedding models, efficient vector databases, and a capable Large Language Model (LLM) like SahabatAI, which is proficient in Bahasa Indonesia and regional dialects. Fine-tuning this LLM for domain-specificity and empathetic response generation will be crucial.

Global case studies from the US, Singapore, Estonia, and others demonstrate AI's potential in automating routine tasks, enhancing information accessibility, and improving service efficiency. However, these examples also highlight significant challenges, including data privacy, algorithmic bias, the risk of AI "hallucinations," and the need for strong ethical governance and public trust.

Strategic recommendations include a phased implementation of SARI, continuous investment in a high-quality multilingual knowledge base, iterative fine-tuning and rigorous evaluation of the AI system using both automated metrics and human

oversight, establishment of robust ethical and data privacy protocols, and capacity building within MoFA. Ultimately, the success of this AI-driven transformation will depend on a human-centric approach that prioritizes user needs, builds trust, and ensures responsible deployment, particularly when serving vulnerable populations.

II. Enhancing Consular Services for Indonesian Citizens Through Artificial Intelligence

A. The Evolving Landscape of Public Service Delivery

The global landscape of public service delivery is undergoing a significant transformation, driven by rapid advancements in digital technologies and evolving citizen expectations. Governments worldwide are increasingly exploring and adopting Artificial Intelligence (AI) to enhance the accessibility, efficiency, and inclusivity of public services, often within the constraints of limited budgets.¹ This shift is not merely a technological trend but a strategic response to the growing demand from citizens for services that are available 24/7, personalized to their needs, capable of providing instant responses, and often offered in multiple languages.³

The Indonesian Ministry of Foreign Affairs (MoFA) has demonstrated a clear commitment to this digital transformation agenda, particularly in its efforts to improve services and protection for Indonesian citizens (WNI) abroad.⁵ Existing digital platforms such as Portal Peduli WNI and the Safe Travel application are testaments to this commitment. The introduction of initiatives like SARI (Sahabat Artifisial Migrant Indonesia) further underscores MoFA's strategic direction towards leveraging AI to meet the complex needs of its diaspora.

The global adoption of AI in public services, especially through tools like chatbots, offers valuable precedents. For instance, Singapore's GovTech agency has successfully deployed AI-powered chatbots across numerous government websites, handling routine inquiries and providing multilingual support, thereby extending service availability beyond traditional office hours and alleviating the burden on human agents.² Similarly, Estonia has integrated AI into various sectors, including public administration, using chatbots for citizen engagement.² These international examples highlight a common objective: to harness AI's capacity for automation and information processing to create more responsive and citizen-centric public services.

However, this technological advancement is not without its challenges. A critical consideration, especially when serving vulnerable populations such as migrant workers, is the digital divide. While AI can theoretically enhance accessibility, its practical effectiveness is contingent upon users' digital literacy levels and their access

to the necessary technology. MoFA's own strategic assessments have identified low digital literacy among WNI abroad as a significant hurdle.⁹ Therefore, the deployment of AI solutions, even those designed with empathy like SARI, must be accompanied by comprehensive strategies to improve digital inclusion, ensuring that the benefits of these technological advancements are equitably distributed.

B. Potential of AI in Addressing Consular Challenges

Consular services inherently involve a wide array of complex tasks and challenges. These include managing a high volume of diverse inquiries, ensuring the provision of accurate and up-to-date information on matters such as passport and visa applications or document legalization, offering timely support to WNI in distress, and upholding the broad mandate of WNI protection abroad.⁹ The dynamic nature of international regulations, travel advisories, and security situations further complicates these responsibilities.

Artificial Intelligence, particularly through the deployment of advanced Large Language Models (LLMs) and systems built upon Retrieval-Augmented Generation (RAG) architecture, presents significant potential to address these multifaceted challenges. AI-powered chatbots can provide citizens with 24/7 access to information and frequently asked questions (FAQs), significantly improving service availability.³ By automating responses to routine and high-volume inquiries, AI can free up human consular staff to concentrate on more complex, sensitive, or emergency cases that require direct human intervention and nuanced judgment.²

Furthermore, AI offers robust multilingual support capabilities, a crucial feature for a global diaspora. It can be employed to summarize critical security information and potential threats relevant to WNI in specific regions, enhancing proactive protection efforts.⁵ AI tools can also assist in guiding users through complex application processes for visas or passports and aid in document verification, streamlining administrative workflows.

Despite these promising applications, the implementation of AI in governmental contexts is fraught with challenges that demand careful consideration. Data privacy and security are paramount, especially when handling sensitive citizen data. Algorithmic bias, if present in the AI models or the data they are trained on, can lead to unfair or discriminatory outcomes. A significant concern with generative AI models is the risk of "hallucinations"—the generation of plausible but incorrect or fabricated information—which can have serious consequences in a consular context where accuracy is critical. Additionally, the integration of AI with legacy systems, the need for a skilled workforce to develop and manage these AI tools, and the imperative for

robust human oversight and ethical frameworks are substantial hurdles.¹

The successful integration of AI into consular services, therefore, necessitates a carefully balanced approach. It requires leveraging AI's potential for efficiency and enhanced service delivery while concurrently establishing strong data governance, ethical guidelines, and continuous monitoring mechanisms. This is particularly true when AI systems are tasked with providing information or support in critical situations or to vulnerable individuals, where the margin for error is minimal and the demand for accuracy and fairness is at its peak. The development of RAG architectures, which aim to ground AI-generated responses in verified, up-to-date information sources, is a key technological strategy to mitigate risks like hallucinations and improve the reliability of AI in these sensitive domains.²⁰

III. SARI (Sahabat Artifisila Migran Indonesia): A Dedicated AI for WNI Protection

A. Objectives, Target Users, and Collaborative Development

The SARI (Sahabat Artifisila Migran Indonesia) initiative represents a focused application of Artificial Intelligence by the Indonesian Ministry of Foreign Affairs (MoFA) aimed at enhancing the protection and support services for Indonesian citizens abroad, with a particular emphasis on Indonesian migrant workers (PMI).⁵ Launched in collaboration with UN Women and supported by the Migration Multi-Partner Trust Fund (MMPTF), SARI's primary objective is to provide accessible, unbiased, and non-discriminatory information and initial support services.²⁴ It is specifically designed to cater to the needs of vulnerable groups, most notably female migrant workers who may be susceptible to violence and exploitation.²⁴ SARI is envisioned not as a replacement for existing in-person consular services but as a complementary tool that leverages technology to extend reach and improve responsiveness.²⁴

The target users for SARI are primarily Indonesian migrant workers.²⁴ The collaboration with UN Women brings a crucial human-rights and gender-sensitive perspective to the initiative's design and implementation.²⁴ This partnership underscores a commitment to developing AI solutions that are not only technologically advanced but also ethically grounded and tailored to the specific vulnerabilities and needs of its intended users. The broader SahabatAI LLM ecosystem, which could potentially power SARI, involves a consortium of Indonesian technology and telecommunication companies like Indosat Ooredoo Hutchison and GoTo, with support from international research centers and tech experts such as AI Singapore and Tech Mahindra.²⁸ This

multi-stakeholder approach signifies a comprehensive effort to build sovereign AI capabilities tailored to Indonesian linguistic and cultural contexts. The SARI initiative, therefore, is more than just a technological deployment; it is a strategic endeavor to utilize AI for proactive protection and empowerment, particularly for a demographic that often faces significant challenges abroad.

B. Key Features: AI Chatbot, Language Capabilities, and Empathetic Interaction

SARI is manifested as an AI-powered chatbot, designed to serve as an interactive and accessible first point of contact for Indonesian migrant workers seeking information or assistance.⁵ A key technological aspect of SARI is its language capabilities. It is expected to detect the language used by the user⁵ and may support various regional Indonesian languages in addition to Bahasa Indonesia.⁶ This aligns with the capabilities of the underlying SahabatAI LLM, which has been pre-trained and instruction-tuned for Bahasa Indonesia, Javanese, Sundanese, and English, demonstrating an understanding of local context and informal language.²⁸

A distinguishing feature of SARI is its emphasis on empathetic interaction. The system is being developed to deliver responses that are not only factually accurate but also empathetic, free from stigma or prejudice, particularly when dealing with sensitive issues such as gender-based violence.⁵ This is described as a form of "digital empathy" and "digital solidarity," aiming to create a supportive and non-judgmental environment for users.²⁴ To achieve this, SARI's development incorporates gender bias-free data and a human-centered, participatory design process. This process involves direct consultations with female migrant workers, service providers, civil society organizations (CSOs), and youth representatives to ensure the chatbot's responses and functionalities are aligned with actual user needs and sensitivities.²⁷

In terms of information provision, SARI is designed to offer real-time information delivery, improve access to vital services, and summarize security information, including potential dangers and crime threats relevant to WNI abroad.⁵ It also serves as a channel for users to seek help without the fear of stigma that might be associated with direct human interaction in some sensitive cases.²⁵ The goal of creating an AI capable of genuine empathy presents a significant technological and ethical challenge. It necessitates sophisticated natural language understanding (NLU) and natural language generation (NLG) capabilities, coupled with an LLM that has been carefully fine-tuned on datasets reflecting empathetic communication styles and culturally appropriate responses. This moves beyond simple information retrieval to a more nuanced form of AI-human interaction, critical for supporting individuals in potentially distressing situations.

