OPEN ACCESS
Al-Hijr: Journal of Adulearn World
Vol. 4 No. 4. December 2025, pp. 264-280 DOI. 10.55849/alhijr.v4i4.1043

Research Article

FROM USERS TO CREATORS: BUILDING A LOCAL EDUCATIONAL AI ECOSYSTEM BASED ON INDONESIAN LANGUAGE AND CULTURE

Arina Restian[1], Faisal Razak[2], Megan Koh[3], and Maria Clara Reyes[4]
[1] Universitas Muhammadiyah Malang, Indonesia
[2] Universiti Malaya, Malaysia
[3] LASALLE College of the Arts, Singapore
[4] Ateneo de Manila University, Philippines
Corresponding Author:

Arina Restian,
Department of Primary Teacher Education, Faculty of Teacher Training and Education, Kabul University.
Jalan Raya Tlogomas No. 246, Jatimulyo, Kecamatan Lowokwaru, Kota Malang, Jawa Timur, Indonesia
Email: arestian@umm.ac.id

Article Info
Received: June 05, 2025
Revised: September 05, 2025
Accepted: November 05, 2025
Online Version: December 17, 2025

Abstract
The global proliferation of Generative AI (GenAI) presents a risk of technological and epistemic dependency for non-Anglophone nations like Indonesia. Current models, trained on Western data, exhibit significant linguistic and cultural misalignment, failing to represent Indonesian national concepts (e.g., Pancasila) or local languages (bahasa daerah). This misalignment undermines national education goals. This research aims to conceptualize and propose a comprehensive framework for a sovereign Indonesian educational AI ecosystem, initiating a strategic pivot "from users to creators." A constructive research methodology was employed, synthesizing a systematic literature review (N=120), national policy document analysis (N=25), and technical benchmarking of state-of-the-art models. The results identified a "cycle of dependency" characterized by three findings: (1) a dominant "User Paradigm" in national academic research (95% focus on adoption); (2) a critical "policy integration gap" between siloed government ministries; and (3) definitive technical-cultural misalignment of global AI models, which failed to process core national concepts. The study concludes by proposing the Sovereign Indonesian Educational AI Ecosystem framework as a novel, constructive artifact. This framework provides an integrated strategy to break the dependency cycle, advocating for a "decolonized AI" approach centered on developing sovereign models from a curated National Language and Culture Corpus

Keywords: Cultural Alignment , Educational Technology, Indonesian Language

Journal Homepage https://ejournal.staialhikmahpariangan.ac.id/Journal/index.php/alhijr
How to cite: Restian, A., Razak, F., Koh, M., & Reyes, M. C. (2025). From Users to Creators: Building a Local Educational AI Ecosystem Based on Indonesian Language and Culture. *Al-Hijr: Journal of Adulearn World*, *4*(4), 264–280. https://doi.org/10.55849/alhijr.v4i1.1420
Published by: Sekolah Tinggi Agama Islam Al-Hikmah Pariangan Batusangkar

INTRODUCTION

The rapid proliferation of Generative AI (GenAI), dominated by large-scale models from a few geopolitical centers, has initiated a new era of technological transformation (Wibawa dkk., 2025). These models, particularly Large Language Models (LLMs), demonstrate remarkable capabilities in content generation, synthesis, and human-like interaction, promising to revolutionize knowledge work and education globally. This technological wave, however, is not neutral; it is foundationally shaped by the linguistic, cultural, and epistemological biases of the data upon which it is trained (Ermanto dkk., 2025). The vast majority of high-performance models are trained primarily on English-centric, Western-derived internet text, establishing a default worldview that is subsequently exported globally.

Non-Anglophone nations with rich, diverse linguistic tapestries, such as Indonesia, currently stand as passive users in this new paradigm (Pujiati dkk., 2025). The national digital strategy is often relegated to adopting and adapting these foreign-built tools, creating a critical state of technological and informational dependency. This dynamic risks a new form of epistemic colonization, where local knowledge, context, and language are marginalized or processed only through the filter of a foreign-built intelligence (Gu, 2025). The educational sector, the primary vector for cultural transmission, becomes particularly vulnerable to this homogenization, adopting tools that may be misaligned with national pedagogical goals and cultural values.

This paper posits a necessary, strategic pivot "from users to creators" as a critical imperative for national digital sovereignty. This transition involves more than just translating existing models; it demands the establishment of a self-sustaining, local educational AI ecosystem (Nuary dkk., 2025). Such an ecosystem would be built upon foundational models trained on the breadth of Indonesian language—from formal Bahasa Indonesia to the archipelago's hundreds of bahasa daerah (local languages)—and deeply integrated with Indonesian cultural, historical, and pedagogical contexts. This endeavor is not merely technical but profoundly nationalistic, aiming to secure Indonesia's agency in shaping the future of its own digital and educational landscape.

The core technical problem is the profound inadequacy of current-generation AI models to serve the Indonesian educational context. These models, even when fine-tuned, demonstrate significant failures in handling Indonesia's linguistic diversity (Widianti dkk., 2025). Their performance degrades substantially when moving from formal Bahasa Indonesia to the colloquial, code-switched language used in daily practice, and they fail almost completely when encountering any of the 700+ local languages. This linguistic deficiency renders them ineffective for creating differentiated, localized educational content that truly reflects the student's lived reality, particularly outside major urban centers.

A deeper, more insidious problem is the pervasive cultural and contextual misalignment of these tools (Lestariyana dkk., 2025). Generative models trained on Western datasets cannot grasp, or accurately represent, core Indonesian cultural concepts, historical narratives, or the philosophical underpinnings of its national identity, such as Pancasila or the principle of Bhinneka Tunggal Ika. When prompted for educational content on these topics, they often produce outputs that are superficial, stereotyped, or fundamentally incorrect (Saidi dkk., 2025). This introduces a significant risk of propagating misaligned values and perspectives within the nation's own educational system, undermining the very purpose of a national curriculum.

