Hiraia: An Offline, On-Device AI Science Tutor for Filipino Learners

Whitepaper — September 2026 (release 0.4.11)

Luis Buenaventura, primary proponent

Abstract

Hiraia is a free, offline, on-device artificial-intelligence science tutor for Filipino grade-school students. The system operates entirely on the low-cost Android devices already present in Filipino households: it requires no internet connection after installation, no subscription, and no user account. A single application package contains a Filipino-language large language model developed by the project and executed locally through the QVAC SDK from Tether; a curriculum-aligned knowledge bank exceeding twenty-one thousand facts in Tagalog, Cebuano, and English; a bank of 27,783 practice questions; an illustrated science feed of nearly nineteen thousand items; and neural text-to-speech voices for Tagalog and English. All components are aligned to the Department of Education's MATATAG curriculum, and all function in the absence of connectivity. This paper presents the motivating context, the system's positioning, its technical architecture — including the training of its language model — and its pedagogical design.

1. Context: The Philippine Education Crisis

The scale of the Philippine education crisis is documented most starkly in international assessment. In the 2022 cycle of the OECD's Programme for International Student Assessment (PISA), Filipino fifteen-year-olds recorded mean scores of approximately 355 in mathematics, 356 in science, and 347 in reading — among the lowest of the 81 participating education systems, against OECD means in the range of 470 to 485. These results confirmed rather than surprised: the country's first PISA participation, in 2018, produced comparable placements. The World Bank's complementary indicator is more direct still: its 2022 estimate placed learning poverty in the Philippines — the proportion of ten-year-olds unable to read and comprehend an age-appropriate text — at approximately ninety percent.

The most recent assessment cycle offers the first genuinely encouraging signal. In the PISA 2025 results, released in September 2026, Filipino students recorded mean scores of 367 in reading, 371 in mathematics, and 373 in science — improvements of twenty, sixteen, and seventeen points respectively over 2022, lifting the country out of the bottom ten for the first time, with the OECD describing the Philippines as the fastest-improving country in reading. The Department of Education

has attributed the gains to its reform programme, including reduced administrative loads on teachers, curricular improvement, and an emphasis on remediation. These gains deserve recognition; they do not amount to recovery. Filipino fifteen-year-olds remain roughly one hundred points below OECD averages in every domain — on the standard conversion of score points to time in school, a deficit of nearly three academic years. A trajectory this promising, against a gap this wide, argues for more of what works, delivered to more children, at lower cost per child: precisely the multiplication problem to which this project is addressed.

These outcomes rest on well-documented structural deficits. The Department of Education has acknowledged a national classroom shortage on the order of one hundred and fifty thousand rooms; class sizes in urban schools routinely exceed forty learners, and many schools operate on double or triple shifts. Tens of thousands of teaching positions remain unfilled, while serving teachers carry administrative burdens that displace instructional time. The COVID-19 pandemic imposed on Filipino learners one of the longest school closures in the world, compounding deficits that predated it. The MATATAG curriculum reform, introduced in 2023, represents a substantive effort to decongest and refocus instruction; a curriculum, however, cannot compensate for the system's binding constraint, which is adult instructional attention per child.

Hiraia is addressed to that constraint. It does not propose to substitute for teachers. It proposes to place a patient, factually careful science companion in the hands of every child whose household possesses a telephone — which is to say, in the hands of most Filipino children — at zero recurring cost.

2. Positioning: Why Offline Operation Is Foundational

Conventional educational technology presumes connectivity: an application in the hand, a model in the cloud. In the Philippine context this presumption fails at the household level.

Fixed household internet penetration in the Philippines remains below one in five households. The dominant mode of access is prepaid mobile data purchased in small denominations, under which recurring data expenditure competes directly with household essentials; coverage degrades rapidly outside urban centres and is intermittent even within them. An AI tutor that requires a server round-trip for every interaction is therefore, for the majority of Filipino learners, a tutor that functions occasionally, imposes a marginal cost on each use, and fails precisely in the environments where it is most needed.

What Filipino households do reliably possess is the device itself. Low-cost Android smartphones — frequently secondhand, typically equipped with four to eight gigabytes of memory — are broadly present even in unconnected homes. Hiraia's founding premise is that the device, rather than the network, is the viable point of delivery: if the entire tutor resides on the telephone, then a single download — at a school, a public terminal, or a neighbour's connection — provides a child with unlimited subsequent use at zero marginal cost, in the household where study actually occurs.