C. Integration with Existing Digital Platforms (Portal Peduli WNI, Safe Travel) and Complementary Roles

The SARI initiative is strategically designed to integrate with and complement MoFA's existing digital service platforms for Indonesian citizens abroad: the **Portal Peduli WNI** and the **Safe Travel** mobile application.⁵ The Portal Peduli WNI is a web-based platform primarily for WNI residing overseas, offering features such as online self-reporting (Lapor Diri), access to consular services (Pelayanan Kekonsuleran), and a mechanism for reporting cases or grievances (Pengaduan Kasus).⁵ Safe Travel, on the other hand, is a mobile application geared towards Indonesian citizens undertaking short-term travel abroad, though also usable by expatriates. It provides practical travel information, security alerts, emergency contact details for Indonesian missions, and features like location sharing in crises.⁵

SARI is explicitly mentioned as being available on the Safe Travel app.²⁵ This integration is a key aspect of its deployment strategy, leveraging an established mobile platform to reach its target audience. While Portal Peduli WNI and Safe Travel provide crucial information dissemination and reporting functionalities, they are largely non-interactive or rely on users navigating static information and completing forms.⁵ SARI, as an AI-powered chatbot, aims to overcome some of these limitations by offering a more dynamic, conversational, and immediate mode of interaction.

The complementary role of SARI is particularly evident in its focus on providing empathetic support for sensitive issues, such as those faced by female migrant workers who may be victims of gender-based violence or exploitation.²⁴ In such situations, the anonymity and non-judgmental nature of a well-designed chatbot can be less intimidating than direct human interaction, encouraging individuals to seek initial help and information. By being accessible through the Safe Travel mobile app, SARI can offer a more readily available and interactive support channel for migrant workers who predominantly use mobile devices. This approach can enhance the overall WNI protection ecosystem by adding a layer of immediate, empathetic, and specialized AI-driven assistance to the existing information and reporting tools.

IV. AI Adoption in Global Public and Consular Services: A Comparative Analysis

A. Case Studies: AI Initiatives by Governments and International Organizations

The adoption of Artificial Intelligence in public and consular services is a growing global trend, with numerous governments and international organizations exploring its

potential to enhance efficiency, accessibility, and citizen engagement.

The **United States Department of Homeland Security (DHS)** utilizes AI across various immigration and border control functions. A notable initiative is the "Machine Learning Translation Technology Initiative" (DHS-197), which is developing a real-time translation tool to help staff communicate with individuals with limited English proficiency. This tool, intended for non-critical conversations, will offer voice-to-text, text-to-voice, and voice-to-voice translations in at least 21 languages and is expected to be operational by January 2026. DHS also employs Generative AI (GAI) for training asylum and refugee officers in interview skills and for summarizing documents in investigative processes.¹⁹ Furthermore, Customs and Border Protection (CBP) uses AI for biometric identification and traveler screening.

The **United States Department of State (DoS)** has several planned AI use cases. These include the "Consular Affairs Photo Quality Service (PQS) / FaceVACS" for automatically checking passport photo quality during online renewals, "Evaluating Customer Feedback and Sentiments with AI" using NLP and LLMs, "Translation of Consular Content using AI" for its websites, and an enhanced "Travel.State.Gov (TSG) Enhanced Search and Chatbot" to handle FAQs and provide predefined responses.⁴⁵

Singapore's GovTech agency has been a pioneer in deploying AI-powered chatbots. "Ask Jamie" is a virtual assistant available on over 70 government websites, alongside specialized chatbots like "HealthBuddy" for healthcare inquiries and "CPF Chatbot" for financial questions. These NLP-driven systems provide multilingual, instant responses and have reportedly led to a 50% reduction in call center workload and 80% faster response times.²

Estonia has integrated AI into multiple public service domains. Its AI-powered health information system facilitates patient data management and predictive health analytics. In transportation, AI-based systems manage traffic flow and support on-demand public transport. Estonian public administration also utilizes AI chatbots for citizen engagement and AI algorithms for data-driven policy development.²

Other notable examples include **Canada's Revenue Agency**, which uses an AI-driven system for tax fraud detection², **Australia's Visa Finder**, an AI tool guiding users to appropriate visa options, and the **European Union's iBorderCtrl**, an AI-enhanced border security system employing facial recognition and AI lie-detection.⁸

Internationally, **UN Women** is collaborating with MoFA Indonesia on the SARI AI chatbot, highlighting the role of international organizations in fostering AI for social

good, particularly for vulnerable groups like female migrant workers.²⁴

These case studies reveal a predominant trend: governments are leveraging AI primarily to automate high-volume, routine, or information-intensive tasks. This strategy aims to improve operational efficiency and allow human personnel to focus on more complex or sensitive issues that require direct human judgment and empathy. The success of Singapore's chatbots in reducing call center load ² and the US DHS's use of GAI for document summarization ¹⁹ exemplify this approach. This pragmatic application of AI targets areas where automation offers clear efficiency benefits without entailing the high risks associated with complex human decision-making.

Table 1: Comparative Analysis of AI Chatbots/Systems in Public/Consular Services

Country /Organization	System Name/Initiative	Primary Function	Target Users	AI Technology Used	Key Features	Reported/Anticipated Impact/Benefits	Key Challenges/Considerations
USA (DHS)	Machine Learning Translation Initiative (DHS-197)	Real-time translation	DHS Staff, Non-English proficient individuals	Machine Learning , NLP	Voice-to-text, text-to-voice, voice-to-voice, 21+ languages	Faster communication, improved service access for LEP individuals	Accuracy for critical conversations, integration
USA (DoS)	TSG Enhanced Search and Chatbot (Planned)	FAQ handling , information provision	Travelers , general public	Chatbot, LLM (likely)	Predefined responses, enhanced search	Improved information access on Travel.State.Gov	Ensuring accuracy, managing diverse queries ⁴⁵
Singapore	Ask Jamie,	Routine inquiry	Citizens	NLP, AI Chatbot	Multilingual,	50% call center	Maintaining

(GovTech)	HealthBuddy, CPF Chatbot	handling , information provision		s	instant responses, 24/7 availability	workload reduction, 80% faster responses	accuracy across diverse services ²
Estonia	Citizen Engagement Chatbots	Information provision, common inquiries	Citizens	AI Chatbots, NLP	Instant responses	Reduced workload on public servants , enhanced service accessibility	Data privacy, need for skilled AI professionals ²
Australia	Visa Finder	Visa guidance, application assistance	Prospective visa applicants	AI-powered questioning	Targeted questions, suitable visa options	Reduced application errors, improved user guidance	Keeping information up-to-date with policy changes
Indonesia/UN Women	SARI	Information, support for migrant workers	Indonesian female migrant workers	AI Chatbot, NLP, (SahabatAI LLM likely)	Language detection, empathetic responses, information on rights & safety, help-seeking	Accessible, non-discriminatory support, violence prevention	Digital literacy of users, data sensitivity, ensuring true empathy ²⁴

This comparative table allows for a structured understanding of how different entities

are approaching AI in public and consular services. It highlights commonalities in leveraging AI for efficiency and accessibility, while also pointing to the unique, targeted approach of initiatives like SARI that focus on vulnerable populations and empathetic interaction. Learning from these international experiences can inform MoFA's strategy for developing and deploying AI solutions effectively.

B. Common Applications in Consular Domains

The application of AI in consular services globally is expanding, addressing various aspects of citizen support and administrative processes. Common applications include:

- **Information Provision and FAQ Handling:** AI-powered chatbots are widely used to provide citizens with 24/7 access to answers for frequently asked questions regarding consular services, visa requirements, passport procedures, and local regulations. This automates responses to high-volume, repetitive inquiries.³
- **Application Assistance:** AI tools are being developed to guide users through complex visa and passport application processes. This includes assistance with form filling, checking for document completeness, and even verifying the quality of submitted photographs, aiming to reduce errors and processing times.
- **Emergency Support and Alerts:** Mobile applications, sometimes integrated with AI features, can provide real-time security alerts, location-based safety information, and facilitate emergency contact with embassies or consulates. The Safe Travel app, with which SARI is integrated, serves such functions.³⁸
- **Language Translation Services:** AI-driven real-time translation tools are being implemented to bridge communication gaps between consular staff and citizens who have limited proficiency in the official language of the host country or the consular office.
- **Citizen Feedback and Engagement:** AI can be used to collect, categorize, and analyze citizen feedback on consular services, helping identify areas for improvement and gauge public sentiment.³
- **Document Summarization and Analysis:** For internal consular operations, AI can assist staff in processing and understanding large volumes of documents, such as legal texts, policy documents, or investigative reports related to WNI cases.¹⁹

While AI demonstrates considerable utility in these informational and transactional aspects of consular work, its application in areas requiring deep empathy, nuanced understanding of complex human situations, and crisis intervention—such as the support SARI aims to provide for victims of violence and exploitation—represents a more advanced and challenging domain. Most current AI applications in consular

services focus on structured information and process automation. SARI's objective to offer empathetic support ²⁴ demands a higher level of AI sophistication, including advanced natural language understanding, sentiment analysis, and robust ethical safeguards, distinguishing it from standard informational chatbots.

