A Global Hierarchical LLM Framework for Precise Responses through Subscriptions to Country LLMs

Anthony Jairam, Tamika Ramkissoon, Kevin Baboolal and Patrick Hosein
Department of Computer Science
The University of the West Indies
St. Augustine, Trinidad
anthony.jairam@my.uwi.edu, tamika.ramkissoon@my.uwi.edu, kevin@lab.tt, patrick.hosein@uwi.edu

***Abstract*—The increasing adoption of Large Language Models (LLM) and Retrieval-Augmented Generation (RAG) pipelines is transforming information access, yet their efficacy is significantly hampered in developing regions by a scarcity of relevant, localized data and compute. Furthermore, reliance on centralized RAG architectures raises critical data governance concerns, as organizations and governments are reluctant to forfeit control of sensitive data, thereby centralizing economic benefits and exacerbating regional digital divides. To address these challenges, this paper introduces a novel framework that leverages a distributed network of specialized LLMs rooted in the principles of Distributed Retrieval-Augmented Generation (DRAG). Our proposed architecture eliminates the need for a centralized knowledge base, allowing data owners to retain full control over their assets. We detail an intelligent query-routing mechanism for efficient knowledge discovery within the network and, critically, propose a comprehensive incentive model to encourage the active participation of regional entities. By fostering a collaborative, yet decentralized, ecosystem, this framework facilitates equitable data governance and paves the way for a sustainable, regional data economy, empowering local communities and enhancing the accuracy of LLMs.**

***Index Terms*—data governance, RAG, aggregator network, orchestration, incentive model, MCP network, sovereign network, hierarchical LLM framework**

I. Introduction

Small Island Developing States (SIDS) often face unique challenges in building their digital economies. These challenges include limited technological infrastructure, small domestic markets and a high dependence on external actors all of which often slow their ability to take advantage of emerging technologies [1]. While internet access in SIDS has improved with 67% of the population being online in 2023 similar to the global average, barriers such as poor infrastructure quality, affordability and limited local capacity remain [2]. A recent UNESCO assessment makes this gap even clearer by highlighting over 70% of SIDS lack quality datasets for AI research, some do not have official AI strategies and nearly half lack data governance frameworks [3] These issues make it difficult for SIDS to adopt advanced AI systems such as Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) which are already reshaping knowledge-intensive tasks around the world [4].

At the same time global adoption of LLMs has been rapid. Tools like ChatGPT now process more than 2.5 billion prompts every day and reach over 800 million weekly active users [5]. Surveys show that nearly two-thirds of internet users worldwide have tried an AI chatbot in the past year [6]. This widespread adoption highlights a stark contrast: while the rest of the world is increasingly engaging with these technologies, SIDS remain underserved due to structural gaps in data, infrastructure and governance. If these gaps are not addressed, SIDS risk being left behind as passive consumers rather than active participants in the global AI economy.

One promising step toward addressing these issues is Distributed Retrieval-Augmented Generation (DRAG), introduced by Xu et al. [7]. DRAG tackles the limitations of centralized RAG by distributing knowledge and computation across peer-to-peer networks which improves privacy, scalability and resilience. It uses a Topic-Aware Random Walk (TARW) algorithm for efficient query routing and has been shown to achieve performance close to centralized RAG while reducing communication overhead. However as Xu et al. note, DRAG can only be sustainable if participants are motivated to share their knowledge and resources. Without clear incentives, regional actors may have little reason to join or maintain such networks.

This paper builds on DRAG as a technical foundation and adapts it to the context of SIDS by introducing a framework of incentives to encourage long-term participation. The idea is to move beyond DRAG as a technical system and frame it as both a technical and socio-economic model for SIDS. In our vision, users interact with state-of-the-art foundation models that connect to a regional and general-purpose model hosted within a SIDS. This regional hub ensures sovereignty and governance while linking out to domain-specific RAG pipelines in areas such as health, fintech, agriculture and education. Each of these pipelines maintains control of its local data while still contributing to the broader distributed ecosystem. Incentives, ranging from reputational benefits to financial rewards or opportunities for collaboration are built in to encourage organizations within SIDS to contribute. In doing so, this approach fosters a sustainable local data economy while helping SIDS close the gap with global AI adoption.