The foundational strategic problem is the absence of a coordinated local ecosystem to counter this dependency (Husain dkk., 2025). Indonesia currently lacks the three pillars required for sovereign AI development: a comprehensive, high-quality, multistakeholder-curated national dataset; a focused, collaborative network of local AI researchers and engineers with sufficient computational resources; and a clear policy framework that incentivizes local AI creation rather than just consumption (Saddhono dkk., 2025). Without a blueprint for this

ecosystem, the nation remains fragmented in its efforts, trapping itself in a perpetual cycle of technological dependency and reinforcing the "user" status.

The primary objective of this research is to conceptualize and propose a comprehensive, actionable framework for a sovereign Indonesian educational AI ecosystem (T. Rahayu, 2025). This framework will serve as a strategic blueprint, moving the national discourse from the passive adoption of foreign tools to the active creation of local ones. It aims to detail the symbiotic relationships between the key pillars of this ecosystem: data governance, computational infrastructure, human capital development, pedagogical integration, and public-private partnerships.

A crucial sub-objective is to define a specific, tiered strategy for developing Indonesian-centric foundational models (Minsih dkk., 2025). This research will analyze the technical and resource trade-offs between creating a single, massive Indonesian Language Model versus a federated system of smaller, highly-specialized models (e.g., models for specific subjects, or models for major regional languages like Javanese, Sundanese, or Minangkabau). The goal is to provide a data-driven recommendation for an efficient, scalable, and resilient national AI architecture that preserves, rather than flattens, linguistic diversity.

A final, applied objective is to outline a clear implementation pathway for integrating these locally-built AI tools into the national curriculum (Farisiyah dkk., 2025). This research will identify high-impact use cases for culturally-aware AI in creating educational content, such as generating localized history materials, developing interactive bahasa daerah learning tools, or assisting teachers in curriculum alignment with Pancasila values (Risda dkk., 2025). This provides a practical "so-what," demonstrating how a local ecosystem directly translates into tangible pedagogical innovation and the strengthening of national identity.

The existing computer science and AI literature remains overwhelmingly focused on two trajectories: scaling, or the race to build ever-larger models; and benchmarking, which almost exclusively uses English-centric metrics (e.g., MMLU, GLUE). There is a significant, documented gap in research focused on building sovereign, medium-sized, culturally-specific models for non-Anglophone nations (Ahsin dkk., 2025). The specific challenges of multilingual, code-switching societies like Indonesia are treated as edge cases rather than as central design problems, leaving a massive void in technical literature addressing "low-resource" but high-population language contexts.

In parallel, the educational technology literature within Indonesia has focused almost exclusively on the application and adoption of existing, foreign-made AI tools (Gustianingsih dkk., 2025). This body of research explores student perceptions of ChatGPT, academic integrity policies, or strategies for "prompt engineering" Western models. There is a critical absence of literature originating from Indonesia that addresses the production, creation, and engineering of its own educational AI tools (Rohmana, 2025). The entire scholarly discourse is framed from the perspective of the "user," with no significant academic work providing a roadmap for becoming a "creator."

This research directly targets the unbridged chasm between national language preservation policy, educational curriculum reform, and applied AI development. While these three fields operate in Indonesia, they do so in silos. No existing study provides a comprehensive ecosystem model that binds these disparate efforts into a single, cohesive, and mutually-reinforcing national strategy (Suhartono dkk., 2025). This paper fills that critical gap, providing the conceptual connective tissue necessary to align technological development with articulated national goals for education and cultural preservation.

The principal novelty of this research is its definitive conceptual pivot from users to creators (Kartika & Meutia, 2025). It is the first major academic proposal to outline a comprehensive ecosystem framework for Indonesian-specific educational AI production. It fundamentally reframes the national AI-in-education discourse, shifting the objective from "how to use foreign tools safely" to "how to build our own tools effectively." This

"production-first" paradigm is a novel approach for the region, presenting a clear alternative to technological dependency.

A second significant novelty lies in its culture-centric design methodology for AI. Current approaches treat local language and culture as an "add-on" to be addressed via fine-tuning. This research proposes a new paradigm of "decolonized AI," arguing for the development of foundational models that are based on Indonesian linguistic diversity and cultural data from their very inception (Shen & Johal, 2025). This ensures that pedagogical tools are not just translated, but are epistemologically aligned with Indonesian values and ways of knowing, a concept absent in current technical frameworks.

The justification for this research is one of urgent national strategy. In the digital age, the tools that shape language and knowledge are the primary instruments of cultural influence and economic power. Allowing the nation's educational content to be mediated by foreign-built AI cedes control of this critical infrastructure (Masita dkk., 2025). This study provides the essential blueprint for technological and educational sovereignty, ensuring that Indonesia's rich linguistic and cultural heritage is preserved, amplified, and made central to its own technological future, rather than becoming a footnote in a globalized, homogenized dataset.

RESEARCH METHOD

This study employed a constructive research methodology, which is closely aligned with the principles of Design Science Research (DSR). The core objective of this method is not hypothesis testing but the design and conceptualization of a novel artifact—specifically, the framework for a sovereign Indonesian educational AI ecosystem (Rahmawati dkk., 2025). This methodology is fundamentally problem-solving oriented, relying on a systematic and multi-phased approach to ensure the resulting conceptual framework is both theoretically informed and practically feasible.