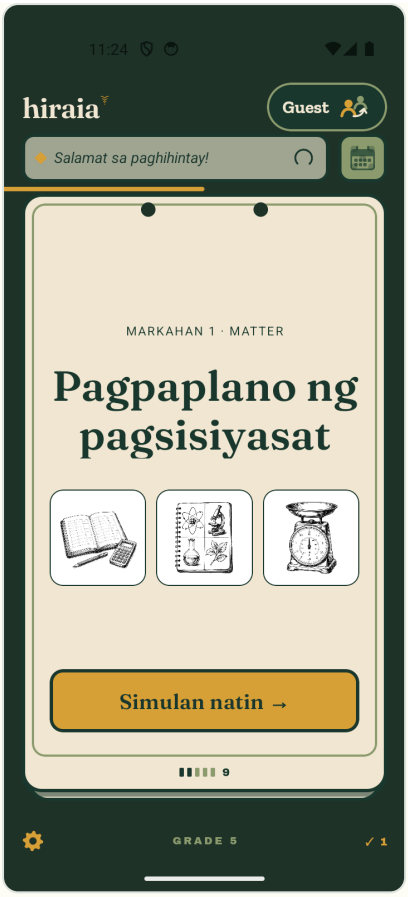
Hiraia is accordingly offline-first in the strict sense. Inference is performed locally; the knowledge bank, question bank, imagery, and speech synthesis are all resident on the device. Following the initial download, connectivity is never required. Updates are offered opportunistically when a connection happens to exist and are never a precondition of use.

3. The System at Release 0.4.11

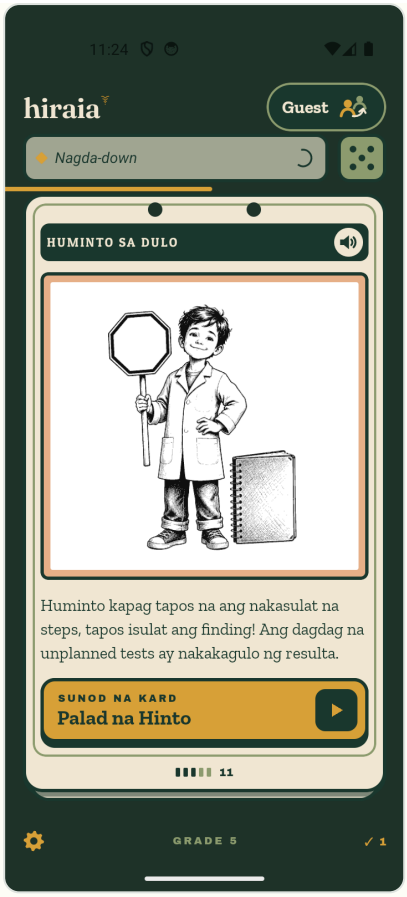
As of release 0.4.11, Hiraia is distributed as a single Android application package of approximately four hundred megabytes comprising:

- **Hiraia-2B**, the project's Filipino-language large language model (Section 4.2), executed on-device, which composes single, self-contained answer cards in Tagalog, Cebuano, or English in response to learner questions;
- a **grounded knowledge bank** of more than twenty-one thousand curriculum-aligned science facts in three languages, with a hybrid retrieval layer that constrains the model to answer from vetted content;
- a **practice bank of 27,783 multiple-choice questions**, likewise trilingual, interleaved into the learning experience;
- an **illustrated science feed** of nearly nineteen thousand engraving-style images paired with curriculum-tagged factual items;
- **neural text-to-speech**: Tagalog and English voices developed by the project, rendering any card aloud with a synchronised reading guide, in real time, on the central processor of an entry-level device; and
- a **language-identification layer** that routes Tagalog, Cebuano, and English input with measured accuracy between 94 and 99 percent.

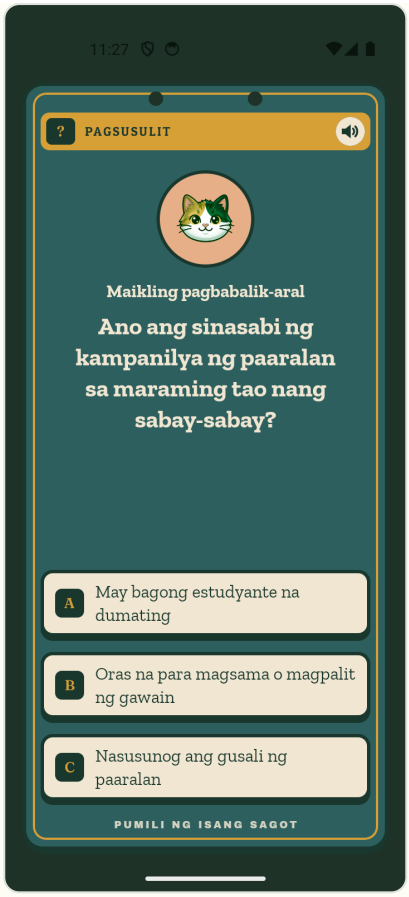
The application carries no advertising, collects no payments, requires no account, and transmits no user data as a condition of use. Hiraia is a not-for-profit undertaking.



