

# DEMOCRATIZING CLIMATE INTELLIGENCE THROUGH LOCALIZED LARGE LANGUAGE MODELS FOR EDUCATION AND GOVERNANCE

Zulfikar Aji Kusworo<sup>1\*</sup>  
Widyana Verawaty Siregar<sup>2</sup>  
Baharuddin Ismail<sup>3</sup>  
Defry Hamdhana<sup>4</sup>

<sup>1</sup>Program Studi Teknologi Pertambangan, Politeknik Energi dan Pertambangan Bandung  
Gedung PEP Bandung, Jl. Jend. Sudirman No.623, Kota Bandung, Jawa Barat 40211, INDONESIA

<sup>2</sup>Program Studi Manajemen, Universitas Malikussaleh  
Kampus Bukit Indah, Kalimantan, Blang Pulo, Lhokseumawe 24355, INDONESIA

<sup>3</sup>Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis  
Pauh Putra Campus 02600 Arau, Perlis, MALAYSIA

<sup>4</sup>Program Studi Teknik Informatika, Universitas Malikussaleh  
Kampus Bukit Indah, Kalimantan, Blang Pulo, Lhokseumawe 24355, INDONESIA

## Abstract

*Climate change presents complex challenges in Indonesia, where local governments and communities frequently experience information asymmetry and limited access to expert knowledge, particularly in low-resource and low-connectivity regions. This study aims to develop and evaluate a localized, domain-adapted Large Language Model (LLM) that functions as offline-capable climate knowledge infrastructure for education and local governance in Indonesia. The research method employs a design science research methodology comprising four stages: (1) selection of Qwen3-4B as the base model, (2) curation of an Indonesian climate and energy transition corpus containing approximately 12,400 instruction-response pairs (~38 MB) drawn from national climate policy documents, NDC/RPJMN frameworks, renewable energy guidelines, and educational climate science texts, (3) parameter-efficient fine-tuning using QLoRA with LoRA rank  $r=16$ ,  $\alpha=32$ , learning rate  $2e-4$ , 3 epochs, per-device batch size 2 with gradient accumulation 4, and 4-bit NF4 quantization, and (4) offline deployment on consumer-grade hardware with task-oriented evaluation against three baseline models (Qwen3-4B-Thinking, Gemma-3-4B, LLaMa-3.1-8B). The results show that the fine-tuned model (Qwen3-4B-REnewbie v1) achieved a 15.4% perplexity reduction on domain-specific test data and an average qualitative score of 9.3/10 across factual accuracy, reasoning structure, and Bahasa Indonesia language compliance, outperforming all baselines (score range 7.0–8.2). The system operates fully offline on consumer-grade hardware with acceptable inference latency. The conclusion drawn from this study is that localized, resource-efficient LLMs can function as practical climate knowledge infrastructure for vocational education and local governance in Indonesia, aligning with Green AI principles and supporting the democratization of climate intelligence in low-connectivity settings.*

## Keywords:

*Large language model; climate change; education; governance*

## Abstrak

Perubahan iklim menghadirkan tantangan kompleks di Indonesia, khususnya bagi pemerintah daerah dan komunitas lokal yang sering mengalami asimetri informasi dan keterbatasan akses terhadap pengetahuan pakar di wilayah dengan sumber daya dan konektivitas terbatas. Penelitian ini bertujuan mengembangkan dan mengevaluasi *Large Language Model (LLM)* yang dilokalkan dan diadaptasi ke domain iklim sebagai infrastruktur pengetahuan iklim berbasis *offline* untuk pendidikan dan tata kelola lokal di Indonesia. Metode penelitian ini menggunakan pendekatan *design science research* yang meliputi (1) pemilihan Qwen3-4B sebagai *base model*, (2) kurasi korpus iklim dan transisi energi Indonesia berisi sekitar 12.400 pasangan instruksi-respons (~38 MB) dari dokumen kebijakan iklim nasional, kerangka NDC/RPJMN, panduan energi terbarukan, serta teks ilmiah iklim, (3) *parameter-efficient fine-tuning* berbasis QLoRA (LoRA rank  $r=16$ ,  $\alpha=32$ , *learning rate*  $2e-4$ , 3 epoch, *batch size* 2 per perangkat dengan *gradient accumulation* 4, dan

**Citation in APA Style:** Kusworo, Z. A., Siregar, W. V., Ismail, B., Hamdhana, D. (2026). Democratizing climate intelligence through localized Large Language Models for education and governance. *VOCATECH: Vocational Education and Technology Journal*, 8(1), 202-215.

kuantisasi 4-bit NF4), dan (4) deployment *offline* pada perangkat keras kelas konsumen dengan evaluasi berorientasi tugas terhadap tiga *baseline* (Qwen3-4B-Thinking, Gemma-3-4B, LLaMa-3.1-8B). Hasil penelitian ini menunjukkan model hasil *fine-tuning* (Qwen3-4B-REnewbie v1) menghasilkan penurunan *perplexity* sebesar 15,4% pada data uji domain dan skor kualitatif rata-rata 9,3/10 pada dimensi akurasi faktual, struktur penalaran, dan kepatuhan Bahasa Indonesia, mengungguli seluruh *baseline* (kisaran 7,0–8,2), serta beroperasi sepenuhnya secara *offline* pada perangkat konsumen dengan latensi inferensi yang dapat diterima. Kesimpulan yang diperoleh dari penelitian ini adalah LLM yang dilokalkan dan hemat sumber daya dapat berfungsi sebagai infrastruktur pengetahuan iklim yang praktis bagi pendidikan vokasi dan tata kelola lokal di Indonesia, selaras dengan prinsip *Green AI* dan mendukung demokratisasi kecerdasan iklim di wilayah konektivitas terbatas.

#### Kata Kunci:

Model bahasa besar; perubahan iklim; pendidikan; tata kelola.

DOI: [10.38038/vocatech.v8i1.325](https://doi.org/10.38038/vocatech.v8i1.325)

Received: 04 April 2026; Accepted: 27 April 2026; Published: 30 April 2026

#### \*Corresponding author:

Nama lengkap, Program Studi Teknologi Pertambangan, Politeknik Energi dan Pertambangan Bandung, Gedung PEP Bandung, Jl. Jend. Sudirman No.623, Kota Bandung, Jawa Barat 40211, INDONESIA  
Email: [zakusworo@gmail.com](mailto:zakusworo@gmail.com)

## 1. INTRODUCTION

Indonesia is among the most climate-vulnerable countries in the world due to its archipelagic geography, tropical climate regime, and high dependence on climate-sensitive ecosystems. Over the past decade, extreme hydrometeorological events such as floods, landslides, and severe storms have increased in both frequency and intensity, posing important risks to human safety, infrastructure, and regional development. These risks are particularly acute in Sumatra, where complex interactions between climate variability, land-use change, and governance capacity shape disaster outcomes (Masson-Delmotte et al., 2023). In late November 2025, a rare tropical system known as Tropical Cyclone Senyar formed unusually close to the equator over the Malacca Strait and made landfall on Sumatra Island. The cyclone generated exceptionally intense and prolonged rainfall, triggering catastrophic floods and landslides across Aceh, North Sumatra, and West Sumatra. The disaster resulted in widespread displacement, severe damage to settlements and critical infrastructure, and substantial loss of life (International Federation of Red Cross & IFRC, 2025; NASA Earth Observatory, 2025; Zahro, 2025).