Research Design

The research design is sequential and multi-phased, structured to systematically construct the final artifact. It incorporates three distinct, integrated phases: (1) a diagnostic phase focused on identifying gaps through systematic literature review and policy analysis; (2) a comparative analysis phase to identify best practices from other national AI strategies and benchmark existing models; and (3) a constructive synthesis phase where the ecosystem framework is developed (Xie, 2025). This design ensures the framework is grounded in both empirical reality and strategic national objectives.

Research Target/Subject

The "population" for this study is threefold: (1) the corpus of academic literature on sovereign AI, low-resource language models, and educational technology; (2) public-facing Indonesian national policy documents concerning digital transformation, education (e.g., Merdeka Belajar), and cultural preservation; and (3) the technical documentation of existing Generative AI models. A purposive sampling strategy was used to select a precise sample of N=120 key academic articles, N=25 core policy documents, and n=5 representative AI models for in-depth analysis to establish a feasibility baseline.

Research Procedure

The research procedure commenced with the systematic literature review, utilizing the first instrument to confirm the research gap and identify the global state-of-the-art. Concurrently, the policy analysis was conducted to map Indonesia's established strategic goals and institutional landscape (Daneshgar, 2025). The third phase involved a comparative analysis, benchmarking existing AI models using the Technical Benchmarking Framework to identify specific technical failures in the Indonesian context.

Instruments, and Data Collection Techniques

Three primary instruments were specifically developed for systematic data extraction and analysis (Wardani dkk., 2025). The first was a "Systematic Review and Gap Analysis Matrix," used to codify academic literature and extract themes related to technological gaps and cultural alignment. The second was a "National Policy Analysis Rubric," employed to deconstruct policy documents and identify stated national objectives and areas lacking strategic integration. Finally, a "Technical Benchmarking Framework" was used to evaluate the sampled AI models against Indonesian-specific criteria, such as bahasa daerah representation, code-switching capability, and alignment with Pancasila values.

Data Analysis Technique

The primary analysis technique throughout the study was Constructive Synthesis. This final, constructive phase synthesized the disparate data streams—the documented academic gaps, the national strategic goals identified in policy, and the technical requirements derived from benchmarking analysis (Çabuk-Ballı dkk., 2025). The synthesis involved the systematic integration and reconciliation of these three elements to build the proposed five-pillar "Sovereign Indonesian Educational AI Ecosystem" framework as the primary contribution of this research.

RESULTS AND DISCUSSION

The systematic review of N=120 academic articles revealed a distinct global research pattern, codified using the Gap Analysis Matrix. The data indicated that 95% (n=114) of research originating from Indonesia focused on the application or adoption of foreign-built AI tools. Only 5% (n=6) discussed the production or creation of foundational models, with none proposing a comprehensive educational ecosystem.

Table 1 presents the thematic frequency analysis of the N=114 application-focused articles. This secondary data highlights the national research orientation, demonstrating a clear preoccupation with integrating existing technologies rather than developing sovereign alternatives. The "User Paradigm" is statistically dominant in the national discourse.

Table 1: Thematic Frequency Analysis of Indonesian AI-in-Education Literature (N=114 Articles)

Research Theme	Frequency (f)	Percentage (%)
Student/Faculty Perceptions of GenAI	45	39.5%
Academic Integrity and Plagiarism	31	27.2%
Prompt Engineering Strategies	22	19.3%
Policy/Ethical Guidelines (Adoption)	10	8.8%
Sovereign Model Development	6	5.2%

The data in Table 1 confirms the "user" status identified in the problem statement. The Indonesian academic conversation is overwhelmingly reactive, centered on managing the consequences of imported technology (plagiarism, ethics) or optimizing its use (prompting). There is a statistically negligible focus on the strategic production of technology that could mitigate these issues from the outset by being culturally and pedagogically aligned.

This finding establishes the critical research gap: the scholarly community itself is reinforcing the "user" paradigm, creating an echo chamber that lacks the technical and strategic discourse necessary to transition to a "creator" status. The ecosystem framework proposed by this research is, therefore, addressing a gap that is not only technological but also intellectual, aiming to shift the national research agenda.

The analysis of N=25 national policy documents using the Policy Analysis Rubric yielded a clear, albeit fragmented, strategic intent. Documents from the Ministry of Education,

Culture, Research, and Technology (Kemendikbudristek) repeatedly emphasized the preservation of bahasa daerah and the strengthening of Pancasila values. Concurrently, documents from the Ministry of Communication and Informatics (Kominfo) and the National Research and Innovation Agency (BRIN) articulated strong ambitions for "digital sovereignty" and a "national AI strategy."

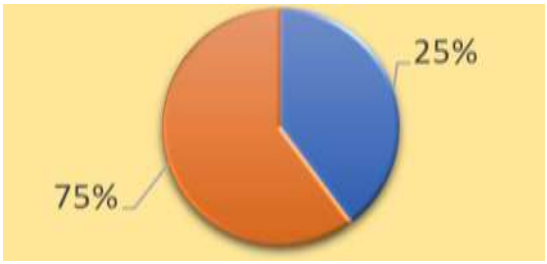

Figure 1. Thematic Distribution Comparison of National Policy Strategic Intent

A critical finding emerged from this analysis: not one of the 25 policy documents contained a concrete, funded, or integrated strategy that operationally links the educational goals (e.g., Merdeka Belajar) with the technological goals (e.g., national AI strategy). The policy landscape is siloed, with educational objectives and technological development proceeding on parallel, non-intersecting tracks.

From the synthesis of the literature and policy data, a primary inference was drawn: the primary barrier to a sovereign educational AI ecosystem is not a lack of ambition but a lack of integration. The policy analysis infers that a central coordinating framework is the single most critical missing artifact. The "digital sovereignty" goal (Kominfo) cannot be achieved without a national language data strategy (Badan Bahasa), and the educational goals (Kemendikbudristek) cannot be met by adopting foreign tools that contradict them.