C. Lessons Learned: Benefits Achieved and Challenges Encountered Globally

The global adoption of AI in public and consular services has yielded significant benefits while also surfacing common challenges. Understanding these lessons is crucial for MoFA as it expands its AI initiatives.

Benefits Achieved:

The primary benefits reported include increased operational efficiency and potential cost savings through automation of routine tasks.¹ AI systems enable 24/7 service availability and faster response times, improving citizen access to information and services.³ Multilingual support and tailored assistance enhance accessibility for diverse populations, including those with disabilities. Furthermore, AI facilitates data-driven decision-making and policy development, leading to more effective governance and improved citizen engagement and satisfaction.¹ RAG systems, in particular, promise more accurate and up-to-date information by grounding responses in verified knowledge bases.²⁰

Challenges Encountered:

A major category of challenges revolves around data. Ensuring data privacy and security, especially when dealing with sensitive citizen information, is a paramount concern. Data sovereignty issues can arise if data is processed or stored in foreign jurisdictions.⁵⁰ Moreover, the quality and representativeness of training data are critical; biased or incomplete data can lead to AI systems that perpetuate or even amplify societal biases.

Regarding **AI models themselves**, the risk of "hallucinations"—where AI generates incorrect or fabricated information—is a significant concern, particularly in governmental contexts where accuracy is vital.⁵¹ Algorithmic bias can lead to unfair or discriminatory outcomes. The "black box" nature of many complex AI models, especially LLMs, poses challenges for transparency, interpretability, and accountability, which can erode public trust if AI-driven decisions or information cannot be adequately explained.⁵³

Implementation and operational challenges include the difficulty of integrating new AI systems with legacy IT infrastructure.² There is often a shortage of skilled AI professionals and a need for comprehensive workforce training to manage and utilize these new technologies effectively.² Initial implementation costs can be substantial.² Establishing clear ethical guidelines, ensuring public trust and acceptance, and developing robust methods for managing and measuring AI performance are ongoing tasks.¹

A critical lesson emerging from global AI adoption in government is that technological advancements must be accompanied by strong governance frameworks. These frameworks need to address ethical use, data privacy, transparency in algorithmic decision-making, and mechanisms for public accountability. The opacity of some AI models, particularly sophisticated LLMs, can undermine public trust if the information or decisions generated by AI systems cannot be clearly understood, verified, or appealed. Initiatives like the US DHS AI Playbook ⁴⁴ and Estonia's focus on data protection regulations ² illustrate the necessity of a regulated and responsible approach to AI in the public sector.

V. Data Foundation for the Consular AI System

A. Identifying and Sourcing Relevant Knowledge for Consular Q&A

The effectiveness of an AI-powered consular Q&A system, such as SARI, is fundamentally dependent on the quality, comprehensiveness, and accuracy of its underlying knowledge base. This knowledge base must be meticulously curated from a variety of authoritative sources within and related to the Indonesian Ministry of Foreign Affairs (MoFA).

Key sources include:

1. **Official MoFA Regulations (Permenlu):** These ministerial regulations form the legal and procedural backbone for consular services. Examples include Permenlu No. 5 Tahun 2018 concerning the Protection of Indonesian Citizens Abroad ⁹, and regulations governing document legalization, such as Permenlu No. 14 Tahun 2022 and No. 13 Tahun 2019.⁵⁵ Regulations pertaining to passport and visa services are also critical.⁵⁵
2. **MoFA Strategic Plans (Renstra) and Performance Reports (LAKIP):** Documents like the *Rencana Strategis Kementerian Luar Negeri 2020-2024* ⁹ and *Laporan Akuntabilitas Kinerja Instansi Pemerintah* ⁷³ provide essential context on MoFA's strategic priorities, ongoing challenges in service delivery, digital transformation objectives, and statistical data regarding WNI cases and the demand for consular services. This information helps align the AI's purpose with broader ministerial goals.
3. **Frequently Asked Questions (FAQs) and Official Website Content:** Information from the primary Kemlu.go.id portal, the specialized consular services portal (layanandiplomatik.kemlu.go.id), and the websites of Indonesian embassies and consulates worldwide are rich sources of answers to common citizen inquiries.⁴³
4. **WNI Protection Information:** Specific guidelines, information on common issues

faced by WNI (e.g., human trafficking, labor disputes, legal aid requirements), and data managed through the Portal Peduli WNI are vital, especially for tailoring the SARI chatbot's knowledge base.⁹

5. **Safe Travel Application Content:** Data from the Safe Travel app, including information on security conditions in various countries, travel advisories, local immigration rules, health facilities, and emergency contact details, can provide dynamic and location-specific context.¹¹
6. **UN Women and CSO Resources:** For the SARI initiative, materials from UN Women and relevant Civil Society Organizations concerning the protection of female migrant workers, guidelines on handling gender-based violence, and examples of empathetic communication will be crucial for developing its specialized supportive capabilities.²⁴

The creation of a comprehensive and reliable knowledge base for a consular RAG system necessitates a robust data strategy. This strategy must address the ingestion of diverse data formats (e.g., PDFs for regulations, HTML for web content, potentially structured data from case management systems). Furthermore, given the dynamic nature of consular information (e.g., travel advisories, visa policies), mechanisms for continuous updates and verification are paramount to ensure the AI system provides consistently accurate and timely information.²⁰

B. Structuring the Dataset: Q&A Pairs, Document Snippets, and Metadata

To effectively power a RAG-based consular Q&A system, the sourced knowledge must be meticulously structured. This involves transforming raw information into formats that are optimized for retrieval and subsequent generation by an LLM.

1. **Question-Answer (Q&A) Pairs:** A significant portion of the dataset should consist of well-defined Q&A pairs. These can be extracted directly from existing FAQs on MoFA and embassy websites⁵⁷ and from analyses of common inquiries received by consular staff.¹⁰ For the SARI chatbot, specialized Q&A pairs must be developed, focusing on issues pertinent to migrant workers, including guidance on rights, emergency assistance protocols, and how to report violence or exploitation. These pairs should also incorporate examples of empathetic and supportive responses, potentially derived from consultations with target users and support organizations.²⁴
2. **Document Snippets (Chunks):** Lengthy official documents, such as Permenlu, Renstra, and international agreements, need to be segmented into smaller, semantically coherent and contextually complete snippets or "chunks." This process is critical for effective retrieval, as LLMs have context window limitations, and providing overly long, undifferentiated text can dilute relevance.⁸² The

chunking strategy must be carefully chosen to preserve the meaning and integrity of legal articles, procedural steps, and policy statements.

3. **Metadata Enrichment:** Each Q&A pair and document chunk must be tagged with rich metadata to facilitate accurate retrieval and contextual understanding by the AI system. Essential metadata fields include:

- `chunk_id` or `qa_pair_id`: A unique identifier.
- `original_document_id`: A reference to the source document (e.g., "Permenlu No. 5 Tahun 2018," "FAQ_Kemlu_Visa").
- `document_type`: Classification of the source (e.g., Regulation, FAQ, Guideline, News Report, Strategic Plan).
- `text_snippet_indonesian`: The actual textual content of the chunk or answer in Bahasa Indonesia.
- `text_snippet_english`: The English translation, if available or deemed necessary for multilingual capabilities.
- `keywords`: A list of relevant keywords (e.g., "paspor hilang," "visa kerja," "legalisasi ijazah," "kekerasan PMI," "lapor diri").
- `category`: A primary categorization (e.g., Passport Services, Visa Services, Document Legalization, WNI Protection, Emergency Assistance, SARI-Specific Support).
- `sub_category` (optional): More granular categorization (e.g., "WNI Protection - Human Trafficking," "Visa Services - Schengen").
- `target_audience` (optional): Specifying if the information is for WNI umum (general public), PMI wanita (female migrant workers), MoFA staff, etc.
- `related_questions`: Examples of user questions that this chunk or Q&A pair could effectively answer.
- `source_url`: If applicable, the URL of the source document or webpage.
- `publication_date`: Date of the original document's publication.
- `last_verified_date`: Date when the information in the chunk was last verified for accuracy and currency.
- `empathy_level_required` (for SARI): A tag indicating if the response should incorporate a specific level or style of empathy.

