ISSN 3111-4296

Canadian Journal of Artificial Intelligence for Learning and Engineering Innovation

CJAILEI | Published by CuspideScience

Vol. 1 No. 1 (2026)

Special Issue — AEEF 2026

AI, Education, Environment and Future Mobility: Machine Learning Applications for Sustainable Intelligent Systems

CONFERENCE PROCEEDINGS

Peer-Reviewed Contributions from the African Engineering Education Forum 2026 (AEEF 2026)



CuspideScience
Montréal, Québec, Canada
2026

Publication Information

Journal	Canadian Journal of Artificial Intelligence for Learning and Engineering Innovation (CJAILEI)
Issue	Volume 1, Number 1 (2026)
Special Issue	AEEF 2026
Proceedings title	*AI, Education, Environment and Future Mobility: Machine Learning Applications for Sustainable Intelligent Systems*
ISSN	3111-4296
Publisher	CuspideScience – Montréal, Québec, Canada
Publication medium	Online, open access
Language	English
Journal frequency	Semiannual
Conference	African Engineering Education Forum 2026 (AEEF 2026)
Venue and dates	Faculty of Science, University of Douala, Douala, Cameroon – June 8–11, 2026
Special Issue URL	https://cuspidescience.com/index.php/cjailei/SI_AEEF2026
Journal URL	https://cuspidescience.com/index.php/cjailei
Editorial contact	cjailei@cuspidescience.com

Copyright and Licensing

© 2026 CuspideScience. Authors retain copyright in their individual articles. Published articles are distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0), unless otherwise stated on the corresponding article landing page.

The views expressed in individual contributions are those of the authors and do not necessarily represent the views of CJAILEI, CuspideScience, AEEF 2026 organizers, guest editors, or affiliated institutions.

Recommended Citation and Identifiers

Recommended citation for the complete proceedings

Canadian Journal of Artificial Intelligence for Learning and Engineering Innovation. (2026). *AI, Education, Environment and Future Mobility: Machine Learning Applications for Sustainable Intelligent Systems*. Vol. 1, No. 1. Special Issue – AEEF 2026. CuspideScience, Montréal, Québec, Canada.

Persistent identifiers

- **ISSN** 3111-4296
- **Publisher DOI prefix:** 10.65269
- Each research article retains its own permanent DOI. Assignment from Forthcoming to this final issue does not change the article DOI.

Open-access statement

CJAILEI provides immediate open access to its published content. Articles are freely available online to support broad dissemination, accessibility, and responsible reuse of scholarly knowledge.

How to cite individual articles

For individual papers, cite the final volume, issue, page range, and DOI provided in this issue and on the corresponding OJS article landing page.

Journal Editorial Leadership

Editor-in-Chief
Prof. Justin Moskolaï Ngossaha
University of Douala, Cameroon

Founder, Publisher and Managing Editor
Dr. Bappa Muktar
Université du Québec en Outaouais (UQO), Canada

Associate Editors
Dr. Vincent Fono – Université du Québec en Outaouais, Canada
Dr. Adama Nouboukpo – Université du Québec en Outaouais, Canada
Dr. Zongo Meyo – Concordia University, Canada

Managing Editor
Ms. Bekrar Marwa

For the current and complete editorial-board listing, consult the CJAILEI website.

Special Issue Editorial Team

Lead Guest Editors

Prof. Funso Falade
Professor of Civil Engineering, University of Lagos, Nigeria; President, African Engineering Education Association (AEEA)

Prof. Justin Moskolaï Ngossaha
Faculty of Sciences, University of Douala, Cameroon; General Co-Chair, AEEF 2026

Prof. Raymond Houe Ngouna
Tarbes University of Technology, France; General Co-Chair, AEEF 2026; Vice-President for Central Africa, AEEA

Prof. Marie Joseph Ntamak Nida
Dean, Faculty of Sciences, University of Douala, Cameroon; General Co-Chair, AEEF 2026

Dr. Bappa Muktar
Université du Québec en Outaouais (UQO), Canada

Dr. Adama Nouboukpo
Université du Québec en Outaouais (UQO), Canada

Guest Editors

Prof. Ado Adamou Abba Ari
University of Garoua, Cameroon

Prof. Franklin Tchakounte
University of Ngaoundéré, Cameroon

Dr. Igor Tchappi
Research Associate, University of Luxembourg, Luxembourg

About AEEF 2026

The African Engineering Education Forum 2026 (AEEF 2026) was held at the Faculty of Science, University of Douala, Douala, Cameroon, from June 8 to 11, 2026.