The contributions of this paper are threefold, first, we analyze the barriers that Small Island Developing States (SIDS) face in adopting centralized RAG systems and demonstrate why the Distributed Retrieval-Augmented Generation (DRAG) paradigm provides a stronger foundation for sovereign and equitable AI adoption. Second, we introduce the Distributed Sovereign Retrieval-Augmented Generation (DS-RAG) framework, a multi-layered architecture specifically adapted for the needs of SIDS. This framework integrates four core components; Global LLM Layer, Sovereign Participant Registry, Distributed Sovereign Network, and Reward Node, ensuring data sovereignty, efficient query routing, and incentive-aligned participation. Third, we evaluate how DS-RAG compares to baseline architectures, highlighting its potential to promote fair data governance, strengthen regional digital economies, and foster long-term sustainability in the global AI ecosystem.

Section II reviews related work on centralized and distributed RAG, data governance, and incentive mechanisms for distributed networks. Section III presents the DS-RAG framework, with detailed discussion of its layered architecture and system design. Section IV focuses specifically on the Reward System, which underpins the sustainability of DS-RAG by balancing reputational, financial and collaborative incentives. Section VI discusses the broader implications for equitable data governance and resilient digital economies in SIDS, while also outlining use cases, limitations, and areas for future work. Then, Section VI concludes by summarizing the paper's contributions and offering directions for further research.

II. Related Work

Although LLMs have been improving exponentially over the past few years, hallucinations persist as one of their most common intrinsic problems [8], [9]. This phenomenon arises from a mismatch between the model's internal approximations of its training data and real-world context [10]. This problem is particularly present in developing regions, which often have underrepresented languages, dialects, and cultural contexts that LLMs are not optimized to handle [11]. Cultural benchmarks created to evaluate knowledge across diverse cultures and dialects, such as BLEnD [12] and CulturalBench [13] have shown that the hallucination problem is further exacerbated in these low-resource regions, leading a higher frequency of hallucinations or culturally irrelevant outputs. This undermines the LLM's reliability and raises fairness and bias concerns [14]

Various mitigation strategies have been proposed over the years [15], [16], including many Prompt Engineering [17] and Fine-Tuning approaches [18]. While these offer some improvements, they may still suffer from hallucinations and lack access to dynamic, culturally relevant data. Retrieval-Augmented Generation (RAG) [4] has emerged as one of the most effective methods for mitigating hallucinations [19]. RAG is a model-agnostic framework that provides the LLMs with external knowledge and context before generating responses. But as outlined in Section I, the limitations of RAG highlight the need for region-specific approaches to data collection, model adaptation, and evaluation metrics tailored to low-resource contexts.

The DRAG [7] framework address these challenges by enabling targeted peer discovery within a peer-to-peer network, facilitating knowledge sharing in decentralized environments. DRAG was seen to be a promising scheme in primarily privacy-sensitive and resource-constrained edge environments. Instead of being limited to edge devices, this framework can be adapted to suit the needs of SIDS organizations.

III. Proposed Framework

To address the challenges of data privacy and lack of real-time culturally-aware knowledge, we introduce the Distributed Sovereign Retrieval-Augmented Generation (DS-RAG) framework, adapted from the principles of DRAG. This framework aims to create a decentralized data economy, independent SIDS participants to provided localized context to Global LLM Platforms. Efficient query routing is done though the use of a Registry node, and a reward system is also proposed to encourage participation in the DS-RAG network. This approach ensures that data owners retain full control over their sensitive information while contributing to a richer, more accurate data ecosystem.

The architecture is composed of four core nodes: Global LLM (Aggregator Node), Distributed Sovereign Network, Distributed Participant Registry, and the Reward System Node. The iteration of these components is illustrated in Figure 1 and is designed to ensure participants are fully in control of their data promoting a sustainable system. The following Subsections will dive deeper into each layer of the framework and detail each of the key steps.