Table 2: Sample Dataset Snippets for Consular QA

chunk_id	original_document_id	document_type	text_snippet_indonesian	keywords	category	related_questions	last_verified_date

PERMEN LU_5_20 18_ART8 _A	Permenl u No. 5 Tahun 2018	Regulati on	"Pelindu ngan Kekonsul eran...pa ling sedikit meliputi: a. melindu ngi kepentin gan Negara dan WNI di Negara Penerim a berdasa rkan ketentua n peratura n perunda ng-unda ngan..."	pelindun gan WNI, kekonsul eran, kepentin gan negara	WNI Protecti on	"Apa saja bentuk pelindun gan kekonsul eran untuk WNI?"	2024-10 -01
FAQ_KE MLU_PA SPOR_O 03	kemlu.g o.id/faq/ paspor	FAQ	"T: Berapa lama proses pembua tan Paspor Diploma tik dan Paspor Dinas? J: Paling lambat 4 (empat) hari kerja setelah semua	paspor diplomat ik, paspor dinas, waktu proses	Passport Services	"Berapa lama bikin paspor dinas?", "Proses paspor diplomat ik berapa hari?"	2025-01 -15

			persyaratan dilengkapi..."				
SARI_KB_EMPAT HY_001	Modul Pelatihan SARI - Empati	Guideline	"Saya memahami ini pasti situasi yang sangat sulit bagi Anda. Mari kita coba cari solusi bersama. Bisakah Anda ceritakan lebih lanjut apa yang terjadi?"	empati, dukungan, kekerasan, pekerja migran, lapor	WNI Protection - Migrant Worker Support	"Saya takut dan butuh bantuan.", "Saya mengalami masalah."	2025-03-01
RENSTR A_KEMLU_2024_ DIGI	Renstra Kemlu 2020-2024, Bab IV	Strategic Plan	"Transformasi digital untuk mewujudkan pelayanan tepat, mudah, murah dan akurat melalui Portal Peduli WNI dan Safe	transformasi digital, Portal Peduli WNI, Safe Travel	MoFA Strategy - Digitalization	"Apa strategi digital Kemlu untuk layanan WNI?"	2024-11-01

			Travel."				
LEGAL_ DOC_IJA ZAH_00 1	Laman Legalisa si Dokume n Kemlu.g o.id	Web Content	"Untuk legalisas i ijazah yang diterbitk an di Indonesi a dan akan digunak an di luar negeri, dokume n harus dilegalis asi terlebih dahulu oleh Kemendi kbud..."	legalisas i, ijazah, dokume n, persyara tan, Kemendi kbud	Docume nt Legaliza tion	"Bagaim ana cara legalisas i ijazah untuk dipakai di luar negeri?"	2025-02 -20

This structured dataset, particularly the sample table, concretely illustrates how diverse consular information can be transformed into a machine-usable format. For MoFA, this clarifies the data preparation lifecycle, emphasizing the crucial role of comprehensive metadata for retrieval accuracy and the inclusion of specialized content, like empathetic dialogue examples, to meet SARI's unique objectives.

The dataset for SARI, in particular, requires careful curation. Beyond factual information, it must incorporate extensive examples of empathetic phrasing, active listening prompts, culturally sensitive language, and dialogue flows designed to guide and support users in distress. This may necessitate the generation of new content or the careful adaptation of existing materials, drawing from the insights gained during SARI's human-centered design process involving consultations with migrant workers and service providers.²⁴ This specialized content is vital for training or fine-tuning the LLM to deliver genuinely supportive and non-judgmental interactions.

C. Considerations for Data Collection, Preprocessing, and Multilingual Support

The development of a robust data foundation for the consular AI system involves

several critical considerations:

1. **Data Quality and Currency:** The foremost priority is ensuring the accuracy, consistency, and up-to-dateness of all information sourced. Consular information, particularly regulations, travel advisories, and service procedures, is subject to change. Therefore, establishing rigorous processes for regular review, verification, and updates of the knowledge base is essential to prevent the AI from providing outdated or incorrect guidance.
2. **Preprocessing Techniques:** Raw data from diverse sources (PDFs, HTML web pages, Word documents) must be converted into clean, usable text. This involves:
 - Effective PDF-to-text conversion, preserving layout where important (e.g., for legal articles).
 - Stripping HTML tags and irrelevant content from web-scraped FAQs and informational pages.
 - Text normalization, including handling variations in spelling, abbreviations, and formatting.
 - Special attention to extracting and representing structured information like tables or lists accurately.
3. **Multilingual Support:** The system must primarily support Bahasa Indonesia, the national language. Given the international context of consular work and the use of English in many official and diplomatic communications, robust English language support is also crucial. The SahabatAI LLM, a potential core for this system, is already trained on Bahasa Indonesia, Javanese, Sundanese, and English, which is a significant advantage.²⁹ For the SARI initiative, the ability to understand and respond to various Indonesian dialects or informal language commonly used by migrant workers could further enhance its effectiveness and accessibility.⁶ The choice of embedding models must also reflect this multilingual requirement, ensuring accurate semantic understanding and retrieval across supported languages.⁸⁷ If the knowledge base contains documents in both Indonesian and English, or if queries are posed in one language while relevant documents exist in another, the RAG system must handle such cross-lingual scenarios effectively. This might involve leveraging multilingual embedding models capable of mapping concepts across languages or incorporating a translation layer, though the latter can introduce its own complexities and potential for error.
4. **Ethical Data Handling and Bias Mitigation:**
 - **Anonymization:** If historical case data or WNI interaction logs are used to enrich the knowledge base or for fine-tuning (e.g., for empathetic response styles), all Personally Identifiable Information (PII) must be rigorously anonymized or pseudonymized to comply with data privacy regulations and

protect citizen confidentiality.

- **Bias Mitigation:** For SARI, in particular, the commitment to providing unbiased and non-discriminatory information necessitates careful curation of the training and knowledge data to be free of gender, ethnic, or other forms of bias.²⁷ This involves actively auditing data sources and potentially augmenting the dataset with balanced perspectives.

5. **Data Governance and Maintenance:** A clear governance framework for the knowledge base is essential. This includes defining responsibilities for data sourcing, verification, updating, and quality control. A continuous improvement loop, potentially incorporating feedback from users and consular staff, should be established to refine and expand the knowledge base over time.

The effectiveness of the multilingual RAG system will heavily depend on the quality of its embedding models in handling both Bahasa Indonesia and English, especially if cross-lingual information retrieval becomes a frequent necessity. While SahabatAI's multilingual foundation is promising, the specific performance of chosen embedding models on Indonesian consular domain-specific terminology and query types will need thorough evaluation using benchmarks like MTEB⁹⁰ or domain-specific test sets.

VI. System Architecture: RAG-Powered Q&A for Consular Assistance

A. Conceptual Workflow: From User Query to AI-Generated Answer

The proposed AI-powered consular Q&A system will leverage a Retrieval-Augmented Generation (RAG) architecture to provide accurate, contextually relevant, and (for SARI) empathetic responses. The workflow, from the moment a user inputs a query to the generation of an answer, involves several interconnected stages:

1. **User Interaction & Query Input:** A WNI or MoFA staff member interacts with the system via a user interface, such as the SARI chatbot integrated into the Safe Travel mobile application or an internal MoFA web portal. The query is submitted in Bahasa Indonesia or English.
2. **Query Preprocessing:** The user's raw query undergoes initial processing. This may include language detection (to route to the appropriate language model or processing pipeline if necessary), spell correction, and normalization of text (e.g., converting to lowercase, handling abbreviations). For more advanced systems, intent recognition might be employed to categorize the query's purpose.
3. **Query Embedding:** The preprocessed query is then transformed into a dense vector representation (embedding) using a specialized embedding model. This model is chosen for its proficiency in capturing the semantic meaning of text in

the supported languages (Bahasa Indonesia, English).

4. **Information Retrieval from Vector Database:** The query embedding is used to search a vector database. This database stores pre-computed embeddings of document chunks and Q&A pairs from the curated consular knowledge base. A similarity search algorithm (e.g., k-Nearest Neighbors (k-NN) or Approximate Nearest Neighbor (ANN) ⁸³) identifies and retrieves the top-K document chunks or Q&A pairs that are most semantically similar to the user's query.
5. **Optional Reranking:** The initially retrieved set of chunks can be further refined by a reranking model. This model applies more nuanced relevance scoring, considering factors beyond simple semantic similarity, to reorder the chunks and prioritize the most pertinent information for the LLM.⁸³
6. **Context Augmentation and Prompt Engineering:** The selected, most relevant retrieved document chunks are then combined with the original user query. This augmented input is formatted into a specific prompt designed to guide the LLM. The prompt includes instructions on how to use the provided context, the desired tone of the response (e.g., empathetic and supportive for SARI, formal and procedural for general consular information), and any constraints on the output format or length.⁵²
7. **LLM-based Answer Generation:** The augmented prompt is fed into a Large Language Model (LLM), such as a fine-tuned version of SahabatAI. The LLM processes this input and generates a coherent, contextually relevant, and factually grounded answer based on the information contained within the retrieved context.
8. **Response Post-processing and Delivery:** The LLM-generated answer may undergo a final post-processing stage. This could include safety checks (to filter out any potentially harmful or inappropriate content), generation of citations if the system is designed to reference sources, and formatting for optimal display. The final response is then delivered to the user through the original chatbot interface.