Bias in language-model outputs — including cultural, linguistic, and data-representation bias — is a critical concern when AI is deployed in public education and governance, because global pre-trained models tend to over-represent English-language and high-resource-country knowledge while under-representing local Indonesian contexts (Bender et al., 2021; Joshi et al., 2020). Addressing this bias through domain- and language-specific localization is therefore essential to ensure fair, contextually relevant, and trustworthy climate communication in Indonesia. Building on these concerns, this study is guided by three research questions: (RQ1) To what extent can a localized, domain-adapted Large Language Model in Bahasa Indonesia reduce predictive uncertainty (perplexity) on Indonesian climate and energy-transition content compared to its general-purpose base model? (RQ2) How does the localized model compare with open-weight baselines (Qwen3-4B-Thinking, Gemma-3-4B, LLaMa-3.1-8B) on factual accuracy, reasoning structure, and language compliance in climate education and governance tasks? (RQ3) Can the fine-tuned model be deployed as a fully offline, resource-efficient system on consumer-grade hardware while maintaining acceptable response quality and latency? Accordingly, this research has three objectives: (1) to develop a fine-tuned Indonesian climate LLM based on Qwen3-4B using QLoRA-based parameter-efficient fine-tuning; (2) to evaluate its domain adaptation performance against the base model and open-weight baselines through task-oriented benchmarks; and (3) to validate its offline deployability on consumer-grade hardware as a practical climate knowledge infrastructure for vocational education and local governance in Indonesia.

## 2. LITERATURE REVIEW

### 2.1. Climate Change and Intensification of Extreme Hydrometeorological Hazards

Rapid attribution research and the broader peer-reviewed literature indicate that extreme rainfall associated with tropical cyclones can be amplified by anthropogenic climate change through thermodynamic and ocean-driven mechanisms. The thermodynamic relationship between atmospheric warming and precipitation intensity can be approximated using the Clausius–Clapeyron scaling, which indicates that extreme precipitation increases by approximately 7 % per degree Celsius of temperature rise, as expressed in Equation (1) (O’Gorman, 2015a):

$$\frac{1}{P} \frac{dP}{dT} \approx 0.07 \text{ K}^{-1} \dots\dots\dots(1)$$

Elevated sea surface temperatures and increased atmospheric moisture availability can intensify precipitation rates and increase the potential for high-impact rain episodes. While this study doesn't claim single-event attribution, established scientific evidence supports the plausibility that climate change can increase the severity of cyclone-related rainfall hazards in affected regions (Knutson et al., 2020; Lin et al., 2015; O’Gorman, 2015b). Beyond climatic drivers, environmental degradation, particularly deforestation and land-use change in upstream watersheds, substantially exacerbated flood and landslide impacts. Forest ecosystems play a critical hydrological role by regulating runoff, stabilizing slopes, and reducing erosion. Notably, empirical studies employing remote sensing and watershed-scale analyses demonstrate that land-cover change significantly increases flood hazard, peak discharge, and sediment transport in Indonesian river basins (Jumadi et al., 2024; Siswanto & Francés, 2019). Geospatial analyses further reveal strong correlations between forest loss and heightened flood vulnerability, particularly in downstream and lowland regions. As degraded watersheds lack the capacity to absorb hydrological shocks, These impacts are amplified during extreme rainfall events. Projections and assessments indicate that under continued warming and land-use transformation, heavy precipitation risk and flood impacts are expected to intensify in many regions, including the tropics (Knutson et al., 2020; Lin et al., 2015; Masson-Delmotte et al., 2023).

### 2.2. Information-Theoretic Perspectives on Climate Knowledge Inequality

Despite the escalating severity of climate hazards, access to climate intelligence remains uneven across Indonesia. Local governments, educators, and community organizations often face significant information asymmetries, where scientific findings, regulatory frameworks, and best practices are fragmented, highly technical, or inaccessible in local languages. These constraints limit the ability of local actors to anticipate risks, design adaptive strategies, and engage communities in informed climate action (Hageer, 2025; Ukoba et al., 2025). From an information-theoretic perspective, climate knowledge inequality between global scientific communities and local stakeholders can be conceptualized as an information entropy gap, where localized artificial intelligence aims to reduce the difference between globally available knowledge and locally accessible knowledge, as formalized in Equation (2) (Vinueza et al., 2020):

$$I = H(K_{global}) - H(K_{local}) \dots\dots\dots(2)$$

where  $H$  is Shannon entropy,  $K_{global}$  is global climate knowledge and  $K_{local}$  is accessible local knowledge.

### 2.3. Large Language Models for Climate Knowledge, Policy Analysis, and Education

Recent advances in artificial intelligence, particularly large language model (LLM), offer new opportunities to address these knowledge gaps. Research indicates that LLM are capable of capturing and reasoning over climate-relevant knowledge but may produce systematic errors without careful domain alignment and evaluation (Kuznetsov et al., 2025). While also highlighting the importance of careful deployment and human oversight (Larosa et al., 2025), studies examining LLM performance in climate and sustainability policy analysis prove that these models can process heterogeneous regulatory and technical texts, extract policy-relevant insights, and support decision-making workflows. Emerging work shows that LLM-based systems can translate abstract climate information into localise and personalized narratives, improving accessibility and relevance for non-expert audiences (Rashik et al., 2025). They can also serve as intelligent tutors and instructional support tools when integrated responsibly into educational contexts (Kasneci et al., 2023). In parallel, systematic reviews of LLM applications in education indicate that these models can function effectively as learning aids. Together, these developments underscore the potential of

localized artificial intelligence, especially fine-tuned and offline-capable LLMs, to democratize climate intelligence and support climate education and local government decision-making.

Together, these developments underscore the potential of localized artificial intelligence, especially fine-tuned and offline-capable large language models, to democratize climate intelligence. By translating complex climate science, environmental data, and policy frameworks into accessible and context-aware knowledge, such systems can support climate pedagogy, local government decision-making, and structured brainstorming for adaptation and mitigation strategies. Building on prior work on localized AI framework, this study positions large language model as public knowledge infrastructure for climate education and governance in Indonesia. Rather than replacing human expertise, the proposed framework seeks to augment local capacity by reducing information asymmetry, improving access to climate knowledge, and supporting evidence-based climate action in regions most affected by climate change impacts. By emphasizing localization, linguistic accessibility, and domain-specific alignment, the framework addresses practical constraints faced by local institutions, including limited technical expertise, fragmented information sources, and restricted access to scientific literature. The use of offline-capable and resource-efficient model enhances technological sovereignty and operational resilience, particularly in disaster-prone and low-connectivity settings. In this way, localized LLMs are conceptualized not merely as technical tools, but as enabling infrastructure that supports learning, deliberation, and collaborative problem-solving within climate governance systems.