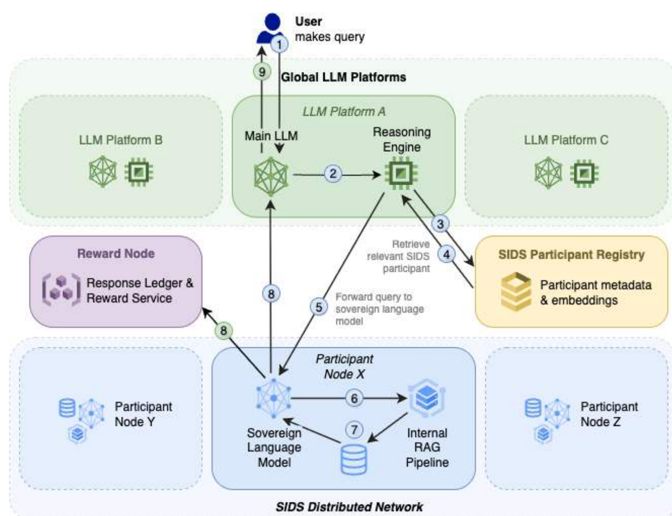

Fig. 1. The query lifecycle of the proposed DS-RAG Framework, illustrating the decentralized interaction from user query to rewarded data retrieval.

A. Global LLM Layer

The lifecycle begins with the Global LLM platform, acting as an Aggregator LLM. As discussed in Section I, while these state-of-the-art models (such as ChatGPT, Gemini, Claude) excel at general knowledge, they lack the real-time and cultural context needed by SIDS. Most Global LLMs have internal orchestrators (or reasoning engines) that serve as the core infrastructure for managing communication and information flow among agents. This is a key component needed to the communication between the Global LLM and the Local LLM Network. In Step 1 and 2, the internal orchestrator analyzes the user's query and identifies a knowledge gap. The engine is then responsible for initializing a call to begin the DS-RAG pipeline.

B. Sovereign Participant Registry

The Global LLM first pings the Sovereign Participant Registry. The Participant Registry's sole function is to maintain a secure database of all participating SIDS organizations. This allows the Aggregator LLM to discover what Sovereign Nodes are accessible. It contains non-sensitive metadata about each Sovereign Node in the network, as well as secure endpoints and a pre-computed embedding vector of their knowledge domain. This Node acts as a DNS-like service for the framework. The Aggregator LLM queries the registry in Step 3, and a similarity search is done to return the most relevant Participant Node in Step 4. If the similarity score falls below a specified threshold then no result will be returned. The Orchestrator Framework will then directly forward the query to the Participant Node in Step 5.

C. Distributed Sovereign Network

The core of our framework is the Distributed Sovereign Network layer. Here contains all of the Sovereign Participant Nodes. Its primary job isn't to be a generalist chatbot but to act as an intelligent 'interface' between the external query and the node's internal database and responsible for query interpretation. Each participant will have full control of their implementation details and infrastructure. For example, using a Small Language Model (SLM) instead of complex Large Langue Models (LLMs) will ease many of the infrastructure gaps present in SIDS, allowing for easier adoption. Additionally, having full control of their RAG pipeline and what data is accessed leads to fair data governance in the region. The node's Local RAG pipeline processes the query internally, securely accessing their data (Steps 6 and 7). In this layer we also have a Sovereign Gateway as a mandatory sub-component of every Participant. This Gateway's responsibilities would include:

- Authentication: Validating incoming requests.
- Protocol Adherence: Handling all LINC/MCP communication.
- Metering: Securely logging usage data and generating the signed receipts for the Reward Node.
- Protection: Acting as a firewall and rate-limiter to protect the internal model.

This architectural separation via a gateway clarifies the security model, simplifies development for participants (who could use a standardized open-source Gateway), and creates a logical point for enforcing network-wide policies and collecting billing data. This approach guarantees data sovereignty, as only sanitized snippets of the participant is returned to the Aggregator LLM.