Under the theme "Artificial Intelligence (AI) for Engineering Education: Innovation and Solutions for Resource-Constrained Environments," the forum brought together educators, researchers, policymakers, professionals, and industry leaders to discuss the future of engineering education in Africa and beyond.

AEEF 2026 provided a multidisciplinary setting for exchange on artificial intelligence, engineering education, digital learning, smart infrastructure, sustainable systems, mobility, and technologies adapted to resource-constrained environments.

Relationship to this Special Issue

This Special Issue extends the scientific discussions initiated at AEEF 2026. It brings selected, independently peer-reviewed journal contributions into a permanent scholarly record published by CJAILEI. Conference participation or presentation did not, by itself, constitute acceptance for publication in the journal.

Conference website

https://aeef2026.com/

About the Special Issue and Peer Review

About the Special Issue

AI, Education, Environment and Future Mobility: Machine Learning Applications for Sustainable Intelligent Systems brings together research at the intersection of artificial intelligence, engineering education, digital learning, intelligent systems, smart infrastructure, sustainable innovation, and future mobility.

The issue emphasizes educational, technological, industrial, and societal relevance, with particular attention to emerging economies and resource-constrained environments. Its scope includes AI for engineering education, learning analytics, intelligent tutoring systems, generative AI, digital twins, smart mobility, telecommunications, IoT and edge intelligence, cybersecurity, data science for urban systems, frugal AI, and responsible and sustainable AI.

Peer Review and Editorial Process

All research contributions included in this Special Issue were subject to editorial screening and the journal's double-anonymous peer-review process. Research manuscripts are evaluated by at least two independent reviewers with relevant expertise. Additional rounds may be conducted when substantial revision is required or reviewer recommendations diverge.

Final publication decisions are made by the assigned editor under the oversight of the Editor-in-Chief and in accordance with CJAILEI policies on editorial independence, conflicts of interest, research integrity, and publication ethics.

Preface

Artificial intelligence is reshaping both the practice of engineering and the way engineers are educated. This transformation is especially consequential in contexts where connectivity, infrastructure, educational resources, and institutional capacity vary widely. The scientific discussions initiated at AEEF 2026 highlighted the need to connect technological innovation with practical educational and societal needs.

This Special Issue brings together peer-reviewed contributions that illustrate this intersection. The papers address curriculum adaptation, generative-AI tutoring in low-connectivity settings, aerospace engineering education, the role of artificial intelligence in strengthening engineering capacity and prosperity, and applied machine-learning systems with implications for intelligent and sustainable environments.

Taken together, the contributions show that artificial intelligence should not be treated as a single technology or educational tool. Its value depends on context, human oversight, infrastructure, affordability, data quality, and alignment with real educational and engineering needs. The issue therefore seeks to connect methodological advances with responsible implementation and long-term capacity building.

We thank the authors for their contributions, the reviewers for their careful and constructive evaluations, the guest editors for their scientific stewardship, and the AEEF 2026 community for the discussions that helped shape this Special Issue.

Special Issue Editors, 2026

Contents

Front Matter

Publication Information i
Recommended Citation and Identifiers ii
Journal Editorial Leadership iii
Special Issue Editorial Team iv
About AEEF 2026 v
About the Special Issue and Peer Review vi
Preface vii
Contents viii

Research Articles

1. Continental Prosperity in Africa Through Engineering Education: The Role and Contribution of Artificial Intelligence
Funso Falade
DOI: 10.65269/k8k2fm46 Pages: 1–10

2. Self-evolving engineering curricula: A Reinforcement Learning architecture to align academic training with industry demands in developing countries
Leslie Erika Kouamouo Ndangang; Justin Moskolaï Ngossaha; Adolphe Ayissi Etémé
DOI: 10.65269/993p4h33 Pages: 11–26

3. Malloum: An SMS-Based Generative AI Tutoring System for Contextualized Learning in Low-Connectivity Environments
Isaac Touza
DOI: 10.65269/5p09zq79 Pages: 27–50