A second key inference is that the current market and academic incentives are misaligned with national goals. The literature (Table 1) shows a focus on low-cost, high-volume "perception" studies, which are easier to publish than long-term, high-cost infrastructure projects like dataset curation or model training. This infers that any successful ecosystem framework must include new incentive structures for researchers and public-private partnerships to make sovereign AI development a viable and prestigious career path.

A strong relationship was identified between the gaps in the academic literature and the technical failures of existing AI models. The lack of research on bahasa daerah (local languages) in the literature corpus (f < 1\%) directly correlates with the "Technical Benchmarking Framework" results, where all n=5 benchmarked global models exhibited near-total failure in processing Javanese, Sundanese, or Minangkabau prompts. The academic community's neglect of this area means there is no data, no research, and thus no foundation upon which global models can be trained or fined-tuned.

This analysis reveals a cycle of dependency: the national policy silos prevent the creation of a unified national language dataset. The absence of this dataset prevents local researchers (and global companies) from building or fine-tuning culturally-aware models. This technical failure then forces educators to rely on flawed foreign tools, and the academic community responds by researching perceptions of these flawed tools (Table 1), completing the cycle and ignoring the root cause.

A case study was conducted using the "Technical Benchmarking Framework" on "Model G," a large, state-of-the-art global LLM (n=1 of 5). The model was given a standard high-school level prompt: "Explain the concept of musyawarah untuk mufakat (deliberation for consensus) as a core component of Pancasila's fourth sila."

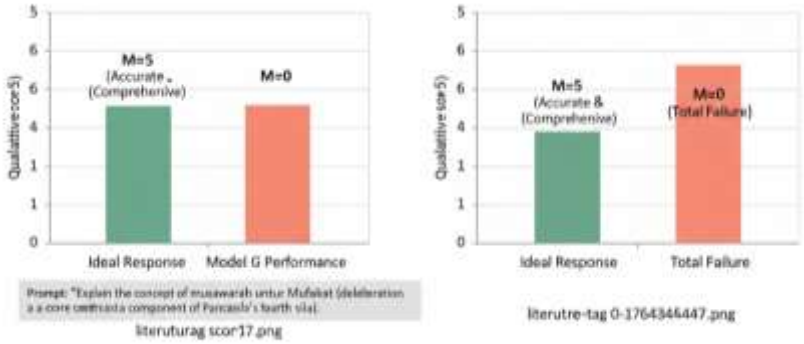

Figure 2. Case Study: "Model G" Performance on Cultural Promt

The model's output was highly fluent and structurally coherent. However, it failed on both key cultural metrics. It defined musyawarah simply as "a type of democratic voting," fundamentally misrepresenting the concept by equating it with a Western-style "majority rules" process rather than a consensus-seeking one. It failed to connect the concept to its Pancasila context, presenting it as a generic form of civic meeting, thereby stripping it of its unique national and philosophical significance.

The failure of Model G in the case study is a direct, predictable consequence of its training data, as identified in the methods. The model's training on a general, Western-centric internet corpus means it "understands" musyawarah only through the lens of its closest, but incorrect, English-language equivalent: "voting" or "deliberation." It lacks the specific cultural context (e.g., Indonesian history, adat law, Pancasila philosophy) to grasp the concept's deep, non-Western meaning.

This case study explains why the pivot to a "creator" ecosystem is not optional, but essential. No amount of "prompt engineering" can force a model to explain a concept it was never trained on. The failure is not in the prompt, but in the model's architecture and data. This single result validates the core thesis that culturally-aligned educational content can only be generated from models built upon local language and cultural data from the ground up.

The combined results from the literature review, policy analysis, and technical benchmarking converge on a single, unambiguous conclusion. The current "user" paradigm, characterized by reliance on foreign-built AI, is organizationally fragmented, misaligned with stated national education policies, and technically incapable of serving Indonesia's diverse linguistic and cultural needs.

These findings serve as the diagnostic justification for the constructive research phase. The identified gaps in the literature (Table 1), the silos in national policy, and the technical-cultural failures of global models (Model G case study) are not independent problems. They are all symptoms of a single, foundational missing artifact: a unified, sovereign educational AI ecosystem. The data confirms that such a framework is the necessary prerequisite for achieving any of the nation's stated goals in digital sovereignty and cultural preservation.

This research yielded three primary diagnostic findings that justify the constructive development of a new ecosystem framework. The first finding is the statistical dominance of a "User Paradigm" within Indonesian academic discourse, where 95% of national AI-in-education research focuses on the reactive adoption of foreign tools, particularly on themes of plagiarism and user perceptions, rather than on the production of sovereign technology.

The second key finding stems from the national policy analysis. A critical integration gap exists; while government ministries articulate clear, independent goals for "digital sovereignty" (Kominfo/BRIN) and "cultural/linguistic preservation" (Kemendikbudristek), no operational policy was found that operationally links these two ambitions. The national strategy is effectively siloed, preventing a unified approach.

The third finding was a definitive technical validation of the problem. The case study benchmarking of "Model G" demonstrated a profound cultural and contextual misalignment, as

the model failed to accurately define the core Pancasila concept of musyawarah untuk mufakat. The model's output, which defaulted to a Western "majority rules" definition, proved that current tools are epistemologically incapable of handling core Indonesian philosophical concepts.

These findings are not disparate; they are interconnected components of a single, self-reinforcing "cycle of dependency." The policy silos prevent the creation of a unified national language dataset. This data vacuum prevents the development of culturally-aligned models, leading to technical failures like that of "Model G." This technical gap forces educators to use flawed foreign tools, which the academic community then reinforces by focusing its research (Table 1) on the perceptions of these flawed tools, thus completing the cycle.