D. Reward Node

Once the Participant Model sends the generated response to the Aggregator LLM in Step 8, it simultaneously logs the transaction in the Reward Node. This component ensures the economic sustainability and fairness promoting usage and wider adoption. The Reward Node is detailed further in Section IV.

IV. Reward System

The reward system represents the socio-economic foundation of the DS-RAG framework, transforming participation into tangible value while ensuring fairness, sustainability, and long-term engagement. By combining reputational recognition, financial compensation and opportunities for collaboration, the system appeals to a wide spectrum of actors including universities, government agencies, startups and NGOs, encouraging them to actively contribute localized knowledge and resources. This approach is particularly critical for SIDS, where incentives must be sufficiently attractive to offset resource limitations while reinforcing principles of data sovereignty.

The following subsections elaborate on the four pillars of the reward system, supported by practical examples that illustrate their potential applications in SIDS contexts.

A. Reputation Scoring

Reputation acts as a non-monetary incentive that enhances trust and prestige within the network. Each participating node is continuously assessed based on response accuracy, contextual relevance, and reliability. High-reputation nodes are prioritized during query routing, increasing their visibility and influence. For example, a university specializing in marine science may consistently provide high-quality oceanographic data. As its reputation score rises, the institution becomes a preferred source for queries related to fisheries or coastal management, cementing its role as a regional knowledge hub.

Similarly, a fintech startup offering anonymized transaction data might develop a reputation for reliability, positioning itself as a credible partner for international researchers exploring digital financial inclusion in SIDS. Over time, reputation itself becomes an institutional asset, comparable to academic citations or professional endorsements, strengthening the standing of participants both regionally and globally.

B. Financial Compensation

Financial incentives ensure that participation remains economically viable, particularly for organizations with limited budgets. Rewards are tied not to the sheer volume of data contributed but to quality-based metrics, including contextual accuracy, timeliness, and compliance with governance standards. A Ministry of Health that provides accurate epidemiological data could, for instance, receive financial compensation, offsetting the costs of maintaining and updating public health databases.

An agricultural cooperative sharing soil quality and crop yield data may also earn monetary rewards that can be reinvested in digital tools or farmer training programs, creating a developmental feedback loop within the community. This model allows smaller organizations with niche but high-quality contributions to compete equitably alongside larger institutions, thereby promoting inclusivity.

C. Opportunities for Collaboration

Beyond direct compensation, the reward system fosters strategic partnerships and shared projects. High-performing nodes may be offered privileged access to collaborative initiatives, which amplify their regional and global presence. A university node with expertise in climate modeling may be invited to join a cross-island consortium focused on disaster resilience, gaining both visibility and research funding. A startup contributing renewable energy data could collaborate with regional energy ministries in pilot projects, strengthening its market position and enabling joint ventures. Likewise, NGOs providing culturally specific datasets, such as indigenous agricultural practices or traditional medicine may be invited to co-author studies with international institutions, ensuring their knowledge is acknowledged and preserved. These opportunities encourage participants to view the DS-RAG not only as a data-sharing framework but also as a gateway to innovation, partnerships, and socio-economic growth.

D. Transparency and Governance

To preserve legitimacy and trust, all reputational updates, financial transactions, and collaborative opportunities are logged within the Reward Node using tamper-resistant mechanisms. Oversight is proposed through neutral international bodies such as the International Telecommunication Union (ITU) or the United Nations Development Programme (UNDP), which can mediate disputes and validate fairness. If two organizations dispute ownership of a fisheries dataset, the oversight body could reference timestamps and metadata within the Reward Node to fairly assign credit and compensation. This governance structure mitigates fears of exploitation, provides institutional legitimacy, and ensures that incentives operate in alignment with principles of fair data governance.