4. Agile Aerospace Engineering Education: GenAI to Navigate Resource Constraints
Piwai Chikasha; Kemlall Ramdass
DOI: 10.65269/qcx2b691 Pages: 51–72

5. Hybrid SVM and CNN System for Cough Detection and Health Classification
Fabien Mouomene Moffo; Auguste Noumsi; Joseph Mvogo; Samuel Bowong
DOI: 10.65269/pky9dc27 Pages: 73–88

The research articles are compiled from their archived LaTeX source packages using the CJAILEI article class. The page ranges above are the final continuous article pagination of this issue. The five articles are typeset from their archived LaTeX source packages with continuous numbering across the complete proceedings.

Canadian Journal of Artificial Intelligence for Learning and Engineering Innovation (2026), Vol. **1**, No. **1**
ISSN 3111-4296
https://doi.org/10.65269/k8k2fm46

CuspideScience
At the cutting edge of science

PROCEEDINGS | *African Engineering Education Forum (AEEF 2026)*

Continental Prosperity in Africa Through Engineering Education: The Role and Contribution of Artificial Intelligence

Funso Falade *

President, African Engineering Education Association (AEEA)

*Corresponding author: president@aeeaonline.org

(Received: 17 June 2026; Revised: 26 August 2026; Accepted: 04 September 2026; Published: 07 September 2026)
(Editor: Dr. Bappa Muktar; Reviewers: Anonymous reviewers)

Abstract

Continental prosperity represents an economically integrated, stable, and developed Africa where citizens enjoy improved living standards and access to productive employment. Engineering education is central to this vision because it develops the technical and problem-solving capacity required to build infrastructure and industrial capability. Artificial intelligence (AI) offers strategic opportunities to enhance engineering teaching, learning, and research, though its potential is contingent on institutional capability and infrastructure readiness. This paper presents a policy-oriented conceptual perspective based on a structured scoping synthesis focused primarily on literature published from 2018 to 2026, supplemented by selected foundational sources. Applying a Context–Capability–Outcome (CCO) theoretical framework, the study synthesizes global and regional evidence, examines key pedagogical applications of AI, evaluates infrastructure and governance limitations using empirical metrics, and presents an actionable implementation roadmap for institutional reform.

Keywords: Engineering Education; Artificial Intelligence; Africa; Continental Prosperity; Context–Capability–Outcome

Abbreviations: AI: Artificial Intelligence; 5G: Fifth-generation mobile network; AEEA: African Engineering Education Association; AEEF: African Engineering Education Forum; AfDB: African Development Bank; AfCFTA: African Continental Free Trade Area; AU: African Union; CAD: Computer-Aided Design; CCO: Context–Capability–Outcome; CRediT: Contributor Roles Taxonomy; DFIs: Development Finance Institutions; ERIC: Education Resources Information Center; GPU: Graphics Processing Unit; IEEE: Institute of Electrical and Electronics Engineers; IITs: Indian Institutes of Technology; LAN: Local Area Network; LMS: Learning Management System; NITs: National Institutes of Technology; RECs: Regional Economic Communities; ROI: Return on Investment; STEAM: Science, Technology, Engineering, Arts and Mathematics; UNESCO: United Nations Educational, Scientific and Cultural Organization

1. Introduction

The vision of continental prosperity across Africa aligned with the African Union's Agenda 2063 depends on sustained economic transformation, poverty reduction, industrial expansion, and regional integration [1]. Engineering education serves as a fundamental catalyst for this transformation by equipping graduates with the scientific, technical, and leadership capabilities needed to address local development challenges [6][7].

However, engineering education systems across the continent face persistent structural constraints, in-

© 2026 by the author. This is an Open Access article, distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/)

cluding severe funding deficits, outdated curricula, inadequate physical laboratories, weak university-industry linkages, and brain drain [2][10]. While global advances in artificial intelligence (AI) offer transformative possibilities for engineering pedagogy and research, AI adoption in African higher education remains uneven and frequently hindered by baseline infrastructure deficits [4].