This multi-stage RAG pipeline ensures that the LLM's responses are not solely based on its pre-trained knowledge (which can be outdated or generic) but are actively informed by specific, relevant, and up-to-date information from the consular domain knowledge base. The effectiveness of this entire workflow is critically dependent on the quality and synergy of its individual components: the curated data, the chunking strategy, the accuracy of the embedding and retrieval processes, the capability of the LLM, and the design of the prompts. A deficiency in any single component can significantly degrade the overall performance and reliability of the system.⁵¹

Diagram 1: High-Level RAG System Architecture for Consular Q&A

Code snippet

graph TD

```
A -- User Query --> B(Query Preprocessing);
B -- Processed Query --> C(Embedding Model for Query);
C -- Query Embedding --> D{Vector Database};
D -- Similarity Search / Retrieval --> E;
subgraph Knowledge Base Ingestion
  F --> G(Document Loaders);
  G --> H(Document Chunking);
  H -- Chunks --> I(Embedding Model for Documents);
  I -- Document Embeddings --> D;
end
E -- Retrieved Context --> J(Prompt Augmentation / Engineering);
B -- Original Query --> J;
J -- Augmented Prompt --> K(Large Language Model - LLM e.g., Fine-tuned SahabatAI);
K -- Generated Answer --> L(Response Post-processing);
L -- Final Response --> A;
M -- Refined Chunks --> J;
E --> M;
```

Diagram 1 illustrates the flow of information in the RAG system. The user query is processed and used to retrieve relevant context from the vector database, which is populated through a separate data ingestion pipeline. This context, along with the query, is then used by the LLM to generate a grounded response.

B. Key Components:

1. Data Ingestion and Preprocessing:

- **Document Loaders:** Capable of handling diverse source formats such as PDFs (for Permenlu and official reports), HTML (for website FAQs and articles), and plain text.
- **Document Chunking Strategies:** The choice of chunking strategy is critical for official MoFA documents, which are often structured and carry precise

legal or procedural meaning. Simple fixed-size chunking risks breaking apart essential clauses or contextually linked information. Therefore, **paragraph-based chunking** or **semantic chunking**⁸⁴ would be more appropriate. Semantic chunking, which groups sentences based on thematic coherence, could be particularly beneficial for ensuring that retrieved snippets are contextually complete. For highly structured documents like regulations with articles and sub-articles, **layout-aware or hierarchical chunking** methods should be investigated to preserve the inherent document structure. This ensures that the LLM receives meaningful and complete units of information, reducing the risk of misinterpretation or generating incomplete answers.⁸⁵

- **Metadata Extraction:** During ingestion, extracting and associating rich metadata (as detailed in Section V.B) with each chunk is crucial for effective retrieval and filtering.

2. **Embedding Models:**

- The selection of an embedding model is vital for the accuracy of the retrieval stage. The model must effectively capture the semantic meaning of consular queries and document content in both Bahasa Indonesia and English.
- Promising candidates include robust multilingual models such as sentence-transformers/paraphrase-multilingual-mpnet-base-v2, which has shown support for Indonesian.⁸⁹
- Alternatively, exploring or developing Indonesian-specific embedding models, potentially by fine-tuning existing multilingual models on a domain-specific consular corpus, could enhance performance on nuanced local terminology. The LUSIFER architecture, which combines multilingual encoders with LLM-based embedding models, also presents a viable approach for adapting high-performing English-centric models for multilingual use.⁸⁷
- Performance should be benchmarked using resources like MTEB (Massive Text Embedding Benchmark)⁹⁰ and custom test sets tailored to consular domain language.

3. **Vector Database and Retrieval Mechanism:**

- A specialized vector database (e.g., Chroma, FAISS, Pinecone, Milvus⁹⁴) is required for efficient storage and indexing of the high-dimensional document embeddings.⁸²
- The retrieval mechanism will employ similarity search algorithms (e.g., k-NN, or more scalable Approximate Nearest Neighbor (ANN) methods like HNSW or IVF) to quickly find the top-k document chunks most relevant to the query embedding.⁸³
- Consideration should be given to implementing a **reranking model** as a

secondary stage after initial retrieval. Rerankers (often cross-encoder models) can examine the top-k retrieved chunks more closely and reorder them based on more sophisticated relevance signals than cosine similarity alone, thereby improving the quality of context fed to the LLM.⁸³

4. **Large Language Model (LLM) for Generation:**

- **SahabatAI:** This Indonesian-focused LLM, based on Llama3 and Gemma2 architectures and pre-trained on Bahasa Indonesia, Javanese, Sundanese, and English, is a strong candidate for the generation component.²⁸ Its familiarity with local languages and cultural context is advantageous.
- **Fine-tuning SahabatAI:** To optimize SahabatAI for the specific demands of consular Q&A and, particularly for SARI, empathetic response generation, fine-tuning will be necessary. Parameter-Efficient Fine-Tuning (PEFT) techniques such as **LoRA (Low-Rank Adaptation)** or **QLoRA (Quantized LoRA)**⁹⁸ are highly recommended. These methods allow for efficient adaptation of large models with significantly fewer trainable parameters and computational resources compared to full fine-tuning, while often achieving comparable performance.¹⁰²
 - QLoRA, in particular, enables fine-tuning of quantized LLMs, further reducing memory footprint. For SARI, the fine-tuning dataset would need to be rich in examples of empathetic consular dialogues and factually accurate consular information. The rank 'r' in LoRA/QLoRA can be adjusted to balance style imitation (empathy) with fact memorization capabilities.¹⁰¹ IR-QLoRA, focusing on information retention during quantization and fine-tuning, could further enhance accuracy.¹⁰⁰

5. **Prompt Engineering for Contextual and Empathetic Responses:**

- The design of effective prompts is crucial for guiding the LLM's generation process. Prompts must clearly instruct the LLM to:
 - Base its answer strictly on the provided retrieved context.
 - Adopt the appropriate tone (e.g., empathetic and supportive for SARI; formal and procedural for general inquiries).
 - Maintain factual accuracy and avoid speculation.
 - Provide answers in the language of the query (Bahasa Indonesia or English).
- Techniques like providing few-shot examples within the prompt, or using structured output instructions, can improve response quality. For complex queries requiring multi-step reasoning, incorporating Chain-of-Thought (CoT) prompting or similar methods could be beneficial.⁵²

6. **Orchestration Frameworks:**

- Open-source frameworks like **LangChain** or **LlamaIndex** can be utilized to

build, chain, and manage the various components of the RAG pipeline (data loaders, chunkers, embedders, vector stores, LLMs, and prompting strategies) in Python, facilitating development and experimentation.⁹⁴

The successful implementation of this RAG architecture will depend on careful selection and integration of each component, robust data engineering practices, and continuous evaluation and refinement to ensure the system meets the specific needs of MoFA and WNI.

VII. Conclusion and Strategic Recommendations

The exploration of Artificial Intelligence, particularly through initiatives like SARI and the implementation of Retrieval-Augmented Generation (RAG) systems, presents a transformative opportunity for the Indonesian Ministry of Foreign Affairs (MoFA) to enhance its consular services and the protection of Indonesian citizens (WNI) abroad. The global trend towards AI adoption in public services underscores the potential for increased efficiency, 24/7 accessibility, personalized support, and improved handling of complex consular challenges. SARI, with its focus on providing empathetic and non-discriminatory support to vulnerable migrant workers, exemplifies a human-centric application of AI.

The foundation of such an advanced consular AI system lies in a meticulously curated, multilingual knowledge base derived from official MoFA documents, FAQs, and specialized content, coupled with a robust RAG architecture. This architecture, powered by an Indonesian-focused LLM like SahabatAI (potentially fine-tuned using techniques like QLoRA for domain specificity and empathetic response generation), and supported by appropriate embedding models and vector database technologies, can deliver contextually relevant and factually grounded answers.

However, the journey towards AI integration is not without its challenges. Lessons from global AI deployments highlight the critical importance of addressing data privacy, security, algorithmic bias, and the potential for AI-generated misinformation ("hallucinations"). Public trust and ethical considerations are paramount, especially when AI interacts with citizens on sensitive matters.

Strategic Recommendations for MoFA:

1. **Phased Implementation and Pilot Testing:** Begin with a phased rollout of the SARI chatbot, initially targeting specific user groups or query types. Conduct thorough pilot testing to gather user feedback, assess performance, and identify areas for refinement before broader deployment.