By intertwining reputation, financial returns, collaboration, and governance, the reward system ensures that SIDS organizations are not only motivated to join but also supported to thrive within the DS-RAG framework. Reputation enhances credibility, financial rewards provide immediate value, collaboration strengthens regional integration, and governance ensures fairness and accountability. Collectively, these mechanisms create a balanced ecosystem that promotes sustainable participation, enabling SIDS to bridge the gap with global AI adoption while retaining sovereignty over their data assets.

V. Comparative Analysis

Consider the following example where a user asks a Global LLM such as ChatGPT:

User: *"List all national court rulings in Trinidad and Tobago from the last 24 months related to intellectual property disputes in the energy sector."*

A. Scenario 1: Without DS-RAG

In Figure 2a, the DS-RAG framework is not used. Therefore given the recent context needed from the users query, the answer lies beyond the model's knowledge cutoff. The model's internal orchestration framework may attempt real-time retrieval via a web search, but the absence of indexed public data results in a retrieval failure. Consequently, the model is generates a response based on irrelevant data or hallucinates a response.

B. Scenario 2: Using DS-RAG network

Now consider Figure 2b, the Global LLM has integrated with the DS-RAG network. In this scenario, the internal orchestration framework realizes it's need for additional context, and the DS-RAG network is invoked to resolve the query. ChatGPT first pings the SIDS Registry Node to determine which Sovereign Node will be best equipped to handle the query. Trinidad's Sovereign LLM ranks the highest, so the query is forwarded. Trinidad's Node then triggers it's internal RAG pipeline to obtain the relevant data. The data is then returned to ChatGPT and it aggregates a response for the user, providing a precise and verifiable answer. Trinidad's Node also logs the transaction in the Reward Layer to receive the incentive.

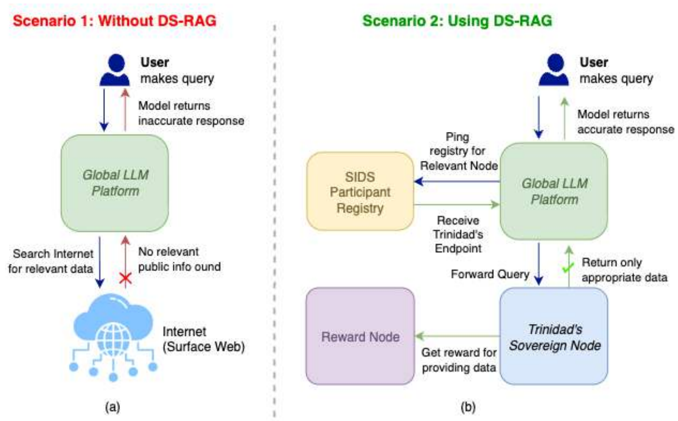

Fig. 2. Comparative Analysis of Query Resolution. (a) Shows the standard process where a Global LLM fails to find relevant public data and returns an inaccurate response. (b) Shows the DS-RAG protocol, which discovers the sovereign node to provide an accurate, contextual answer.

VI. Discussion

The Sovereign Retrieval-Augmented Generation (S-RAG) protocol introduced in this paper presents more than a novel technical architecture; it offers a socio-technical blueprint for a sovereign and sustainable data economy. By integrating a decentralized network of local data providers with a robust incentive mechanism, the framework directly addresses the critical challenges of data sovereignty, economic exclusion, and contextual bias that hinder the equitable adoption of AI in Small Island Developing States (SIDS). This section discusses the broader implications of this work, acknowledges its limitations, and outlines directions for future research.

A. Adoption and Impact

The success of DS-RAG is highly dependent on widespread adoption, creating the 'Network Effect'. As more high-quality nodes join, the network's value increases which then encourages more SIDS organizations and generates greater rewards for all participants. Furthermore, adoption is made more feasible by leveraging emerging open standards like the Model Context Protocol (MCP), which standardizes communication between LLMs and external data nodes and allows for interoperability of models and infrastructure. Using such protocols would dramatically lower the technical barrier to entry for both SIDS organizations and Global LLMs.