This paper provides a policy-oriented conceptual framework examining how AI-enabled engineering education can contribute to continental prosperity. The main contributions of this work are:

i. Formulating an integrated Context–Capability–Outcome (CCO) framework tailored to African engineering institutions [9].
ii. Synthesizing international comparative evidence across standardized dimensions.
iii. Evaluating concrete AI pedagogical use cases against technical and institutional readiness [3].
iv. Delivering an actionable, stakeholder-assigned implementation roadmap with quantitative evaluation metrics [12].

1.1 Research Questions

i. How can engineering education contribute to continental prosperity in Africa?
ii. How do structural, infrastructural, and institutional contexts across African regions influence engineering education outcomes?
iii. What specific AI-enabled pedagogical tools and frameworks can strengthen engineering teaching, learning, and research?
iv. What policy actions, institutional capabilities, and governance frameworks are required to achieve responsible and equitable AI integration?

1.2 Research Contribution

This paper makes four substantive contributions to the scholarship on AI-enabled engineering education in Africa. First, it develops a tailored Context–Capability–Outcome (CCO) framework that links infrastructural realities, institutional capabilities, and developmental outcomes, providing a structured lens for analyzing AI adoption in engineering education. Second, it offers a comparative synthesis of global and regional systems, situating African engineering institutions within standardized international benchmarks and highlighting context-specific constraints and opportunities. Third, it evaluates key AI pedagogical applications, including adaptive tutoring, virtual simulation laboratories, generative computer-aided design (CAD) tools, and learning analytics, against Africa's technical, financial, and governance conditions. Finally, it presents a policy-aligned implementation roadmap with measurable indicators, providing governments, universities, accreditation bodies, and development partners with a structured pathway for responsible, equitable, and context-appropriate AI integration.

2. Review Methodology

This paper utilizes a structured qualitative scoping synthesis to evaluate the intersection of AI, engineering education, and economic development.

2.1 Search Strategy and Database Selection

A structured literature search was conducted across Scopus, Institute of Electrical and Electronics Engineers (IEEE) Xplore, Education Resources Information Center (ERIC), Web of Science, and Google Scholar, supplemented by authoritative institutional repositories from the African Union (AU), United Nations Educational, Scientific and Cultural Organization (UNESCO), the World Bank, and the African Development Bank (AfDB) [1][2][3][12].

- **Primary search terms included:** (“engineering education” OR “technical training”) AND (“artificial intelligence” OR “generative AI” OR “educational technology”) AND (“Africa” OR “sub-Saharan Africa” OR “developing nations”) AND (“infrastructure” OR “industrialization” OR “economic development”) [3].
- **Temporal boundary:** The primary scoping search covered literature published from 2018 through 2026 to capture recent developments in artificial intelligence, engineering education, and related policy frameworks. Earlier publications were included selectively when they provided foundational theoretical frameworks or essential historical context. These sources also offered authoritative baseline policy and engineering education evidence needed to contextualize contemporary literature. Eligible sources included English-language peer-reviewed journal articles, conference proceedings, and authoritative institutional policy reports.

2.2 Inclusion and Exclusion Criteria

- **Inclusion:** Peer-reviewed empirical or conceptual studies on AI in higher/engineering education; comparative education policy studies; national/regional development framework reports with direct technical education data [3][4].
- **Exclusion:** Non-peer-reviewed opinion pieces; studies focusing strictly on primary/secondary basic education without higher education linkage; hardware-only technical engineering papers lacking pedagogical or policy application.

2.3 Selection Process and Synthesis

Sources identified through the structured search were screened for relevance to AI-enabled engineering education, institutional capability, infrastructure constraints, and African development policy. Eligible peer-reviewed studies and authoritative institutional reports were then thematically synthesized using the Context–Capability–Outcome framework. Because the study is a policy-oriented conceptual perspective rather than a systematic review or meta-analysis, the synthesis emphasizes conceptual relevance, comparative evidence, and authoritative policy evidence rather than exhaustive coverage of the literature.

3. Conceptual and Theoretical Framework: Context–Capability–Outcome (CCO)

To analyze the relationship between AI tools, educational institutions, and regional prosperity, this paper applies a Context–Capability–Outcome (CCO) model adapted from policy analysis and organizational capability theory [9]. The framework posits that technological tools do not produce developmental outcomes directly; rather, their potential contribution is shaped by institutional capabilities operating within specific infrastructural and policy contexts [9]. The relationships among these three dimensions are summarized in Figure 1.