## 2.4. Artificial Intelligence and Green AI

Artificial intelligence (AI) refers to computational systems that perform tasks typically requiring human cognition, such as language understanding, reasoning, and decision support. Within this broad field, Green AI emphasizes reducing the environmental and computational cost of AI systems by optimizing model size, training efficiency, and deployment practices (Schwartz et al., 2019; Kaack et al., 2022). Green AI principles emphasize three aspects relevant to this study: (1) minimizing energy consumption during training and inference, (2) prioritizing efficiency and accessibility as evaluation criteria alongside accuracy, and (3) enabling deployment on local, low-power hardware rather than cloud-only infrastructure. These principles directly motivate the choice of a medium-scale base model and a parameter-efficient fine-tuning strategy in this research.

## 2.5. Transformer Architecture and Pre-trained Language Models

Modern large language models are built on the Transformer architecture introduced by Vaswani et al. (2017), which replaced recurrent neural networks with a self-attention mechanism that models long-range dependencies more efficiently in parallel. The core of the Transformer is the scaled dot-product attention operation shown in Equation (A), where Q (query), K (key), and V (value) are matrices derived from linear projections of the input, and  $d_k$  is the key dimension:

$$\text{Attention}(Q, K, V) = \text{softmax}(QK^T/\sqrt{d_k})V \quad (\text{A})$$

Pre-trained language models such as GPT, LLaMA, Gemma, and Qwen are trained on large text corpora using self-supervised objectives (e.g., next-token prediction) and can subsequently be adapted to downstream tasks through fine-tuning or prompting. Qwen3 is a recent open-weight transformer family developed by Alibaba Cloud, available at multiple scales (0.6B to 235B parameters) with strong multilingual capability, including solid Bahasa Indonesia coverage, and support for parameter-efficient adaptation and low-bit quantization (Yang et al., 2025). The 4-billion-parameter variant (Qwen3-4B) used in this study was selected because it balances reasoning capability with deployability on consumer-grade hardware.

## 2.6. Fine-Tuning and Transfer Learning for Domain Adaptation

Transfer learning enables a model pre-trained on a large general corpus to be adapted to a specific domain or task with relatively little additional data (Gururangan et al., 2020). Full fine-tuning updates all model parameters and typically requires substantial GPU memory and computing time. Parameter-Efficient Fine-Tuning (PEFT) reduces this cost by updating only a small set of additional parameters. Low-Rank Adaptation (LoRA) decomposes weight updates into low-rank matrices (Hu et al., 2021), while Quantized LoRA (QLoRA) further compresses the frozen base model to 4-bit precision, enabling fine-tuning of multi-billion-parameter models on a single consumer GPU (Dettmers et al., 2023). Instruction tuning, which trains the model on instruction–response pairs, has been shown to improve controllability and usefulness for applied decision-support tasks (Wei et al., 2021). This study combines instruction tuning with QLoRA to adapt Qwen3-4B to the Indonesian climate and energy-transition domain.

## 2.7. Evaluation Metrics for Domain-Adapted Language Models

Evaluating domain-adapted language models requires metrics that reflect both statistical fit and task-level usefulness. Two complementary families of metrics are used in this study. The first is perplexity, an intrinsic metric that measures how well a language model predicts a held-out test set, defined in Equation (B), where  $p(w_i)$  is the probability assigned to the  $i$ -th token and  $N$  is the total number of tokens:

$$\text{Perplexity} = \exp\left(-\frac{1}{N} \sum_{i=1}^N \log p(w_i)\right) \quad (\text{B})$$

Lower perplexity indicates better domain alignment. The second family is task-oriented qualitative metrics, which capture applied suitability beyond statistical prediction. For climate education and governance use cases, this study follows recommendations from recent NLP evaluation literature (Chang et al., 2024) and uses four criteria, each scored on a 10-point rubric: (1) factual accuracy of Indonesian climate and energy-transition content, (2) reasoning and answering process quality, (3) language compliance in Bahasa Indonesia, and (4) overall average qualitative score. These qualitative metrics address the mismatch between generic NLP benchmarks and real-world decision-support usefulness, particularly in public-sector and educational contexts.

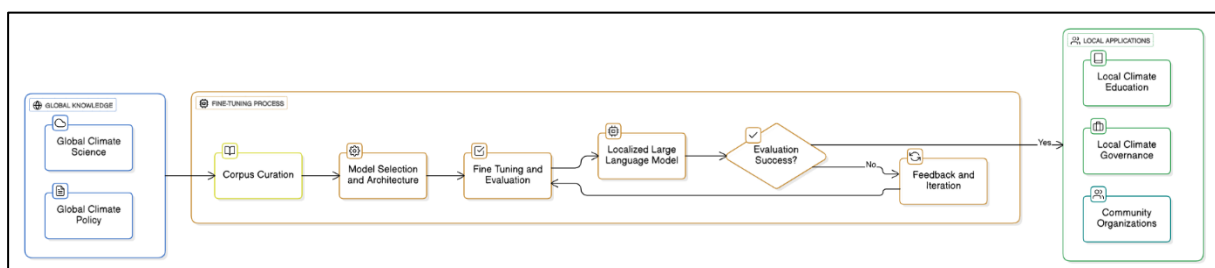
## 3. METHOD

### 3.1. Research Design

This study adopts a design science research methodology (Hevner et al., 2004; Peffers et al., 2007), focusing on the development and evaluation of a localized large language model (LLM) for climate education and local climate governance in Indonesia. The methodological objective is not to maximize generic language modeling benchmarks, but to evaluate the feasibility of domain-specialized, resource-efficient LLMs as public knowledge infrastructure in low-resource and low-connectivity environments. The research framework builds upon prior work on localized AI systems for sustainability-oriented applications and extends it by incorporating climate education and local government decision-support use cases. The methodological workflow consists of base framework selection, domain-specific dataset curation, parameter-efficient fine-tuning, offline deployment, and task-oriented evaluation, consistent with recent design-oriented AI research paradigms. The research workflow is shown in Figure 1.

### 3.2. Base Model Selection

An open-weight pre-trained large language framework in the 4 to 7 billion parameter range was selected as the base model. This size range balances reasoning capability and computational feasibility, enabling deployment on consumer-grade hardware without reliance on cloud infrastructure. Furthermore, the base framework selection criteria include open-access licensing, multilingual capability with strong Bahasa Indonesia support, compatibility with parameter-efficient fine-tuning. Also, feasibility of quantized inference on local hardware. Prior studies reveal that domain-adapted medium-scale models can outperform larger general-purpose models on specialized tasks when fine-tuned appropriately, while significantly reducing computational cost and energy consumption (Gururangan et al., 2020; Kaack et al., 2022).