While top-down integration with global providers is a long-term goal, convincing them to join the network pose a significant barrier to entry. However, global LLM platforms can interact with the DS-RAG framework at the user level. Firstly, users can install the framework as a third-party plug-in that support them. Alternatively, users can add DS-RAG as an MCP Service to LLM client platforms that support MCP. These approaches significantly lower the barrier for entry at the Global LLM layer allowing for Aggregator LLMs to access the soverign networks.

For a SIDS organization, the process is designed to be accessible and scalable. For maximal safety and trust, we propose that the soverign network be overseen by a partnership of neutral international bodies, such as the ITU, UNDP, or a UN-chartered data council, similar to how ICANN governs DNS. An interested participant, like a university or government ministry, would apply to this governing body. Upon successful vetting, their metadata will be added to the Distributed Sovereign Network layer and they will expose a secure endpoint for MCP communication. This governance model would provide legitimacy and technical support needed to efficiently onboard new SIDS participants.

B. Limitations

While the DS-RAG protocol offers a robust theoretical framework, its practical implementation faces significant economic, political, and technical challenges that must be acknowledged. This section will analyze these hurdles and identify areas for improvement.

1) Economic Viability: While reputation and financial rewards provide the mechanism for compensation, the long-term sustainability of the DS-RAG network hinges on its overall economic viability. The framework's financial compensation model assumes that Global LLMs will recognize the superior value of verified, sovereign data and be willing to pay a sustainable price per query. The core limitation is the absence of an established market for this specific type of data exchange. Determining a fair price that is high enough to cover the operational costs and provide meaningful revenue for a diverse range of SIDS organizations, yet competitive enough to ensure widespread adoption by global platforms, presents a formidable economic challenge.

2) Data Governance and Policies: The decentralized architecture of DS-RAG does not eliminate the need for centralized governance. Specifically, a significant limitation is the political and logistical challenge of establishing a neutral, trusted governing body to manage the Sovereign Node Directory which determines which participants are on the network. Creating a consensus among multiple SIDS nations, each with its own priorities and regulations, is a particularly difficult task. The long-term success of the framework is therefore dependent on overcoming these real-world political hurdles to build a transnational governance system capable of maintaining the network's integrity and preventing its exploitation for financial gain.

3) Technical Infrastructure: Finally, the framework's effectiveness is limited by the existing digital divide in SIDS. While the framework grants participants the freedom to manage their own infrastructure, this autonomy requires a baseline technical capacity including reliable power, internet, and skilled developers to deploy and maintain their sovereign node. These prerequisites may constitute significant barriers to entry for under-resourced institutions, such as community-based NGOs. This creates a risk of inadvertently favoring more technologically advanced organizations, potentially undermining the goal of equitable participation. Although the adoption of open-source software and efficient SLMs is intended to lower these barriers, the fundamental infrastructural and human capital requirements remain a key limitation to achieving truly inclusive adoption.

C. Future Work

Before deploying a live pilot, the initial phase of future work will focus on the rigorous simulation and evaluation of the DS-RAG protocol in a controlled environment. We will construct a simulated network of heterogeneous nodes to benchmark the efficiency of the discovery and P2P communication layers, measuring projected query latency and throughput under various load conditions. A critical component of this phase is to evaluate the framework's ability to mitigate bias and improve contextual accuracy using established cultural benchmarks like BLEnD and CulturalBench. Furthermore, we will employ agent-based economic simulations to test the fairness and resilience of the proposed reward and reputation models, ensuring the system can prevent monopolistic behavior and provide sustainable incentives for all participants before real-world capital is involved.

This simulated validation will inform the subsequent practical validation through a multi-faceted pilot program. The first track of this program will be technical, aimed at establishing a testnet with a small consortium of partners in Trinidad and Tobago. This will involve deploying the open-source node software, integrating local RAG pipelines with sovereign databases, and benchmarking real-world performance. A key objective is to measure network latency and test the efficacy of various Small Language Models (SLMs) in retrieving specialized, culturally-specific information. The second track will focus on user experience and utility, partnering with a

Global LLM provider to integrate the testnet as an external tool and gathering qualitative feedback on the relevance and accuracy of the synthesized responses.