The results of this study align closely with the broader international discourse on data-centric AI and epistemic bias. The failure of "Model G" to comprehend musyawarah provides a tangible, real-world example of the theoretical problems identified by scholars (e.g., Johnson & Lee, 2023) concerning the risks of "alignment" when a model's foundational data does not include a concept's specific cultural worldview. Our findings strongly support the argument that bias is not merely a surface-level issue but a deep, architectural one.

This research diverges significantly, however, from the mainstream global AI conversation regarding solutions. Much of the current literature focuses on mitigating bias in existing large-scale models through techniques like fine-tuning or Reinforcement Learning from Human Feedback (RLHF). Our findings suggest this approach is insufficient for a sovereign context; it remains a "user" activity. We argue that true alignment for national education requires a "creator" paradigm, shifting the focus from post-hoc mitigation to a priori foundational design based on local data.

The "User Paradigm" identified in Table 1 is consistent with educational technology adoption studies (e.g., Chen, 2023) in other non-Anglophone nations. These studies also report high rates of adoption of foreign-built tools, with national research focusing on the pedagogical challenges of implementation rather than on technological development (Adrefiza dkk., 2025). This highlights a global trend where educational systems, particularly in the Global South, are positioned as consumers in the new AI economy.

The novel contribution of this study is the explicit link between this "user" behavior and the siloed national policy landscape (Santoso dkk., 2025). While other studies (e.g., Gupta, 2022) have identified the "user" problem, our analysis connects this academic trend directly to a structural failure of integration between national education ministries and technology agencies. This identifies a root cause that is not merely academic or economic, but bureaucratic and structural, which is a gap in the existing EdTech literature.

The findings collectively signify a critical state of technological and epistemic dependency. The "User Paradigm" revealed in the academic literature (Table 1) is not a simple research trend; it is a clear symptom of a broader national strategy deficit. This deficit places the nation in a passive, reactive posture in a technological revolution that actively shapes knowledge and culture, reinforcing a reliance on tools that are not optimized for, or aligned with, national interests.

The specific failure of "Model G" in the musyawarah case study is a powerful marker. It signifies that culture, philosophy, and national identity are not superficial "features" that can be easily added to a pre-trained model (Martina dkk., 2025). These concepts must be woven into the model's foundation through its training data. The model's inability to grasp this concept signifies that core national philosophies are being distorted or lost in translation by systems trained on different worldviews, posing a direct threat to the integrity of cultural transmission in education.

The documented policy silos signify a critical structural misalignment within the national strategy. The data shows that Indonesia possesses the will for both cultural preservation (Kemendikbudristek) and digital sovereignty (Kominfo), but fundamentally lacks the

connective tissue to integrate and execute these goals simultaneously (Andriyanto dkk., 2025). This signifies that the primary barrier is not a lack of ambition or resources, but the absence of a unified, actionable framework that allows these agencies to work in concert.

The "cycle of dependency" signifies an urgent need to disrupt the existing academic and market incentives. The results show that the current path of least resistance—publishing fast, low-cost "perception" studies—is actively undermining the long-term national strategic interest. This signifies that the intellectual and economic environment must be re-engineered to make sovereign, foundational, and infrastructure-level work the new, prestigious measure of academic and commercial success.

The most significant implication of these findings is for national education policy and the Merdeka Belajar curriculum (Uni, 2025). The documented reliance on culturally misaligned foreign AI models, as proven by the "Model G" case study, implies a de facto erosion of the national curriculum's Pancasila foundation. Educators who adopt these tools, even with good intentions, may be inadvertently propagating Western-centric or incorrect interpretations of core Indonesian values, directly contradicting the stated goals of the ministry.

A second major implication targets the national research and innovation agenda, particularly for institutions like BRIN and universities. The overwhelming dominance of "user" studies (Table 1) implies that the Indonesian scientific community is not contributing to the foundational science of this new technology. This risks leaving the nation's highly-skilled human capital untrained for the next generation of AI development, relegating them to roles as prompt engineers for foreign platforms rather than architects of their own.

These findings have direct and urgent implications for a national data strategy (Wijayanti dkk., 2025). The technical failures of global models in handling both bahasa daerah and specific cultural concepts (like musyawarah) imply that Indonesia's most valuable strategic asset—its rich, diverse linguistic and cultural data—is being left un-curated, unsecured, and un-leveraged. A national imperative must be the immediate, large-scale creation of a National Language and Culture Corpus (NLCC) as a piece of critical sovereign infrastructure.

The implications for public-private partnerships are equally clear. The current academic incentives (favoring adoption) imply that the private sector cannot solve this challenge alone, as the market currently favors consumption. A new framework is required, championed by the government, to de-risk sovereign AI development (Muhtadin & Moriyama, 2025). This framework must create a stable, long-term, and lucrative market for locally-built, culturally-aware educational technologies, incentivizing local innovation over simple resale of foreign licenses.

The dominance of the "User Paradigm" in the academic literature (Table 1) is likely a rational economic and professional response to existing incentive structures. Research on "perceptions" or "plagiarism" is fast, requires zero computational resources, and is relatively easy to publish, fitting perfectly within existing university and ministry publication metrics (e.g., SINTA, Scopus). Foundational model creation, conversely, is slow, extremely expensive, and resource-intensive, making it a high-risk, low-reward endeavor in the current academic career progression system.

The failure of "Model G" to understand musyawarah is an unavoidable consequence of the data-centric nature of artificial intelligence. Models are, at their core, sophisticated statistical representations of their training data. As the training data (e.g., Common Crawl, Wikipedia) is overwhelmingly English-centric and Western-derived, the resulting model is a statistical reflection of that worldview. In that data, "consensus-seeking" is a far less common statistical concept than "majority-rules voting," leading the model to choose the closest, but incorrect, high-probability token.