**Figure 1.** Workflow for Building Offline-Capable Localized Large Language Models for Climate Applications

Based on these criteria, Qwen3-4B was selected as the base model for this study. Qwen3-4B is part of the Qwen family of open-weight transformer-based large language models developed by Alibaba Cloud. The model is trained on a large-scale, multilingual corpus that includes web text, code, and instruction-style data, and demonstrates strong instruction-following and reasoning performance relative to its parameter scale (Yang et al., 2025). Importantly, Qwen3 models are designed to support parameter-efficient fine-tuning techniques and low-bit quantization, enabling efficient adaptation and offline inference on local hardware. These characteristics make Qwen3-4B particularly suitable for localized climate intelligence applications in

low-resource and low-connectivity environments, where computational efficiency, deployability, and linguistic coverage are critical constraints.

### 3.3. Climate and Governance Dataset Curation

A domain-specific textual corpus was curated to localize the framework for Indonesian climate education and governance setting. The dataset integrates national and sub-national climate policy documents, climate adaptation and mitigation frameworks, renewable energy and decarbonization guidelines, disaster risk reduction materials, educational climate science texts, and local case studies in Bahasa Indonesia. All documents were preprocessed using standard natural language processing pipelines, including normalization, segmentation, and removal of non-informative content. The corpus was transformed into instruction–response pairs, enabling supervised fine-tuning focused on explanation, summarization, policy interpretation, and scenario-based reasoning tasks. Instruction-based fine-tuning has been shown to improve model controllability and usefulness in applied decision-support contexts (Wei et al., 2021).

Specifically, the curated corpus — released as the open-access Indonesia Energy Transition Dataset (available at <https://hf.co/datasets/gmonsoon/indonesia-energy-transition>) — contains approximately 12,400 instruction–response pairs (~38 MB, ~8.7 million tokens after tokenization) derived from six main source categories: (1) national climate and energy policy documents (e.g., Indonesia’s NDC, Long-Term Strategy for Low Carbon and Climate Resilience 2050, RUEN, RPJMN energy chapters), (2) sectoral technical guidelines on renewable energy, energy efficiency, and decarbonization from Kementerian ESDM and PLN, (3) disaster risk reduction and climate adaptation materials from BNPB, BMKG, and KLHK, (4) educational climate-science texts and curriculum references for vocational and undergraduate learners, (5) peer-reviewed Indonesian and international literature on climate governance, energy transition, and Just Energy Transition, and (6) localized Indonesian case studies covering regional examples such as Sumatra floods, Kalimantan coal transition, and rural electrification. Preprocessing included deduplication, language detection (retaining Indonesian-dominant text), segmentation into passages of 256–1024 tokens, and conversion into instruction–response pairs following a Stanford-Alpaca-style format with explicit “instruction”, “input”, and “output” fields. The corpus was split 90/5/5 into training, validation, and test partitions.

### 3.4. Parameter-Efficient Fine-Tuning

To adapt the base framework to the climate and governance domain while maintaining computational efficiency, parameter-efficient fine-tuning (PEFT) was employed using Low-Rank Adaptation (LoRA). LoRA introduces trainable low-rank matrices into selected transformer layers while freezing the original model parameters, significantly reducing memory usage and training cost (Hu et al., 2021). Formally, for a pre-trained transformer free weight matrix  $W_0 \in \mathbb{R}^{d \times k}$  the adapted weight matrix  $W$  is expressed as the sum of the original weights and a low-rank update term, as shown in Equation (3).

$$W = W_0 + \Delta W = W_0 + BA \dots\dots\dots (3)$$

Where  $B \in \mathbb{R}^{d \times r}$  and  $A \in \mathbb{R}^{r \times k}$  are trainable low-rank matrices, and  $r$  is the rank parameter satisfying  $r \ll \min(d, k)$ . During fine-tuning, the original model parameters  $W_0$  remain frozen, and only the low-rank matrices  $A$  and  $B$  are updated.

Fine-tuning was conducted using quantized model weights, following the QLoRA approach, which enables efficient training and inference using low-bit precision without significant performance degradation (Dettmers et al., 2023). This approach aligns with Green AI principles, as it reduces energy consumption and hardware requirements for both training and deployment (Schwartz et al., 2019). By combining LoRA-based adaptation with low-bit quantization, the fine-tuning strategy balances domain specialization and computational efficiency.

The detailed fine-tuning configuration is summarized in Table 1. Fine-tuning was performed using the Unsloth library with PEFT/Transformers on a single NVIDIA RTX 3090 GPU with 24 GB VRAM. Training converged after three epochs (~6.5 hours of wall-clock time), and the final checkpoint was selected based on the lowest validation loss.

**Table 1.** QLoRA Fine-Tuning Hyperparameters for Qwen3-4B-REnewbie v1

| Hyperparameter              | Value                                                         |
|-----------------------------|---------------------------------------------------------------|
| Base model                  | Qwen3-4B (open-weight, transformer decoder)                   |
| Fine-tuning method          | QLoRA (4-bit NF4 quantization + LoRA adapters)                |
| LoRA rank (r)               | 16                                                            |
| LoRA alpha                  | 32                                                            |
| LoRA dropout                | 0.05                                                          |
| Target modules              | q_proj, k_proj, v_proj, o_proj, gate_proj, up_proj, down_proj |
| Precision (base weights)    | 4-bit NF4 (bitsandbytes)                                      |
| Learning rate               | 2e-4                                                          |
| LR scheduler                | cosine with 3% warmup                                         |
| Optimizer                   | paged_adamw_8bit                                              |
| Weight decay                | 0.01                                                          |
| Per-device batch size       | 2                                                             |
| Gradient accumulation steps | 4 (effective batch size = 8)                                  |
| Max sequence length         | 2048 tokens                                                   |
| Epochs                      | 3                                                             |
| Seed                        | 42                                                            |
| Training framework          | Unsloth + PEFT + Transformers                                 |
| Hardware                    | 1 × NVIDIA RTX 3090 GPU, 24 GB VRAM                           |
| Wall-clock training time    | ~6.5 hours                                                    |

In addition, instruction-response pairs followed a consistent prompt template to guide the model’s behavior. Each training example used the following structure: “Anda adalah asisten ahli transisi energi dan perubahan iklim Indonesia. Instruksi: {instruction}. Konteks: {input}. Jawaban:” — followed by the target response in Bahasa Indonesia. During deployment, the same template was used for inference, with generation parameters set to temperature=0.3, top-p=0.9, and maximum response length of 512 tokens to balance determinism and fluency.