Concurrently, a parallel research effort is required to design the socio-economic and governance framework essential for the network's long-term success. This involves refining the economic models based on simulation outcomes to ensure they are equitable and cover the operational costs of even the smallest participant nodes. Furthermore, this research will produce a comprehensive governance charter for the network. This foundational document will define the legal and ethical standards for data quality, node vetting, dispute resolution, and the formal process for onboarding new SIDS nations. This dual approach, combining simulated evaluation with practical validation and robust socio-economic design, will provide the holistic foundation needed to scale the DS-RAG protocol into a resilient and trusted data ecosystem.

VII. Conclusion

This paper introduced the DS-RAG (Distributed Sovereign RAG) protocol, a novel socio-technical framework designed to address the critical issues of data governance, economic exclusion, and contextual bias in the age of Large Language Models. Our proposed architecture enables Small Island Developing States (SIDS) to participate in the global AI economy on their own terms, transforming them from passive data sources into active, compensated partners. By leveraging a decentralized network of sovereign data nodes, a secure discovery mechanism, and transparent reward system, DS-RAG provides a viable path for SIDS to monetize their unique data assets without compromising ownership. The reward system, incentivizes consistent contributions, strengthens trust among participants, and fosters long-term commitment to regional knowledge ecosystems. The framework not only enhances the accuracy and cultural relevance of global AI platforms but also catalyzes the development of a sustainable regional data economy that encourages collaboration, innovation, and the preservation of local knowledge. Ultimately, DS-RAG serves as a blueprint for a more equitable and inclusive digital future. It demonstrates that with the right architecture and a robust incentive model, a balance can be struck between global technological advancement and the fundamental right of nations to control their own digital destinies. In essence, DS-RAG redefines the role of SIDS in the AI ecosystem, shifting them from data consumers to empowered contributors, while offering a replicable model for equitable participation in the global digital economy.

References

[1] UN, "Small Island Developing States | Office of the High Representative for the Least Developed Countries, Landlocked Developing Countries and Small Island Developing States," 2025. [Online]. Available: https://www.un.org/ohrlls/content/small-island-developing-states

[2] International Telecommunication Union, "ITU Facts and Figures: Focus on Small Island Developing States," Mar. 2024. [Online]. Available: https://www.itu.int/itu-d/reports/statistics/facts-figures-for-sids

[3] UNESCO, "Needs Assessment for Artificial Intelligence, Digital transformation," Oct. 2024. [Online]. Available: https://www.unesco.org/en/articles/needs-assessment-artificial-intelligence-digital-transformation-and-open-data-small-island

[4] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Apr. 2021, arXiv:2005.11401 [cs]. [Online]. Available: http://arxiv.org/abs/2005.11401

[5] Exploding Topics, "Number of ChatGPT Users (July 2025)," Dec. 2024. [Online]. Available: https://explodingtopics.com/blog/chatgpt-users

[6] J. Dupont, V. Ali, D. Baron, A. Price, and S. Wride, "What Does The Public Think About AI?" *Public First*, Jul. 2024.

[7] C. Xu, L. Gao, Y. Miao, and X. Zheng, "Distributed Retrieval-Augmented Generation," May 2025, arXiv:2505.00443 [cs]. [Online]. Available: http://arxiv.org/abs/2505.00443

[8] H. Qiu, A. R. Fabbri, D. Agarwal, K.-H. Huang, S. Tan, N. Peng, and C.-S. Wu, "Evaluating Cultural and Social Awareness of LLM Web Agents," *arXiv preprint arXiv:2410.23252*, 2025.