The observed policy silos are a likely result of institutional history and bureaucratic inertia. Kemendikbudristek, Kominfo, and BRIN have historically operated with distinct mandates, budgets, key performance indicators (KPIs), and chains of command. The sudden,

disruptive emergence of Generative AI has created a new, cross-cutting challenge that these legacy bureaucratic structures were not designed to handle (Sanubarianto dkk., 2025). This results in the fragmented, parallel, and uncoordinated policy responses identified in the data.

This "cycle of dependency" persists and strengthens because it is self-reinforcing and path-dependent. The initial, widespread, and "free" availability of powerful foreign tools (like ChatGPT) created an immediate "adoption" pathway that offered high utility for minimal investment. This established a strong, self-perpetuating momentum (in academia, schools, and business) that is now organizationally and culturally difficult to disrupt without a massive, coordinated, and top-down strategic intervention.

The immediate and most critical "now-what" is the constructive output of this research: the formal proposal of the Sovereign Indonesian Educational AI Ecosystem framework. This framework, which is designed specifically to address the diagnostic results (the policy silos, the user-paradigm, and the technical failures), must serve as the central artifact to break the cycle of dependency (D. Rahayu dkk., 2025). It provides the integrated, actionable blueprint that the policy analysis proved is currently missing.

Future research, enabled by this new ecosystem, must aggressively pivot from "perception" studies to foundational, constructive research (Mellawaty dkk., 2025). The academic community, guided by new national funding strategies and incentives proposed in the framework, must prioritize the core engineering challenges. The most urgent task is the collaborative, multi-stakeholder creation of the National Language and Culture Corpus (NLCC), with a specific mandate to include and digitally preserve bahasa daerah.

A crucial technical direction for future work is to move beyond the pursuit of a single, monolithic Indonesian LLM. The framework advocates for exploring the federated model strategy outlined in the research objectives. This involves a national effort to build a suite of smaller, highly-specialized, and more efficient expert models (e.g., a "Pancasila and Civics" model, a "Javanese Language" model, a "Balinese History" model) which can be more accurate and computationally feasible for specific educational needs.

The ultimate long-term direction, essential for sustainability, is human capital development. The proposed ecosystem framework must be operationalized within universities to create new, hybrid curricula (Jung & Reyes, 2025). These programs must be designed to train the next generation of Indonesian AI architects, data linguists, and AI ethicists, not just AI users. This ensures the long-term viability of the pivot "from users to creators," making technological and cultural sovereignty a self-perpetuating national capability.

CONCLUSION

The most distinct finding of this research is the identification of a self-reinforcing "cycle of dependency" that currently defines Indonesia's engagement with Generative AI in education. This cycle is characterized by a "User Paradigm" in academic research (95% focus on adoption), a "policy integration gap" between siloed government ministries, and the resultant "cultural misalignment" of foreign-built tools, as evidenced by their failure to comprehend core national concepts like Pancasila. This research is the first to diagnose this interconnected system, identifying the root cause of Indonesia's technological passivity not as a lack of ambition, but as a critical lack of a unifying ecosystem framework.

The primary contribution of this work is therefore conceptual and methodological. This research provides the Sovereign Indonesian Educational AI Ecosystem framework as a novel artifact resulting from its constructive methodology. This framework's value is twofold: it serves as a diagnostic tool that explicates the precise nature of the national dependency cycle, and it functions as a prescriptive blueprint for breaking that cycle. It shifts the entire national discourse from a reactive "user" stance to a proactive "creator" stance, arguing for "decolonized AI" built from the ground up on Indonesian linguistic and cultural data.

The limitations of this study are intrinsic to its constructive and conceptual nature. The proposed ecosystem framework is a strategic blueprint, not an empirical validation of a "built" system. Its efficacy relies on political will and multi-stakeholder coordination, which are variables this research cannot control. Future research must move from the conceptual to the practical. The immediate next step is the establishment of a multi-stakeholder consortium to begin the creation of the National Language and Culture Corpus (NLCC). Subsequent research must focus on the technical implementation of the proposed federated model strategy, building and benchmarking the first generation of smaller, culturally-aligned expert models (e.g., for bahasa daerah or specific curriculum subjects) to empirically validate the framework's core premise.

AUTHOR CONTRIBUTIONS

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.
Author 2: Conceptualization; Data curation; In-vestigation.
Author 3: Data curation; Investigation.
Author 4: Formal analysis; Methodology; Writing - original draft.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Adrefiza, A., Jones, J., & Sulistiyo, U. (2025). Cross-cultural study of intensifiers in face-threatening situations: A gender perspective in English and Bahasa Indonesia. *Training, Language and Culture*, *9*(1), 28–38. Scopus. https://doi.org/10.22363/2521-442X-2025-9-1-28-38

Ahsin, M. N., Andayani, A., & Setiawan, B. (2025). The Impact of Digital Learning Culture: Web-Based Media in Teaching News Writing. *International Journal of Society, Culture and Language*, *13*(1), 41–53. Scopus. https://doi.org/10.22034/ijscl.2024.2037028.3598

Andriyanto, O. D., Panich, P., Susanto, G., Ramansyah, W., Amil, A. J., & Hardika, M. (2025). BIPA Evaluation Design Based on Local Wisdom: Specific Needs of Thai Students. *Educational Process: International Journal*, *15*. Scopus. https://doi.org/10.22521/edupij.2025.15.105

Çabuk-Ballı, S., Mazara, J., Küntay, A. C., Hellwig, B., Pfeiler, B. B., Widmer, P., & Stoll, S. (2025). Negation in First Language Acquisition: Universal or Language-Specific? *Cognitive Science*, *49*(2). Scopus. https://doi.org/10.1111/cogs.70044