2. **Investment in Knowledge Base Curation:** Continuously invest in the development, curation, and maintenance of a high-quality, accurate, and up-to-date multilingual knowledge base. Establish clear processes for data verification, updates, and version control, particularly for legal and procedural information.
3. **Iterative LLM Fine-tuning and Evaluation:**
 - Fine-tune the chosen LLM (e.g., SahabatAI) iteratively on domain-specific consular data and dialogues demonstrating empathetic communication for SARI.
 - Implement a comprehensive evaluation framework using a combination of automated metrics (e.g., RAGAS metrics like faithfulness, answer relevance, context precision/recall; and NLG metrics like BLEU, ROUGE for fluency⁵²) and rigorous human evaluation to assess factual accuracy, contextual relevance, empathetic quality, and cultural appropriateness.¹¹⁸
4. **Establish Robust Ethical Guidelines and Data Privacy Protocols:** Develop and enforce clear policies for the ethical use of AI in consular services, addressing data privacy (in compliance with Indonesian and international regulations), data security, bias detection and mitigation, and transparency in how AI-generated information is provided.
5. **Capacity Building and Training:** Invest in training programs for MoFA staff to understand, manage, and effectively utilize AI tools. This includes training on interacting with the AI system, interpreting its outputs, and handling escalations for complex cases that AI cannot resolve.
6. **Explore Advanced RAG Techniques:** As the system matures, explore the integration of more advanced RAG techniques, such as Self-RAG or Corrective RAG (CRAG)¹¹⁸, which can improve robustness by allowing the system to self-assess the quality of retrieved information and generated responses, and adapt its strategy accordingly.
7. **User Feedback and Continuous Improvement:** Implement mechanisms for collecting user feedback on their interactions with the AI system. Use this feedback, alongside performance metrics, to drive a continuous cycle of improvement, refining the knowledge base, LLM responses, and user interface.

The successful deployment of AI in Indonesian consular services, spearheaded by initiatives like SARI, requires more than technological prowess. It demands a sustained commitment to a human-centric design philosophy, ensuring that these advanced tools genuinely serve the needs of WNI, build trust through reliability and transparency, and uphold the highest ethical standards. By carefully navigating the opportunities and challenges, MoFA can significantly enhance its capacity to protect

and serve its citizens around the world.

Works cited

1. wwps.microsoft.com, accessed May 15, 2025, <https://wwps.microsoft.com/wp-content/uploads/2025/02/Making-the-case-for-AI-in-Public-Services-ROI-evaluation.pdf>
2. Case Study: AI Implementation in the Government of Estonia ..., accessed May 15, 2025, <https://publicsectornetwork.com/insight/case-study-ai-implementation-in-the-government-of-estonia>
3. How AI Chatbots Improve Local Government Services - Velaro, accessed May 15, 2025, <https://velaro.com/blog/ai-chatbots-in-local-government-engaging-citizens-across-cities-and-states>
4. Government chat bots: 4 examples of how government agencies use chatbots - VirtualSpirits, accessed May 15, 2025, <https://www.virtualspirits.com/government-chat-bots-4-examples-of-how-government-agencies-use-chatbots.aspx>
5. Kementerian Luar Negeri akan Gunakan AI untuk Pelayanan WNI di ..., accessed May 15, 2025, <https://www.tempo.co/internasional/kementerian-luar-negeri-akan-gunakan-ai-untuk-pelayanan-wni-di-luar-negeri-1207028>
6. Indonesian Ministry Of Foreign Affairs And UN Women Launch Humanist AI SARI Chatbots, Equipped With Various Regional Languages - VOI, accessed May 15, 2025, <https://voi.id/en/news/477218>
7. Foreign Ministry to Utilize AI-Based Services for Indonesian Citizens Abroad, accessed May 15, 2025, <https://en.tempo.co/read/1975389/foreign-ministry-to-utilize-ai-based-services-for-indonesian-citizens-abroad>
8. How Governments are Using AI: 8 Real-World Case Studies, accessed May 15, 2025, <https://blog.govnet.co.uk/technology/ai-in-government-case-studies>
9. kemlu.go.id, accessed May 15, 2025, [https://kemlu.go.id/files-service/storage/repositori/65391/01.%20Renstra%20Dit%20PWN%20BHI%202020-2024%20\(FINAL\).pdf](https://kemlu.go.id/files-service/storage/repositori/65391/01.%20Renstra%20Dit%20PWN%20BHI%202020-2024%20(FINAL).pdf)
10. PERAN KEMENTERIAN LUAR NEGERI TERHADAP PERLINDUNGAN WARGA NEGARA INDONESIA DI LUAR NEGERI PADA MASA PANDEMI COVID-19, accessed May 15, 2025, <https://jurnal.alazhar-university.ac.id/index.php/normatif/article/download/174/168>
11. e-ppid.kemlu.go.id, accessed May 15, 2025, <https://e-ppid.kemlu.go.id/storage/619/Renstra-Kemlu-2020-2024.pdf>
12. Tata Kelola Perlindungan Warga Negara Indonesia dalam Melakukan Peran Diplomasi Digital - Journal of Universitas Airlangga, accessed May 15, 2025, <https://e-journal.unair.ac.id/JHI/article/download/17996/10686/73865>
13. peraturan.bpk.go.id, accessed May 15, 2025,

- <https://peraturan.bpk.go.id/Download/130822/Permenlu%20No.%205%20Tahun%202018.pdf>
14. Pendataan Jadi Tantangan Besar dalam Perlindungan WNI di Luar Negeri - VOA Indonesia, accessed May 15, 2025,
<https://www.voaindonesia.com/a/pendataan-jadi-tantangan-besar-dalam-perindungan-wni-di-luar-negeri/6016183.html>
 15. Indonesia - United States Department of State, accessed May 15, 2025,
<https://www.state.gov/reports/2023-country-reports-on-human-rights-practices/indonesia/>
 16. Travel advice and advisories for Indonesia - Travel.gc.ca, accessed May 15, 2025,
<https://travel.gc.ca/destinations/indonesia>
 17. What To Do If Your Indonesian Visa Application is Rejected - cptcorporate, accessed May 15, 2025,
<https://cptcorporate.com/what-to-do-if-your-indonesian-visa-application-is-rejected/>
 18. Indonesia Updates Multiple-Entry Visa Rules: Stay Up to 180 Days Without Exit - travelobiz, accessed May 15, 2025,
<https://travelobiz.com/indonesia-multiple-entry-visa-update-180-day-stay-without-exit/>
 19. Case Studies of How Generative Artificial Intelligence is Changing Work | Congress.gov, accessed May 15, 2025,
<https://www.congress.gov/crs-product/IN12537>
 20. How RAG Systems Improve Public Sector Management - Datategy, accessed May 15, 2025,
<https://www.datategy.net/2025/04/07/how-rag-systems-improve-public-sector-management/>
 21. Retrieval augmented generation for 10 large language models and ..., accessed May 15, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11971376/>
 22. What Is RAG Architecture? A New Approach to LLMs | Cohere, accessed May 15, 2025, <https://cohere.com/blog/rag-architecture>
 23. 7 Practical Applications of RAG Models and Their Impact on Society ..., accessed May 15, 2025,
<https://hyperight.com/7-practical-applications-of-rag-models-and-their-impact-on-society/>
 24. SARI App: Indonesia and UN Women's Effort to Protect Female Migrant Workers - RRI, accessed May 15, 2025,
<https://www.rri.co.id/internasional/1465199/sari-app-indonesia-and-un-women-s-effort-to-protect-female-migrant-workers>
 25. indonesia.iom.int, accessed May 15, 2025,
https://indonesia.iom.int/sites/g/files/tmzbd1491/files/documents/2025-04/key-results-infosheet_mmpmf_eng.pdf
 26. Migration MPTF Final Report, accessed May 15, 2025,
https://mptf.undp.org/sites/default/files/documents/2025-02/final_report_2024_migration_mptf_indonesia.pdf
 27. Indonesia, UN Women Launch AI Chatbot SARI to Protect Female Migrant