[9] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. Bang, D. Chen, W. Dai, H. S. Chan, A. Madotto, and P. Fung, "Survey of Hallucination in Natural Language Generation," *ACM Computing Surveys*, vol. 55, no. 12, pp. 1–38, Dec. 2023, arXiv:2202.03629 [cs]. [Online]. Available: http://arxiv.org/abs/2202.03629

[10] R. Mehta, A. Hoblitzell, J. O'Keefe, H. Jang, and V. Varma, "MetaCheckGPT – A Multi-task Hallucination Detector Using LLM Uncertainty and Meta-models," Apr. 2024, arXiv:2404.06948 [cs]. [Online]. Available: http://arxiv.org/abs/2404.06948

[11] M. Abdelrahman, "Hallucination in Low-Resource Languages: Amplified Risks and Mitigation Strategies for Multilingual LLMs," *Journal of Applied Big Data Analytics, Decision-Making, and Predictive Modelling Systems*, vol. 8, no. 12, pp. 17–24, Dec. 2024, section: Articles. [Online]. Available: https://polarpublications.com/index.php/JABADP/article/view/2024-12-10

[12] J. Myung, N. Lee, Y. Zhou, J. Jin, R. A. Putri, D. Antypas, H. Borkakoty, E. Kim, C. Perez-Almendros, A. A. Ayele, V. Gutiérrez-Basulto, Y. Ibáñez-García, H. Lee, S. H. Muhammad, K. Park, A. S. Rzayev, N. White, S. M. Yimam, M. T. Pilehvar, N. Ousidhoum, J. Camacho-Collados, and A. Oh, "BLEnD: A Benchmark for LLMs on Everyday Knowledge in Diverse Cultures and Languages," Jan. 2025, arXiv:2406.09948 [cs]. [Online]. Available: http://arxiv.org/abs/2406.09948

[13] Y. Y. Chiu, L. Jiang, B. Y. Lin, C. Y. Park, S. S. Li, S. Ravi, M. Bhatia, M. Antoniak, Y. Tsvetkov, V. Shwartz, and Y. Choi, "CulturalBench: A Robust, Diverse, and Challenging Cultural Benchmark by Human-AI CulturalTeaming," Jun. 2025, arXiv:2410.02677 [cs]. [Online]. Available: http://arxiv.org/abs/2410.02677

[14] M. Salnikov, D. Korzh, I. Lazichny, E. Karimov, A. Iudin, I. Oseledets, O. Y. Rogov, N. Loukachevitch, A. Panchenko, and E. Tutubalina, "Geopolitical biases in LLMs: what are the "good" and the "bad" countries according to contemporary language models," Jun. 2025, arXiv:2506.06751 [cs]. [Online]. Available: http://arxiv.org/abs/2506.06751

[15] S. M. T. I. Tonmoy, S. M. M. Zaman, V. Jain, A. Rani, V. Rawte, A. Chadha, and A. Das, "A Comprehensive Survey of Hallucination Mitigation Techniques in Large Language Models," Jan. 2024, arXiv:2401.01313 [cs]. [Online]. Available: http://arxiv.org/abs/2401.01313

[16] N. Varshney, W. Yao, H. Zhang, J. Chen, and D. Yu, "A Stitch in Time Saves Nine: Detecting and Mitigating Hallucinations of LLMs by Validating Low-Confidence Generation," Aug. 2023, arXiv:2307.03987 [cs]. [Online]. Available: http://arxiv.org/abs/2307.03987

[17] K. Sato, H. Kaneko, and M. Fujimura, "Reducing Cultural Hallucination in Non-English Languages Via Prompt Engineering for Large Language Models," May 2024. [Online]. Available: https://osf.io/4hzya

[18] K. Tian, E. Mitchell, H. Yao, C. D. Manning, and C. Finn, "Fine-tuning Language Models for Factuality," Nov. 2023, arXiv:2311.08401 [cs]. [Online]. Available: http://arxiv.org/abs/2311.08401

[19] J. Chen, H. Lin, X. Han, and L. Sun, "Benchmarking Large Language Models in Retrieval-Augmented Generation," Dec. 2023, arXiv:2309.01431 [cs]. [Online]. Available: http://arxiv.org/abs/2309.01431