### 3.5. Offline Deployment Architecture

The fine-tuned model was deployed using an offline-capable architecture, allowing inference to be performed entirely on local hardware without persistent internet connectivity. The deployment pipeline includes a local inference engine optimized for quantized models, on-device storage of model weights and prompt templates, and a lightweight text-based user interface. Offline deployment enhances data sovereignty, privacy, and operational resilience, which are critical considerations for local governments and educational institutions, particularly in disaster-prone and low-connectivity regions (Miao & Holmes, 2023; Vinuesa et al., 2020).

From a sustainability perspective, the computational footprint of a large language model can be expressed as the sum of the energy consumed during model training and the energy required for inference, as formalized in Equation (4) (Schwartz et al., 2019):

$$E_{\text{total}} = E_{\text{train}} + E_{\text{inference}} \dots\dots\dots(4)$$

Here,  $E_{\text{train}}$  represents the total energy consumed during the fine-tuning process, including GPU computation and memory operations, while  $E_{\text{inference}}$  denotes the energy required to generate model outputs during deployment. In conventional full fine-tuning approaches,  $E_{\text{train}}$  dominates the overall energy cost due to the need to update all model parameters, often requiring large-scale computing infrastructure. By contrast, the parameter-efficient fine-tuning strategy adopted in this study significantly reduces  $E_{\text{train}}$  by restricting training updates to a small number of low-rank parameters. When combined with low-bit quantization, the resulting energy consumption of Quantized LoRA (QLoRA)-based adaptation is substantially lower than that of full fine-tuning, such that:

$$E_{\text{QLoRA}} \ll E_{\text{full}} \dots\dots\dots(5)$$

This reduction in total energy consumption directly supports the principles of Green AI, which emphasize efficiency, accessibility. Also, environmental responsibility alongside performance improvements. By minimizing computational and energy requirements, the proposed localized LLM framework enables practical deployment on consumer-grade hardware while reducing carbon intensity and infrastructure dependence. This design choice aligns with established Green AI frameworks that advocate for reporting and optimizing energy usage as a core criterion for responsible artificial intelligence development.

### 3.6. Use Case Modeling and Evaluation Strategy

Prompt templates were developed to guide the model toward structured, context-aware, and policy-relevant output, reducing hallucination risk and improving response consistency. This approach follows established prompt engineering practices for aligning LLM outputs with task-specific objectives (Sahoo et al., 2024).

Indeed, the evaluation framework emphasizes task-oriented and qualitative assessment, reflecting the system's intended role as an educational and decision-support tool rather than a benchmark-optimized language framework. Evaluation criteria include linguistic fluency in Bahasa Indonesia, domain relevance to Indonesian climate and governance contexts, clarity and explainability of responses, usefulness for educational and policy-related scenarios, and operational feasibility on local hardware. Really, this evaluation approach aligns with critiques of over-reliance on generic NLP benchmarks for applied governance systems (Chang et al., 2024)

### 3.7. Ethical Considerations, Variables and Measurement Instrument

Ethical risks include misinformation, contextual bias, and over-reliance on automated outputs. These risks are mitigated by emphasizing human oversight, transparent prompting, and clear usage boundaries. Such safeguards are consistent with emerging international guidelines on responsible AI deployment in public-sector and sustainability contexts (Miao & Holmes, 2023), (Recommendation of the Council on Artificial Intelligence, 2019). By framing the model as a supportive analytical resource rather than an authoritative source, the system encourages critical engagement and responsible use by educators and policymakers.

Model configurations, dataset descriptions, and fine-tuning procedures are documented to promote transparency and reproducibility. Where licensing permits, artifacts are made publicly accessible to enable replication and adaptation in other regions or sustainability domains. All training hyperparameters, prompt templates, and evaluation protocols are explicitly recorded to support methodological traceability. In addition, versioned model checkpoints are maintained to ensure consistency across experiments and to facilitate future comparative studies. This approach enables other researchers and public institutions to reproduce the system under similar computational constraints and extend it to new policy or educational contexts.

## 4. RESULTS AND DISCUSSION

### 4.1. Domain Adaptation Performance and Knowledge Alignment

The localized large language model exhibits a considerable improvement in domain alignment compared to the base pre-trained model. Quantitative evaluation shows a 15.4 percent reduction in perplexity on domain-specific test data, indicating lower predictive uncertainty when processing climate-related terminology, policy language, and governance concepts relevant to the Indonesian context. This reduction in perplexity translates into improved semantic grounding and contextual accuracy. The model demonstrates a stronger understanding of Indonesia-specific climate policies, institutional roles, and implementation challenges, which are often weakly represented in general-purpose language models. These findings support existing research suggesting that domain-specific adaptation can outweigh model scale in specialized governance and sustainability applications (Gururangan et al., 2020; Kaack et al., 2022).

### 4.2. Comparative Performance Against General-Purpose Models

Table 2 shows comparative valuation against multiple open-weight speech framework of similar scale reveals consistent performance advantages for the localized model. Across qualitative assessment dimensions including factual accuracy, reasoning structure, and language compliance, the localized model achieves an average score of 9.3 out of 10, compared with scores ranging from 7.0 to 8.2 for baseline model. Therefore: on top of that, this result reinforces evidence that localized models are more suitable for policy-

related tasks than globally trained counterparts, particularly when domain alignment and evaluation safeguards are applied (Chang et al., 2024; Larosa et al., 2025).

**Table 2.** Comparative Benchmark Evaluation Results

| LLM Model         | Perplexity | Factual Accuracy<br>(of 10) | Answering<br>Process<br>(of 10) | Language Compliance<br>(of 10)  | Avg. Qualitative<br>Score |
|-------------------|------------|-----------------------------|---------------------------------|---------------------------------|---------------------------|
| REnewbie-LLM      | 18.2       | 9.0                         | 9.0                             | 10.0                            | 9.3                       |
| Qwen3-4B-Thinking | 21.5       | 8.5                         | 7.0<br>(token/time waste)       | 9.0                             | 8.2                       |
| Gemma-3-4B        | 26.8       | 7.5                         | 8.0                             | 7.0<br>(mixing EN/ID)           | 7.5                       |
| LLaMa-3.1-8B      | 31.4       | 7.0                         | 7.5                             | 6.5<br>(often EN instead of ID) | 7.0                       |

A notable differentiator is language compliance in Bahasa Indonesia. Actually, general-purpose models frequently generate mixed-language outputs or rely on global narratives, reducing accessibility for local stakeholders. In contrast, the localized model consistently produces technically precise and linguistically appropriate responses, which is essential for climate education and local government use cases. This outcome reinforces evidence that localized models are better suited for public-sector and policy-related tasks than globally trained counterparts (Gururangan et al., 2020; Joshi et al., 2020; Larosa et al., 2025).