The curriculum feed: a Quarter-1 topic card on planning a scientific investigation.



An illustrated fact card in Tagalog, with the read-aloud control in the title band.



A quiz interjection: retrieval practice gates progress through the deck.

4. Technical Architecture

4.1 On-device inference: the QVAC SDK

The system's central technical commitment — a large language model with no server dependency — is realised through the **QVAC SDK from Tether**, which provides the on-device runtime under which the project's model is loaded and executed on the telephone's CPU. The build is configured deliberately for the realities of the target hardware: GPU compute paths are excluded, as graphics drivers on entry-level Android devices are a documented locus of instability, and inference is profiled directly on the chipsets prevalent in the Philippine secondhand market, including the Qualcomm Snapdragon 685 and MediaTek Helio G85. Predictable CPU inference on the devices families actually own was preferred, on measurement, over nominally faster but unreliable acceleration.

4.2 Hiraia-2B: a Filipino-language model trained by the project

Off-the-shelf open models were evaluated at the outset and found inadequate for the project's languages: base-model Tagalog output was substandard and Cebuano weaker still. Rather than accept this ceiling, the project undertook its own model development programme, which we

summarise here because it is central both to the system's quality and to the project's technical standing.

Corpus construction. The project assembled its own Filipino-language pretraining corpus — to our knowledge among the largest curated for the purpose — by combining the major open multilingual web collections on Hugging Face with Philippine educational sources gathered directly. Web-scale constituents were pulled with revision-pinned manifests for reproducibility: the Filipino and Cebuano partitions of FineWeb-2, MADLAD-400 (including its larger "noisy" Filipino split, recovered and re-filtered by the project), and CulturaX, supplemented by the Tagalog Wikipedia (49,170 articles; the machine-generated Cebuano Wikipedia was deliberately excluded), openly licensed children's books from the Bloom Library, and document-level OPUS collections. To these the project added a purpose-built harvest of Department of Education learning materials: a crawler developed for the task traversed the DepEd Learning Resource Portal and extracted text from approximately twenty-six thousand publicly posted PDF learning resources, of which some ninety-seven percent proved text-bearing. All constituents passed a uniform cleaning pipeline — language identification, repetition and quality gates, and exact plus near-duplicate removal across sources, which eliminated fourteen percent of the raw Tagalog pool — with held-out evaluation sets carved from the final pools and verified leak-free against the training mix.

The training mix. The resulting mix comprises 17.9 million Tagalog documents yielding approximately 4.3 billion Tagalog tokens per epoch — 17.2 billion Tagalog tokens seen across training — alongside an upsampled Cebuano slice and English and Chinese anchor material retained to protect the base model's general competence, an anchor ratio validated empirically in a probe run before the full training was committed:

Slice	Tokens per epoch	Tokens trained	Share
Tagalog	4.30 B	17.19 B	69.4%
Cebuano (8 effective epochs)	0.35 B	1.39 B	5.6%
English anchor	1.03 B	4.12 B	16.6%
Chinese anchor	0.52 B	2.08 B	8.4%
Total	**6.19 B**	**24.78 B**	

Training. Upon this corpus the project performed full-parameter continued pretraining of a Qwen3.5-2B base — approximately 5,900 optimiser steps at a 4.19-million-token global batch on multi-GPU datacentre hardware — in effect re-founding the model's Filipino competence on the project's own data. This was followed by supervised fine-tuning on instruction data authored and quality-gated by the project, with grounding-faithfulness objectives specific to the tutoring task. Validation against held-out Filipino text showed perplexity reductions on the order of sixty percent for both Tagalog and Cebuano relative to the base model, and the resulting model — designated **Hiraia-2B** — supersedes all earlier model lines in the shipping application. The model is quantised

to four-bit weights for on-device execution, where the two-billion-parameter scale represents the measured ceiling of acceptable latency on target hardware.