- Workers - Jakarta Daily, accessed May 15, 2025,
<https://www.jakartadaily.id/international/16214999370/indonesia-un-women-launch-ai-chatbot-sari-to-protect-female-migrant-workers>
28. Sahabat-AI, accessed May 15, 2025, <https://sahabat-ai.com/>
 29. GoToCompany/llama3-8b-cpt-sahabatai-v1-instruct - Hugging Face, accessed May 15, 2025,
<https://huggingface.co/GoToCompany/llama3-8b-cpt-sahabatai-v1-instruct>
 30. GoToCompany/llama3-8b-cpt-sahabatai-v1-base · Hugging Face, accessed May 15, 2025,
<https://huggingface.co/GoToCompany/llama3-8b-cpt-sahabatai-v1-base>
 31. AI Day drives Indonesia's AI sovereignty, with Indosat leading solutions and connectivity, accessed May 15, 2025,
<https://www.thejakartapost.com/business/2024/11/18/ai-day-drives-indonesias-ai-sovereignty-with-indosat-leading-solutions-and-connectivity.html>
 32. Indosat Ooredoo Hutchison & GoTo Launch Indonesian LLM, Sahabat-AI - The Fast Mode, accessed May 15, 2025,
<https://www.thefastmode.com/technology-solutions/38190-indosat-ooredoo-hutchison-goto-launch-indonesian-llm-sahabat-ai>
 33. Llama3 8b Cpt Sahabatai V1 Instruct · Models - Dataloop AI, accessed May 15, 2025,
https://dataloop.ai/library/model/gotocompany_llama3-8b-cpt-sahabatai-v1-instruct/
 34. Evaluating LLMs for Bahasa Indonesia: SEA-LIONv3 vs SahabatAI-v1 – We SUPA AI Blog, accessed May 15, 2025,
<https://blog.supa.so/2025/02/21/evaluating-llms-for-bahasa-indonesia-sea-lionv3-vs-sahabatai-v1/>
 35. Kemlu GO - Apps on Google Play, accessed May 15, 2025,
<https://play.google.com/store/apps/details?id=id.go.kemlu.app>
 36. Portal Peduli WNI Mobile - Apps on Google Play, accessed May 15, 2025,
<https://play.google.com/store/apps/details?id=id.go.kemlu.peduliwni.mobile>
 37. Privacy Policy - Safe Travel, accessed May 15, 2025,
<https://bo-safetravel.kemlu.go.id/privacy-policy>
 38. Safe Travel for Android - Download the APK from Uptodown, accessed May 15, 2025, <https://safe-travel.en.uptodown.com/android>
 39. Social and Political Issues on Sustainable Development in the Post Covid-19 Crisis - OAPEN Library, accessed May 15, 2025,
<https://library.oapen.org/bitstream/id/ab7714c8-fcf0-4d8d-b988-fe4efed50532/9781000594195.pdf>
 40. ipgsc 2018 - Departemen Ilmu Hubungan Internasional FISIP UI - Universitas Indonesia, accessed May 15, 2025,
<https://ir.fisip.ui.ac.id/storage/2020/06/Proceedings-IPGSC-2018.pdf>
 41. Public Health in Asia during the COVID-19 Pandemic - OAPEN Library, accessed May 15, 2025,
<https://library.oapen.org/bitstream/id/88160bce-a480-4b1c-8998-8ab32b123d97/9789048555246.pdf>

42. Copyright - Lintang - Untar, accessed May 15, 2025, https://lintang.untar.ac.id/repository/penelitian/buktipenelitian_10706003_6A031024104306.pdf
43. Perlindungan WNI - Los Angeles - Portal Kemlu, accessed May 15, 2025, <https://kemlu.go.id/losangeles/perlindungan-wni>
44. The DHS Launches Plans to Use AI in Immigration - Docketwise, accessed May 15, 2025, <https://www.docketwise.com/blog/dhs-ai>
45. Department of State AI Inventory 2024 - United States Department of ..., accessed May 15, 2025, <https://2021-2025.state.gov/department-of-state-ai-inventory-2024/>
46. What Is Retrieval Augmented Generation, and How Are State and Local Agencies Using It?, accessed May 15, 2025, <https://statetechmagazine.com/article/2025/02/what-is-rag-perfcon>
47. How Generative-AI can be Effectively used in Government Chatbots —A Joint Experimental Research based on Large Language Model - arXiv, accessed May 15, 2025, <https://arxiv.org/pdf/2312.02181>
48. Understanding the evaluations role in measuring the impact of AI interventions across Government - Digital Trade Blog, accessed May 15, 2025, <https://digitaltrade.blog.gov.uk/2025/04/14/understanding-the-evaluations-role-in-measuring-the-impact-of-ai-interventions-across-government/>
49. AI Risk Assessment Tools for Public Sector - Avero Advisors, accessed May 15, 2025, <https://averoadvisors.com/blog/ai-risk-assessment-tools-for-public-sector/>
50. Navigating the risks of public LLMs: Why government agencies need safe, targeted AI solutions for their investigations - Comtrac, accessed May 15, 2025, <https://www.comtrac.com.au/blog/navigating-the-risks-of-public-llms-why-government-agencies-need-safe-targeted-ai-solutions-for-their-investigations>
51. RAG LLMs are Not Safer: A Safety Analysis of Retrieval-Augmented Generation for Large Language Models - ACL Anthology, accessed May 15, 2025, <https://aclanthology.org/2025.naacl-long.281.pdf>
52. Retrieval Augmented Generation (RAG) Question Answering, accessed May 15, 2025, <https://docs.cloud.deepset.ai/docs/generative-question-answering>
53. Opportunities and Challenges for Large Language Models in Primary Health Care - PMC, accessed May 15, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11960148/>
54. Use of Retrieval-Augmented Large Language Model for COVID-19 ..., accessed May 15, 2025, <https://www.jmir.org/2025/1/e66098/>
55. MENTERI LUAR NEGERI REPUBLIK INDONESIA PERATURAN MENTERI LUAR NEGERI REPUBLIK INDONESIA NOMOR 14 TAHUN 2022 TENTANG TATA CARA L, accessed May 15, 2025, <https://peraturan.bpk.go.id/Download/310940/PERMENLU-14-2022.pdf>
56. Permenlu No. 13 Tahun 2019.pdf - Peraturan BPK, accessed May 15, 2025, <https://peraturan.bpk.go.id/Download/130666/Permenlu%20No.%2013%20Tahun%202019.pdf>
57. FAQ - Portal Kemlu, accessed May 15, 2025, <https://kemlu.go.id/faq>
58. KEDUTAAN BESAR REPUBLIK INDONESIA MEXICO CITY FUNGSI PROTOKOL DAN

- KONSULER LEGALISASI DOKUMEN Legalisasi adalah pengesahan tand, accessed May 15, 2025,
https://kemlu.go.id/files-service/storage/submenu/additional_file/172972013067196f4219a4c_2024_LEGALISASI_Dokumen_Ind_.pdf
59. Legalisasi Dokumen - Lima, accessed May 15, 2025,
<https://kemlu.go.id/lima/pelayanan-perwakilan/legalisasi-dokumen->
 60. Legalisasi Dokumen Warga Negara Indonesia - Den Haag, accessed May 15, 2025,
<https://kemlu.go.id/denhaag/pelayanan-perwakilan/pelayanan-konsuler/legalisasi-dokumen-warga-negara-indonesia->
 61. Kementerian Luar Negeri RI - Portal Kemlu, accessed May 15, 2025,
<https://fe-non-production.apps.opppd2-dev.layanan.go.id/kontak>
 62. Layanan Diplomatik Terpadu Satu Pintu - Ditjen Protkons Kemlu RI, accessed May 15, 2025, <https://layanandiplomatik.kemlu.go.id/guest/faq>
 63. KEDUTAAN BESAR REPUBLIK INDONESIA MEXICO CITY FUNGSI PROTOKOL DAN KONSULER PERSYARATAN PEMBUATAN VISA DINAS DAN DIPLOMATIK Peme, accessed May 15, 2025,
https://kemlu.go.id/files-service/storage/submenu/additional_file/1729875925671bcfd597a5b_2024_VISA_Informasi_Visa_Official_Diplomatik.pdf
 64. Indonesia - Travel.gov - U.S. Department of State, accessed May 15, 2025,
<https://travel.state.gov/content/special-issuance-agency-home/en/spec-issuance-agency/official-diplomatic-visa-information/indonesia.html>
 65. The Embassy of the Republic of Indonesia - Washington, accessed May 15, 2025,
<https://kemlu.go.id/washington/tentang-perwakilan/kontak-kami>
 66. accessed January 1, 1970, <https://kbriparis.fr/layanan-konsuler/paspor>
 67. FAQ | Indonesia - KBRI Den Haag, accessed May 15, 2025,
<https://indonesia.nl/en/faq/>
 68. Welcome to the Indonesian Directorate General of Immigration Website - Direktorat Jenderal Imigrasi, accessed May 15, 2025,
<https://www.imigrasi.go.id/index?lang=en-US>
 69. accessed January 1, 1970,
<https://kemlu.go.id/singapore/pages/layanan-konsuler/id>
 70. accessed January 1, 1970,
<https://kemlu.go.id/kualalumpur/pages/layanan-kekonsuleran/id>
 71. 0 RENCANA STRATEGIS (RENSTRA) KEDUTAAN BESAR REPUBLIK INDONESIA LONDON, INGGRIS TAHUN 2020 - Portal Kemlu, accessed May 15, 2025,
https://kemlu.go.id/files/submenu/additional_file/4355-files-statis-pages-perwakilan-London-href-contains-pdf-5-shared.pdf
 72. Renstra Kemlu 2020 - 2024 | PDF | Bisnis | Pengelolaan Keuangan & Uang - Scribd, accessed May 15, 2025,
<https://id.scribd.com/document/539794959/Renstra-Kemlu-2020-2024>
 73. LAPORAN KINERJA - Kementerian Imigrasi dan Pemasarakatan RI, accessed May 15, 2025,
https://www.kemenimipas.go.id/attachments/Produk_Perencanaan/LAKIP/LKIP_KE_MENKUMHAM_2021.pdf
 74. MENTERI LUAR NEGERI REPUBLIK INDONESIA PERATURAN MENTERI LUAR