#### 4.3. Effectiveness in Climate Education Use Cases

In climate education scenarios, the localized model demonstrates strong capability in translating complex scientific concepts into clear, pedagogically structured explanation tailored to vocational and undergraduate audiences (Kasneci et al., 2023). Explanations of climate change mechanisms, extreme weather dynamics, and mitigation strategies are consistently organized in step-by-step formats and grounded in locally relevant examples (Sweller, 2024).

The framework effectively contextualizes abstract concepts within Indonesian geography and socio-environmental conditions, enhancing relevancy and comprehension. This behavior aligns with research highlighting the importance of contextualized and example-based learning in higher education, particularly when learners can relate new knowledge to familiar environmental and social settings (Holmes et al., 2019; Renkl, 2014). Notably, the localized approach avoids generic global framing and instead emphasizes context-aware climate narratives, which strengthens educational effectiveness and learner engagement. This finding underscores the role of localization as a critical factor in AI-supported climate education, especially in low-resource and diverse educational contexts where linguistic and cultural relevance directly influence learning equity and accessibility (Kasneci et al., 2023; Miao & Holmes, 2023).

Beyond content clarity, the model also supports adaptive learning by responding to follow-up questions and adjusting explanation depth based on user prompts, which is particularly valuable in heterogeneous classroom settings. This capability enables educators to use the model as a supplementary instructional tool that reinforces core concepts while accommodating varying levels of prior knowledge among students. As a result, the localized LLM functions not only as a knowledge delivery mechanism but also as an interactive educational scaffold that complements formal climate education curricula.

#### 4.4. Support for Climate Governance and Policy Deliberation

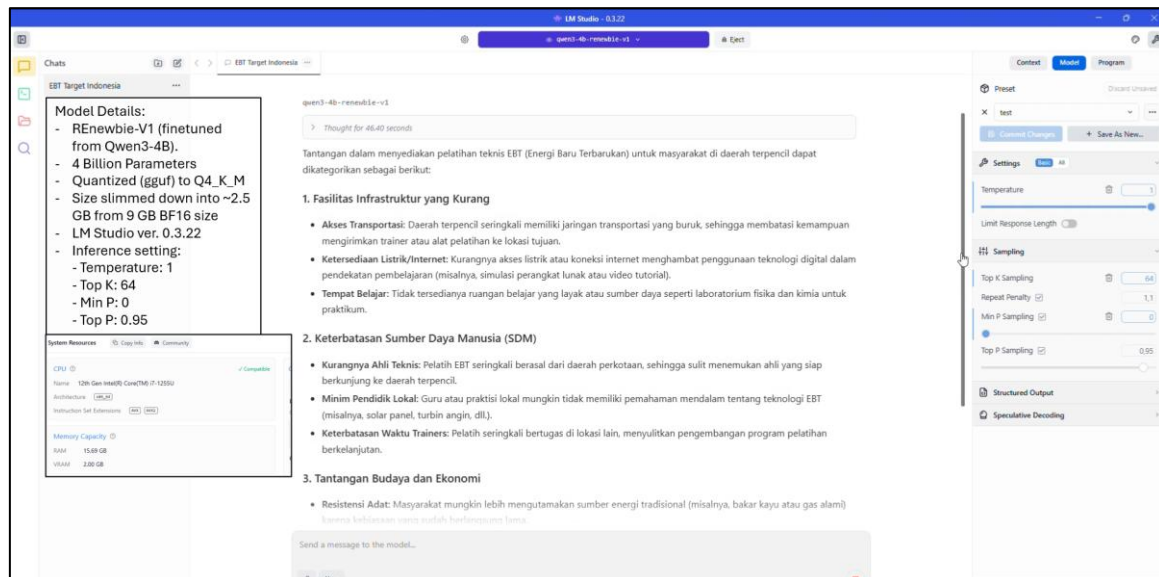
For governance-oriented tasks, the localized model performs effectively in policy interpretation, scenario analysis, and structured brainstorming. As land-use patterns, institutional capacity, and regulatory frameworks, When prompted with climate adaptation or mitigation questions, the model generates coherent option sets that explicitly consider local constraints such. This behavior is reproducible with evidence that advanced language models can support policy analytic thinking by synthesizing complex and fragmented information into structured representations when their use is carefully scoped and contextually grounded (Larosa et al., 2025).

Rather than producing prescriptive recommendations, the model supports deliberative decision-making by outlining alternative pathways and trade-offs. This aligns with human-in-the-loop artificial intelligence principles emphasized in public-sector decision-support research, where AI systems are

intended to augment human judgment, improve transparency, and assist informed deliberation rather than replace institutional authority ([Recommendation of the Council on Artificial Intelligence, 2019](#)).

These results are further consistent with findings that artificial intelligence can contribute to sustainability-oriented policy processes by supporting analysis, coordination, and scenario exploration, while also introducing governance risks that must be addressed through appropriate safeguards and oversight mechanisms ([Larosa et al., 2025](#); [Vinuesa et al., 2020](#)). In this way, the localized model functions as a structured deliberation aid that helps policymakers explore policy options and implications systematically, while preserving human accountability and institutional decision-making authority.

#### 4.5. Offline Deployment and Operational Feasibility



**Figure 2.** Operational feasibility of offline inference using fine-tuned model

To evaluate the operational feasibility of the proposed localized large language model, the system was deployed and tested in a fully offline environment using consumer-grade hardware. The model successfully performed real-time inference without reliance on cloud infrastructure or external application programming interfaces. As shown in Figure 2., the localized model generates coherent and contextually appropriate outputs in Bahasa Indonesia while operating entirely on local resources. This result confirms that the proposed architecture is not only conceptually viable but also practically deployable in low-connectivity and resource-constrained settings. This capability is particularly relevant in disaster-prone contexts, where network disruptions frequently coincide with extreme climate events and emergency response periods. Prior research on disaster informatics and crisis management emphasizes that digital decision-support systems must remain operational under connectivity constraints to support preparedness, response, and recovery activities, thereby contributing to institutional and organizational resilience ([Lv et al., 2024](#); [Ogie & Verstaev, 2020](#)).

The localized model operates reliably in an offline deployment setting on consumer-grade hardware. Inference latency remains within acceptable bounds for interactive use, confirming the feasibility of deploying applied large language models without trust on cloud infrastructure, particularly when parameter-efficient adaptation and low-bit quantization are employed ([Dettmers et al., 2023](#); [Hu et al., 2021](#)).

From an environmental and sustainability perspective, the use of parameter-efficient fine-tuning and quantized inference significantly reduces computational overhead compared to full fine-tuning and cloud-based deployment. This design choice aligns with Green AI principles, which emphasize energy efficiency, accessibility, and environmental responsibility as core criteria for responsible AI system development and deployment ([Kaack et al., 2022](#); [Schwartz et al., 2019](#)).