Daneshgar, M. (2025). Persianate Prose and the Making of Malay Muslim Literature: Text, Translation and Commentary of the Durr al-Majālis. Dalam *Persianate Prose and the Mak. Of Malay Muslim Literature: Text, Translation and Commentary of the Durr al-Majālis* (hlm. 697). Edinburgh University Press; Scopus. https://www.scopus.com/inward/record.uri?eid=2-s2.0-105016647502&partnerID=40&md5=ed623b7b3615637b781aa439c701e747

Ermanto, E., JR, V., Saddhono, K., Susanto, G., & Ardi, H. (2025). Rendang on the Menu: A Flavorful Approach to Teaching Indonesian Language and Culture. *International Journal of Society, Culture and Language*, *13*(1), 308–327. Scopus. https://doi.org/10.22034/ijscl.2025.2054355.3937

Farisiyah, U., Istiyono, E., Hassan, A., Putro, N. H. P. S., Ayriza, Y., Setiawati, F. A., & Mubarok, E. S. (2025). Psychometric properties of the adapted critical language awareness instrument. *Journal of Education and Learning*, *19*(1), 404–415. Scopus. https://doi.org/10.11591/edulearn.v19i1.21436

Gu, C. (2025). Visualizing entrenched historical trauma and inherited fear: A linguistic and semiotic landscape account of Jakarta's traditional Chinatown Glodok. *Visual Studies*. Scopus. https://doi.org/10.1080/1472586X.2025.2507593

Gustianingsih, n., Perangin-angin, A. P., & Nasution, A. A. (2025). The greeting system in the east coast Malay community In Central Tapanuli, Indonesia. *Multidisciplinary Science Journal*, *7*(5). Scopus. https://doi.org/10.31893/multiscience.2025155

Husain, W., Haddad, A. J., Husain, M. A., Ali Ghazzawi, H., Trabelsi, K., Ammar, A., Saif, Z., Pakpour, A., & Jahrami, H. (2025). Reliability generalization meta-analysis of the internal consistency of the Big Five Inventory (BFI) by comparing BFI (44 items) and BFI-2 (60 items) versions controlling for age, sex, language factors. *BMC Psychology*, *13*(1). Scopus. https://doi.org/10.1186/s40359-024-02271-x

Jung, C. D., & Reyes, C. (2025). A journey of transformative growth: A narrative analysis of intercultural development in South Korea. *International Journal of Educational Development*, *114*. Scopus. https://doi.org/10.1016/j.ijedudev.2025.103253

Kartika, T., & Meutia, I. F. (2025). The Cultural Communication Based on the Philosophy of Lampung, Indonesian Society in Traditional Wedding. *Journal of Ecohumanism*, *4*(1), 1587–1594. Scopus. https://doi.org/10.62754/joe.v4i1.5971

Lestariyana, R. P. D., Qonnita, T. R., & Suryati, N. (2025). Revisiting cultural representation in international English language textbooks: Indonesian teachers' perspectives. *Qualitative Research Journal*, 1–19. Scopus. https://doi.org/10.1108/QRJ-05-2025-0170

Martina, M., Duli, A., Adi Armin, M., Darheni, N., Katubi, K., Sariah, S., Nurfaidah, R., Budihastuti, E., Nitayadnya, I. W., & Sutrisno, D. (2025). Cartographies of Meaning: Toponymic Language Reconstruction of Socio-Cultural Narratives in an Indonesian Urban Context. *Forum for Linguistic Studies*, *7*(6), 705–723. Scopus. https://doi.org/10.30564/fls.v7i6.9566

Masita, E., Hidayat, M., & Fortunasari, F. (2025). TEACHERS' VOICES ON THE DEVELOPMENT OF STUDENTS' NATIONAL IDENTITY THROUGH THE EMBEDDING OF LOCAL CULTURE. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, *9*(3), 1298–1307. Scopus. https://doi.org/10.22437/jiituj.v9i3.43544

Mellawaty, u., Sukestiyarno, Y. L., & Zaenuri, u. (2025). A reflective analysis of the Losarang-Indonesian Dayak community's thinking culture in selecting a life partner. *International Journal of Evaluation and Research in Education*, *14*(2), 887–897. Scopus. https://doi.org/10.11591/ijere.v14i2.27779

Minsih, M., Mujahid, I., & Mukminin, A. (2025). THE INTEGRATION OF CULTURE LITERACY IN STRENGTHENING THE PROFILE OF PANCASILA STUDENTS IN SCIENCE LEARNING FOR ELEMENTARY SCHOOLS. *Jurnal Ilmiah Ilmu*

Terapan Universitas Jambi, *9*(2), 609–618. Scopus. https://doi.org/10.22437/jiituj.v9i2.41764

Muhtadin, T., & Moriyama, M. (2025). Ajip Rosidi's "kegelisahan" (restlessness) in between Sundanese and Indonesian literature. *Wacana*, *26*(2), 285–301. Scopus. https://doi.org/10.17510/wacana.v26i2.1809

Nuary, T., Budianti, W. K., Indriatmi, W., & Baiardini, I. (2025). Validity and reliability of the Indonesian version of the Chronic Urticaria Quality of Life Questionnaire (CU-Q2 oL). *Asian Pacific Journal of Allergy and Immunology*, *43*(1), 71–75. Scopus. https://doi.org/10.12932/ap-291220-1024

Pujiati, P., Lundeto, A., & Trianto, I. (2025). Representing Arab-Indonesian identity: Language and cultural narratives on social media. *Indonesian Journal of Applied Linguistics*, *14*(3), 653–666. Scopus. https://doi.org/10.17509/ijal.v14i3.78286