NEGERI REPUBLIK INDONESIA NOMOR 2 TAHUN 2022 TENTANG KLASIFIKASI,
accessed May 15, 2025,

<https://peraturan.bpk.go.id/Download/339023/Permenlu%20No.%202%20Tahun%202022.pdf>

75. Portal Kemlu, accessed May 15, 2025, <https://kemlu.go.id/portal/id>
76. Portal Kemlu, accessed May 15, 2025, <https://kemlu.go.id/>
77. Ministry of Foreign Affairs (Indonesia) - Wikipedia, accessed May 15, 2025, [https://en.wikipedia.org/wiki/Ministry_of_Foreign_Affairs_\(Indonesia\)](https://en.wikipedia.org/wiki/Ministry_of_Foreign_Affairs_(Indonesia))
78. Document Legalization - e-Consular Service KBRI WDC, accessed May 15, 2025, <https://consular.embassyofindonesia.org/page/documentlegalization.html>
79. What you'll need to bring - U.S. Embassy & Consulates in Indonesia, accessed May 15, 2025, <https://id.usembassy.gov/what-youll-need-to-bring/>
80. AI for Gender Equality: UN Women AI School Opens for Changemakers, accessed May 15, 2025, <https://asiapacific.unwomen.org/en/stories/announcement/2025/03/un-women-ai-school-opens-for-changemakers>
81. UN Women Indonesia, accessed May 15, 2025, <https://asiapacific.unwomen.org/en/countries/indonesia>
82. What is RAG? - Retrieval-Augmented Generation AI Explained - AWS, accessed May 15, 2025, <https://aws.amazon.com/what-is/retrieval-augmented-generation/>
83. What is Retrieval-Augmented Generation (RAG)? | NVIDIA, accessed May 15, 2025, <https://www.nvidia.com/en-us/glossary/retrieval-augmented-generation/>
84. Chunking Strategies for LLM Applications | Pinecone, accessed May 15, 2025, <https://www.pinecone.io/learn/chunking-strategies/>
85. 15 Chunking Techniques to Build Exceptional RAGs Systems - Analytics Vidhya, accessed May 15, 2025, <https://www.analyticsvidhya.com/blog/2024/10/chunking-techniques-to-build-exceptional-rag-systems/>
86. 7 Chunking Strategies in RAG You Need To Know - F22 Labs, accessed May 15, 2025, <https://www.f22labs.com/blogs/7-chunking-strategies-in-rag-you-need-to-know/>
87. LUSIFER: Language Universal Space Integration for Enhanced Multilingual Embeddings with Large Language Models - arXiv, accessed May 15, 2025, <https://arxiv.org/html/2501.00874v1>
88. LUSIFER: Language Universal Space Integration for Enhanced Multilingual Embeddings with Large Language Models - arXiv, accessed May 15, 2025, <https://arxiv.org/html/2501.00874v3>
89. sentence-transformers/paraphrase-multilingual-mpnet-base-v2 ..., accessed May 15, 2025, <https://huggingface.co/sentence-transformers/paraphrase-multilingual-mpnet-base-v2>
90. MTEB Leaderboard - a Hugging Face Space by mteb, accessed May 15, 2025, <https://huggingface.co/spaces/mteb/leaderboard>
91. RAG vector database explained - Writer, accessed May 15, 2025, <https://writer.com/engineering/rag-vector-database/>

92. LegalBench-RAG: A Benchmark for Retrieval-Augmented Generation in the Legal Domain - arXiv, accessed May 15, 2025, <http://arxiv.org/pdf/2408.10343>
93. QuIM-RAG: Advancing Retrieval-Augmented Generation with Inverted Question Matching for Enhanced QA Performance - arXiv, accessed May 15, 2025, <https://arxiv.org/html/2501.02702v1>
94. 15 Best Open-Source RAG Frameworks in 2025 - Firecrawl, accessed May 15, 2025, <https://www.firecrawl.dev/blog/best-open-source-rag-frameworks>
95. Sahabat-AI - a GoToCompany Collection - Hugging Face, accessed May 15, 2025, <https://huggingface.co/collections/GoToCompany/sahabat-ai-672af7b248f5dfd39ae2403>
96. Speaking in Code: Contextualizing Large Language Models in Southeast Asia, accessed May 15, 2025, <https://carnegieendowment.org/research/2025/01/speaking-in-code-contextualizing-large-language-models-in-southeast-asia>
97. Gemma2 9b Cpt Sahabatai V1 Base · Models - Dataloop AI, accessed May 15, 2025, https://dataloop.ai/library/model/gotocompany_gemma2-9b-cpt-sahabatai-v1-base/
98. LoRA - Hugging Face, accessed May 15, 2025, https://huggingface.co/docs/peft/main/en/conceptual_guides/lora
99. Fine-tuning 20B LLMs with RLHF on a 24GB consumer GPU, accessed May 15, 2025, <https://huggingface.co/blog/trl-peft>
100. arxiv.org, accessed May 15, 2025, <https://arxiv.org/abs/2402.05445>
101. arxiv.org, accessed May 15, 2025, <https://arxiv.org/abs/2406.08582>
102. Fine-Tuning LLMs: A Guide With Examples - DataCamp, accessed May 15, 2025, <https://www.datacamp.com/tutorial/fine-tuning-large-language-models>
103. Custom Fine-Tuning for Domain-Specific LLMs - MachineLearningMastery.com, accessed May 15, 2025, <https://machinelearningmastery.com/custom-fine-tuning-for-domain-specific-llms/>
104. Think-on-Graph 2.0: Deep and Faithful Large Language Model Reasoning with Knowledge-guided Retrieval Augmented Generation - arXiv, accessed May 15, 2025, <https://arxiv.org/html/2407.10805v7>
105. Llamaindex vs Langchain: What's the difference? - IBM, accessed May 15, 2025, <https://www.ibm.com/think/topics/llamaindex-vs-langchain>
106. arxiv.org, accessed May 15, 2025, <https://arxiv.org/abs/2005.11401>
107. Metrics - Ragas, accessed May 15, 2025, <https://docs.ragas.io/en/stable/concepts/metrics/index.html>
108. arxiv.org, accessed May 15, 2025, <https://arxiv.org/abs/2402.11562>
109. arxiv.org, accessed May 15, 2025, <https://arxiv.org/abs/2307.03109>
110. Yasser Shohoud Maged Shoman Sarah Abdelazim - arXiv, accessed May 15, 2025, <https://arxiv.org/abs/2311.05120>
111. arxiv.org, accessed May 15, 2025, <https://arxiv.org/abs/2401.05563>
112. arxiv.org, accessed May 15, 2025, <https://arxiv.org/abs/2312.05248>
113. Evaluating - LlamaIndex, accessed May 15, 2025,

- https://docs.llamaindex.ai/en/stable/module_guides/evaluating/root.html
114. Evaluation of RAG Metrics for Question Answering in the Telecom Domain - arXiv, accessed May 15, 2025, <https://arxiv.org/html/2407.12873v1>
 115. Evaluation of RAG Metrics for Question Answering in the Telecom Domain - arXiv, accessed May 15, 2025, <https://arxiv.org/abs/2407.12873>
 116. grading conversational responses of chatbots - arXiv, accessed May 15, 2025, <https://arxiv.org/pdf/2303.12038>
 117. Evaluating the performance of multilingual models in answer extraction and question generation - PMC, accessed May 15, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11226668/>
 118. WixQA: A Multi-Dataset Benchmark for Enterprise Retrieval-Augmented Generation - arXiv, accessed May 15, 2025, <https://arxiv.org/html/2505.08643v1>
 119. Retrieval-Augmented Generation with Vector Stores, Knowledge Graphs, and Hierarchical Non-negative Matrix Factorization - arXiv, accessed May 15, 2025, <https://arxiv.org/html/2502.20364v2>
 120. WixQA: A Multi-Dataset Benchmark for Enterprise Retrieval-Augmented Generation - arXiv, accessed May 15, 2025, <https://arxiv.org/pdf/2505.08643?>
 121. openreview.net, accessed May 15, 2025, <https://openreview.net/pdf?id=JnWJbrnaUE>
 122. A Survey on Knowledge-Oriented Retrieval-Augmented Generation - arXiv, accessed May 15, 2025, <https://arxiv.org/html/2503.10677v2>