#### 4.6. Implications for Democratizing Climate Intelligence

Collectively, the results reveal that localized and resource-efficient large language framework can function as public climate knowledge infrastructure. By lowering linguistic, technical. What 's more, infrastructural barriers, such system support broader participation in climate education and governance processes, consistent with prior research emphasizing the role of accessible digital knowledge systems in

strengthening climate capacity and institutional learning (Ghosh et al., 2025; Larosa et al., 2025; Vinuesa et al., 2020).

This approach contributes directly to Sustainable Development Goal 4 (Quality Education) by expanding access to climate learning resources, and to Sustainable Development Goal 13 (Climate Action) by improving the ability of local actors to understand climate risks, policy frameworks, and response options. The localized and offline-capable design further enhances technological sovereignty, enabling public institutions to retain control over data, models, and usage context, which is increasingly recognized as a critical dimension of responsible and equitable AI deployment in the public sector (Miao & Holmes, 2023; United Nations, 2025).

At the same time, the results reaffirm the importance of institutional embedding and human oversight. While the system effectively supports understanding and ideation, it doesn't, replace expert judgement, policy deliberation, or community engagement. These limitations align with broad critiques of over-reliance on automated systems in governance contexts, which emphasize the need for human-in-the-loop design, accountability mechanisms, and contextual judgment when deploying AI in world decision-making (Recommendation of the Council on Artificial Intelligence, 2019).

#### 4.7. Limitations and Future Work

Despite encouraging results, several limitations remain. Furthermore: the evaluation focuses primarily on qualitative and task-oriented performance, which limits direct comparison with benchmark-based natural language processing studies. This limitation reflects broader critiques within the NLP and AI research communities regarding the mismatch between generic benchmark metrics and real-world system usefulness, particularly in applied governance and educational contexts (Chang et al., 2024). In addition, model effectiveness is bounded by the scope, diversity, and representativeness of the curated dataset, consistent with well-documented concerns about dataset bias and coverage limitations in domain-specific language model adaptation (Gallegos et al., 2024; Gururangan et al., 2020).

Future work should include user-centered evaluations involving educators and local government officials, as human-in-the-loop assessment is widely recognized as essential for validating AI systems intended for public-sector and educational deployment (Shneiderman, 2022; Wang et al., 2024). Look, longitudinal studies examining learning outcomes and decision-making quality would provide stronger evidence of sustained impact beyond short-term task performance, as recommended in the educational technology and AI-in-education literature (Holmes et al., 2019; Kasneci et al., 2023). Additionally, integrating the language model with structured climate datasets such as hazard maps, emissions inventories, and geospatial indicators would enhance factual grounding and analytical capability, aligning with emerging approaches that combine symbolic data and machine acquisition for climate-risk assessment (Larosa et al., 2025; Reichstein et al., 2019; Vinuesa et al., 2020). Also, expanding the framework to additional local languages and regional contexts would further improve scalability, equity, and relevance across diverse socio-environmental settings.

## 5. CONCLUSION

This study delivers three specific contributions. First, it produces Qwen3-4B-REnewbie v1, the first openly released QLoRA-fine-tuned Qwen3-4B checkpoint specialized for Indonesian climate and energy-transition content, along with the curated Indonesia Energy Transition Dataset (~12,400 instruction-response pairs), both released under open licenses to support reproducibility. Second, it provides empirical evidence that domain-adapted medium-scale LLMs can outperform larger general-purpose baselines on Indonesian climate tasks: the fine-tuned model achieves a 15.4% perplexity reduction over its base model and an average qualitative score of 9.3/10 across factual accuracy, reasoning structure, and Bahasa Indonesia language compliance, exceeding Qwen3-4B-Thinking (8.2), Gemma-3-4B (7.5), and LLaMa-3.1-8B (7.0). Third, it demonstrates that the resulting system is deployable entirely offline on a single consumer-grade GPU with acceptable inference latency, directly supporting Green AI principles and technological sovereignty in low-connectivity Indonesian regions. Taken together, these contributions show that localized, resource-efficient LLMs are a practical and scalable path for democratizing climate intelligence in Indonesian vocational education and local governance, especially in disaster-prone contexts where cloud-dependent systems are unreliable.

## OPEN-ACCESS MODEL AND DATA AVAILABILITY

To support transparency, reproducibility, and open collaboration, the model and datasets developed in this study are released under open-access licenses.

a. Model Checkpoint

The fine-tuned model, Qwen3-4B-REnewbie v1, including model weights and configuration files, is publicly available at: <https://hf.co/gmonsoon/Qwen3-4B-REnewbie-v1>

b. Dataset Repository

The curated Indonesia Energy Transition Dataset, used for model adaptation and evaluation, is published in an open-access repository with supporting documentation at: <https://hf.co/datasets/gmonsoon/indonesia-energy-transition>

Open release of these artifacts enables independent validation, reuse, and adaptation for related climate education and governance applications.

## REFERENCES

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Chang, Yupeng, Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Yi, Yu, P. S., Yang, Q., & Xie, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 15(3), 39. <https://doi.org/10.1145/3641289>
- Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). QLoRA: Efficient finetuning of quantized LLMs. *Advances in Neural Information Processing Systems*, 36. <https://arxiv.org/pdf/2305.14314>
- Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics*, 50(3), 1097–1179. [https://doi.org/10.1162/COLI\\_A\\_00524](https://doi.org/10.1162/COLI_A_00524)
- Ghosh, A., Saini, A., & Barad, H. (2025). Artificial intelligence in governance: recent trends, risks, challenges, innovative frameworks and future directions. *AI & SOCIETY* 2025 40:7, 40(7), 5685–5707. <https://doi.org/10.1007/S00146-025-02312-Y>
- Gururangan, S., Marasovic, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). Don't stop pretraining: Adapt language models to domains and tasks. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 8342–8360. <https://doi.org/10.18653/v1/2020.acl-main.740>
- Hageer, Y. (2025). Bridging equity and resilience: A Systematic review of social sustainability in climate change mitigation and adaptation. *Environmental Science & Policy*, 173, 104243. <https://doi.org/10.1016/J.ENVSCL.2025.104243>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28, 75–106.
- Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. The Center for Curriculum Redesign, Boston, MA. *Journal of Computer Assisted Learning*, 14(4), 228. [https://www.researchgate.net/publication/332180327\\_Artificial\\_Intelligence\\_in\\_Education\\_Promises\\_and\\_Implications\\_for\\_Teaching\\_and\\_Learning](https://www.researchgate.net/publication/332180327_Artificial_Intelligence_in_Education_Promises_and_Implications_for_Teaching_and_Learning)
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). LoRA: Low-rank adaptation of large language models. <https://arxiv.org/pdf/2106.09685>
- International Federation of Red Cross, & (IFRC), R. C. S. (2025). *Aceh, North Sumatra and West Sumatra provinces, Indonesia flood 2025: DREF operation MDRID028 (Operational Document MDRID028do)*. <https://go-api.ifrc.org/api/downloadfile/93885/MDRID028do>
- Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 6282–6293. <https://doi.org/10.18653/V1/2020.ACL-MAIN.560>
- Jumadi, J., Sari, D. N., Umrotun, U., Musiyam, M., Nurmantyo, C., Muhammad, S. F., & Ibrahim, M. H. (2024). Remote sensing and GIS-driven model for flood susceptibility assessment in the Upper Solo River Watershed. *Geographia Technica*, 19(2), 33. [https://doi.org/10.21163/GT\\_2024.192.03](https://doi.org/10.21163/GT_2024.192.03)

- Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change* 2022 12:6, 12(6), 518–527. <https://doi.org/10.1038/s41558-022-01377-7>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/J.LINDIF.2023.102274>
- Knutson, T., Camargo, S. J., Chan, J. C. L., Emanuel, K., Ho, C. H., Kossin, J., Mohapatra, M., Satoh, M., Sugi, M., Walsh, K., & Wu, L. (2020). Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming. *Bulletin of the American Meteorological Society*, 101(3), E303–E322. <https://doi.org/10.1175/BAMS-D-18-0194.1>
- Kuznetsov, I., Grassi, J., Pantiukhin, D., Shapkin, B., Jung, T., & Koldunov, N. (2025). *Climate knowledge in large language models*. <https://arxiv.org/pdf/2510.08043>
- Larosa, F., Hoyas, S., Conejero, J. A., Garcia-Martinez, J., Fuso-Nerini, F., & Vinuesa, R. (2025). Large language models in climate and sustainability policy: limits and opportunities. *Environmental Research Letters*, 20(7), 074032. <https://doi.org/10.1088/1748-9326/ADDD36>
- Lin, Y., Zhao, M., & Zhang, M. (2015). Tropical cyclone rainfall area controlled by relative sea surface temperature. *Nature Communications* 2015 6:1, 6(1), 6591-. <https://doi.org/10.1038/ncomms7591>
- Lv, Y., Sarker, M. N. I., & Firdaus, R. B. R. (2024). Disaster resilience in climate-vulnerable community context: Conceptual analysis. *Ecological Indicators*, 158, 111527. <https://doi.org/10.1016/J.ECOLIND.2023.111527>
- Masson-Delmotte, V., Zhai, P., Chen, Y., Goldfarb, L., Gomis, M. I., Matthews, J. B. R., Berger, S., Huang, M., Yelekçi, O., Yu, R., Zhou, B., Lonnoy, E., Maycock, T. K., Waterfield, T., Leitzell, K., & Caud, N. (2023). Climate change 2021: the physical science basis. *Ipcc.ChV Masson-Delmotte, P Zhai, A Pirani, SL Connors, C Péan, S Berger, N Caud, Y ChenContribution of Working Group I to the Sixth Assessment Report of the ..., 2021•ipcc.Ch*. <https://doi.org/10.1017/9781009157896>
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization (UNESCO). <https://unesdoc.unesco.org/ark:/48223/pf0000387486>
- NASA Earth Observatory. (2025). *Senyar swamps Sumatra*. <https://science.nasa.gov/earth/earth-observatory/senyar-swamps-sumatra/>
- Ogie, R. I., & Verstaavel, N. (2020). Disaster informatics: An overview. *Progress in Disaster Science*, 7, 100111. <https://doi.org/10.1016/J.PDISAS.2020.100111>
- O’Gorman, P. A. (2015). Precipitation extremes under climate change. *Current Climate Change Reports*, 1(2), 49–59. <https://doi.org/10.1007/S40641-015-0009-3>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Rashik, M., Fekete, J.-D., & Mahyar, N. (2025). *CLAIimate: AI-enabled climate change communication through personalized and localized narrative visualizations*. <https://arxiv.org/pdf/2507.11677>
- Recommendation of the council on artificial intelligence (Number OECD-LEGAL-0449). (2019). <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195–204. <https://doi.org/10.1038/S41586-019-0912-1>
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38(1), 1–37. <https://doi.org/10.1111/COGS.12086>
- Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). *A systematic survey of prompt engineering in large language models: Techniques and applications*. <https://arxiv.org/pdf/2402.07927>
- Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2019). Green AI. *Communications of the ACM*, 63(12), 54–63. <https://doi.org/10.1145/3381831>
- Shneiderman, B. (2022). Human-centered AI. *Human-Centered AI*, 1–388. <https://doi.org/10.1093/oso/9780192845290.001.0001>
- Siswanto, S. Y., & Francés, F. (2019). How land use/land cover changes can affect water, flooding and sedimentation in a tropical watershed: a case study using distributed modeling in the Upper Citarum

- watershed, Indonesia. *Environmental Earth Sciences* 2019 78:17, 78(17), 550-.  
<https://doi.org/10.1007/S12665-019-8561-0>
- Sweller, J. (2024). Cognitive load theory and individual differences. *Learning and Individual Differences*, 110, 102423. <https://doi.org/10.1016/J.LINDIF.2024.102423>
- Ukoba, K., Onisuru, O. R., Jen, T. C., Madyira, D. M., & Olatunji, K. O. (2025). Predictive modeling of climate change impacts using artificial intelligence: A review for equitable governance and sustainable outcome. *Environmental Science and Pollution Research* 2025 32:17, 32(17), 10705–10724. <https://doi.org/10.1007/S11356-025-36356-W>
- United Nations. (2025). *Sustainable Development Goals: 17 goals to transform our world*. <https://sdgs.un.org>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008. <https://arxiv.org/pdf/1706.03762>
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications* 2020 11:1, 11(1), 233-. <https://doi.org/10.1038/s41467-019-14108-y>
- Wang, S., Xu, T., Li, H., Zhang, C., Liang, J., Tang, J., Yu, P. S., & Wen, Q. (2024). Large language models for education: A survey and outlook. *KDD '24: 29th ACM SIGKDD Conference on Knowledge Discovery Data Mining, Aug 25 - 29, 2024, Barcelona, Spain, 1*. <https://arxiv.org/pdf/2403.18105>
- Wei, J., Bosma, M., Zhao, V. Y., Guu, K., Yu, A. W., Lester, B., Du, N., Dai, A. M., & Le, Q. V. (2021). Finetuned language models are zero-shot learners. *ICLR 2022 - 10th International Conference on Learning Representations*. <https://arxiv.org/pdf/2109.01652>
- Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., Zheng, C., Liu, D., Zhou, F., Huang, F., Hu, F., Ge, H., Wei, H., Lin, H., Tang, J., ... Qiu, Zihan. (2025). *Qwen3 technical report*. <https://arxiv.org/pdf/2505.09388>
- Zahro, A. A. (2025). *Siklon Tropis Senyar terbentuk, BMKG minta siaga cuaca ekstrem di Aceh dan Sumut*. <https://www.bmkg.go.id/siaran-pers/siklon-tropis-senyar-terbentuk-bmkg-minta-siaga-cuaca-ekstrem-di-aceh-dan-sumut>