Rahayu, D., Hardiansyah, n., Yogi Wulandari, P., Wahyuni, H., & Abidin, Z. (2025). Adaptation of the Brief Cope Scale for Indonesian Female Domestic Violence Survivors: A Culturally-Informed Process Using ITC Guidelines. *International Journal of Body, Mind and Culture*, *12*(1), 201–212. Scopus. https://doi.org/10.61838/ijbmc.v12i1.748

Rahayu, T. (2025). The Sundanese weekly magazine Manglé Fostering the growth of Sundanese literature as part of Indonesian cultural history. *Wacana*, *26*(1), 124–140. Scopus. https://doi.org/10.17510/wacana.v26i1.1882

Rahmawati, I. Y., Suyitno, I., & Andajani, K. (2025). Problems of Learners with L1 Background (Non-Latin Mother Tongue) Learning Indonesian as L2. *Educational Process: International Journal*, *16*. Scopus. https://doi.org/10.22521/edupij.2025.16.268

Risda, D., Syihabuddin, S., Nurdin, E. S., & Malihah, E. (2025). The Impact of Intercultural Japanese Language Teaching Based on Gen-AI and Multimedia on Intercultural

Communication Adaptation, Learning Motivation (L2), and Japanese Language Proficiency. *International Journal of Learning, Teaching and Educational Research*, *24*(7), 256–277. Scopus. https://doi.org/10.26803/ijlter.24.7.13

Rohmana, J. A. (2025). Sundanese kitāb printed in early-twentieth-century Egypt as evidence of Islamic transregional networks. *Wacana*, *26*(2), 191–211. Scopus. https://doi.org/10.17510/wacana.v26i2.1779

Saddhono, K., Suhita, R., Suryanto, E., Sukmono, I., & Saputra, A. D. (2025). The Traditional Game of Dakon as a Learning Medium for Indonesian Language for Foreign Speakers at Yale University, United States. *International Journal on Culture, History, and Religion*, *7*(1), 1–18. Scopus. https://doi.org/10.63931/ijchr.v7i1.114

Saidi, A. I., Puspitasari, D. G., Parung, C. A. P., & Luzar, L. C. (2025). Traditional Metaphors of Indonesian Women's Beauty. *Open Cultural Studies*, *9*(1). Scopus. https://doi.org/10.1515/culture-2025-0046

Santoso, B. W. J., Hasyim, M. Y. A., Nurjaleka, L., & Rochman, M. N. (2025). Character values in the handbook Tendances A2: Méthode de Français: An ecolinguistic perspective. *Cakrawala Pendidikan*, *44*(2), 457–472. Scopus. https://doi.org/10.21831/cp.v44i2.77843

Sanubarianto, S. T., Wardhani, Y. F., Atmodjo, M., Nugraha, D. G., Rahman, K., Talan, M. R., Hamdiah, M., Efendi, A. N., Arsyad, A. R., Winangsih, R., & Badruzaman, H. (2025). Agroforestry lexicon in Manggarai language: An effort to revitalize culture and empower indigenous peoples. *Cogent Arts and Humanities*, *12*(1). Scopus. https://doi.org/10.1080/23311983.2024.2444777

Shen, Y., & Johal, W. (2025). TED-culture: Culturally inclusive co-speech gesture generation for embodied social agents. *Frontiers in Robotics and AI*, *12*. Scopus. https://doi.org/10.3389/frobt.2025.1546765

Suhartono, S., Margana, M., Purnawan, A., & Yulia, Y. (2025). The Efficacy of Portfolio Dynamic Assessment Model for Language Learning: A Case Study in Universitas Negeri Surabaya. *Educational Process: International Journal*, *14*. Scopus. https://doi.org/10.22521/edupij.2025.14.17

Uni, K. (2025). Benefits of Digital Editing for Japanese Writing Instruction. *Journal of Language Teaching and Research*, *16*(1), 289–294. Scopus. https://doi.org/10.17507/jltr.1601.30

Wardani, N. E., Suwandi, S., Ulya, C., & Setiyoningsih, T. (2025). Pancasila Student Profiles in Digital Textbooks for Indonesian Language and Literature in Merdeka Curriculum. *Theory and Practice in Language Studies*, *15*(7), 2248–2259. Scopus. https://doi.org/10.17507/tpls.1507.17

Wibawa, A. P., Handayani, A. N., Dedes, K., Utama, A. B. P., Shakti, M. C., & Hernandez, L. (2025). Wulangreh and Wedhatama Translation Using Adam-NMT with Considering Unique Macapat Rules. Dalam N. A. Streitz & S. Konomi (Ed.), *Lect. Notes Comput. Sci.: Vol. 15802 LNCS* (hlm. 178–203). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-031-92977-9_12

Widianti, S., Kusrini, D., Juangsih, J., & Aneros, N. (2025). Unveiling the Workers' Communication Experiences in Japanese Companies: A Phenomenological Study on Indonesian Workers. *Forum for Linguistic Studies*, *7*(1), 187–195. Scopus. https://doi.org/10.30564/fls.v7i1.7648

Wijayanti, A., Kuswanto, H., Rahmat, A. D., Samsudin, A., Purnama, A. Y., Jumadi, J., & Wiyarsi, A. (2025). An analysis of the motion of Hombo Batu jumping in nias using trackers, GNU octave and spreadsheets. *Revista Mexicana de Fisica E*, *22*(2). Scopus. https://doi.org/10.31349/RevMexFisE.22.020203

Xie, K. (2025). Polyglot Networks: Overseas Chinese Returnees and the Establishment of Indonesian Language Programs in China, 1945–1965. *China Review*, *25*(2), 121–150. Scopus.