Task design. Hiraia-2B is deployed not as an open-ended conversational agent but as a **single-turn card writer**: given a learner's question and retrieved reference facts, it composes one bounded, age-appropriate answer card. Open dialogue was removed by design. It enlarges the surface for hallucination and unsafe drift, whereas a single grounded card is verifiable, cacheable, and pedagogically honest. Retrieval over the knowledge bank is hybrid — distinctive lexical terms in combination with quantised multilingual sentence embeddings — and where the bank contains no adequate grounding, the model is trained to decline rather than improvise. Throughout the project a single ordering principle applies: where a trade-off is forced, factual accuracy ranks above linguistic fluency.

4.3 Speech synthesis on commodity hardware

Read-aloud capability is essential for a learner population in which emergent readers predominate, and it must function without a cloud speech service. The project's two voices were fine-tuned from Meta's MMS text-to-speech models on a recorded corpus produced for the purpose, then re-architected: the computationally dominant vocoder was replaced with a multi-stream inverse short-time-Fourier-transform design, warm-started from the fine-tuned weights, reducing decoder cost severalfold and bringing synthesis to real time on a Helio G85-class CPU. The voices ship as 55-megabyte half-precision ONNX models with a character-level tokeniser compact enough to implement in application code — no phonemiser and no auxiliary language data are required. Attribution to Meta's MMS work is carried within the application.

4.4 Quality assurance

Two evaluation instruments gate every release. A **regression harness** boots the identical model files that the application ships and executes retrieval stress tests and behavioural assertions; it must pass in full before any build proceeds to a device. A separate **capability benchmark** of approximately 140 probes, scored by an independent model-based judge against a published rubric, is deliberately seeded with items the current system fails, so that improvement is measurable rather than asserted. Releases additionally undergo adversarial role-play evaluation — safety-sensitive questions posed as children actually pose them — prior to shipping.

5. Pedagogical Design

Instruction meets the learner's actual level. Hiraia's default register targets approximately a fifth-grade reading level irrespective of the learner's nominal grade — a deliberate response to the learning-poverty evidence. Content is tagged to MATATAG curriculum competencies, and the feed weights material associated with the current school-year quarter, so that the system shadows the classroom calendar.

Mother tongue first. Instruction defaults to Tagalog, with Cebuano and English as co-equal options; a learner may ask in any of the three and is answered in kind. Scientific vocabulary is introduced bilingually, in recognition that national assessment and upper-level instruction are conducted substantially in English.

Short units, tight loops. The system's atomic surfaces — the answer card, the factual card, and the quiz item — are each self-contained and require under a minute. Interleaved multiple-choice practice operationalises retrieval practice, among the most robustly evidenced techniques in the learning sciences; previously seen items decay in frequency without being retired, approximating spaced repetition. The feed format deliberately adopts the interaction pattern of short-form media, redirecting an attention habit children already possess toward curriculum-bearing material.

Dual coding and voice. Factual items pair text with imagery, and every card can be spoken aloud with a synchronised reading guide — supporting struggling readers, auditory learners, and the shared, sibling-mediated device use typical of Filipino households.

Epistemic honesty as pedagogy. A tutor that fabricates fluently teaches children that fluent fabrication is what knowledge sounds like. The system's grounding discipline, its willingness to abstain, and its bounded single-turn format are therefore pedagogical commitments as much as safety measures: at the scale of a single card, the product models intellectual honesty.

First-run language selection: Tagalog, English, and Cebuano as co-equal options.

Grade selection in Filipino, with grade five preselected.

6. Status and Direction

Release 0.4.11 is publicly and freely downloadable with the full feature set described above. Active development includes a Cebuano voice, for which the recording corpus is prepared; continued scaling of the project's Filipino pretraining programme; content-pack delivery so that the fact and question banks can grow without full reinstallation; and progressive expansion toward complete MATATAG coverage across grade levels.

The project's thesis is unchanged: the Philippines does not lack curious children; it lacks instructional attention per child, and the connectivity to rent substitutes. The devices are already in the households. Hiraia's purpose is to make each of them the most patient science teacher a child has ever had — without cost, in the child's own language, and with the internet switched off.

Hiraia is a not-for-profit project; its primary proponent is Luis Buenaventura. On-device inference is powered by the QVAC SDK from Tether. Hiraia-2B derives from continued pretraining of the

Qwen3.5-2B base (Apache 2.0) on the project's own Filipino corpus; text-to-speech voices are fine-tuned from Meta's MMS models (CC BY-NC, credited in-application). PISA figures: OECD, PISA 2022 and PISA 2025 (results released September 2026). Learning-poverty figure: World Bank, 2022 estimate.