- **Context (C):** The baseline infrastructural, financial, and policy environment (e.g., electricity stability, broadband access, national educational expenditure, regional trade alignment under the African Continental Free Trade Area (AfCFTA)) [1].
- **Capability (Ca):** The internal institutional capacity to absorb and deploy resources effectively (e.g., faculty digital literacy, virtual laboratory availability, industry co-designed curricula, AI governance frameworks) [5].
- **Outcome (O):** The resulting educational and socio-economic achievements (e.g., graduate technical competence, locally relevant technological innovations, resilient infrastructure delivery, regional economic integration) [2].

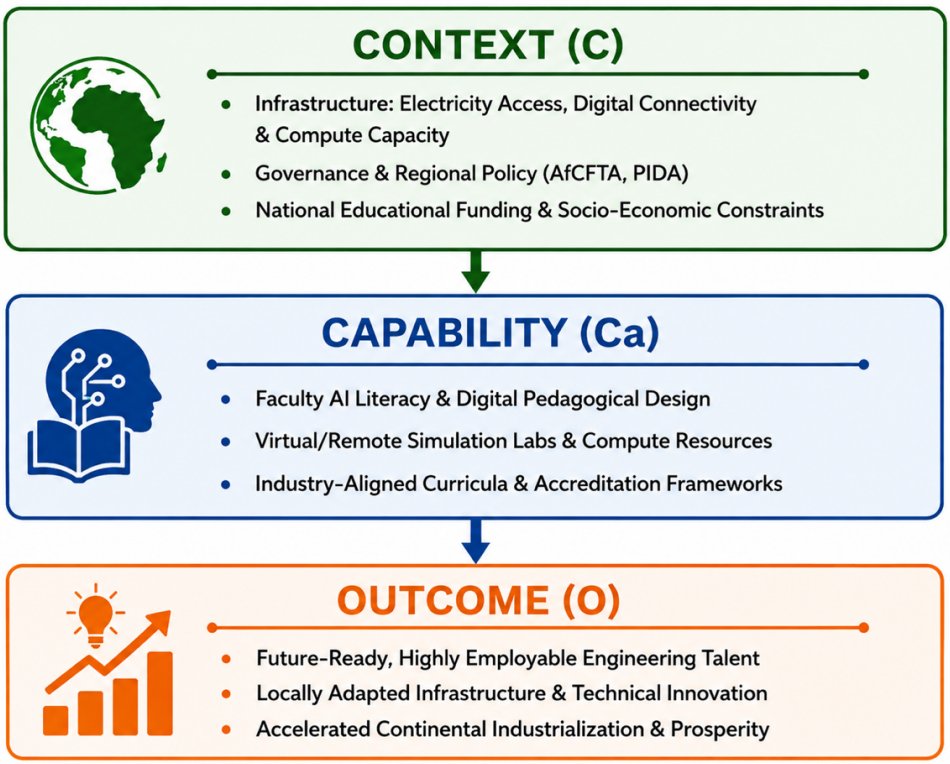


Figure 1. Context–Capability–Outcome (CCO) Framework for AI-Enabled Engineering Education and Continental Prosperity in Africa

The literature-search and selection protocol underlying the synthesis is summarized in Table 1.

Table 1. Literature Search and Selection Protocol for the Structured Scoping Synthesis

Selection Stage	Process & Criteria
Identification	Initial search across Scopus, IEEE Xplore, ERIC, Web of Science, Google Scholar, AU, UNESCO, World Bank, and AfDB.
Screening	Title and abstract screening for relevance to higher engineering education, AI, and policy.
Eligibility	Full-text evaluation against inclusion criteria (excluding secondary basic ed, non-peer-reviewed, and hardware-only papers).
Inclusion	Relevant peer reviewed studies and authoritative policy reports retained for thematic synthesis using the CCO framework.

4. Literature Review and Comparative Analysis

Rather than treating AI as a generic technology, engineering programs must align specific AI capabilities with defined pedagogical objectives and required institutional capacities [3][4]. The principal international and African system-level contrasts are summarized in Table 2.

Table 2. International Comparative Systems Matrix for AI Enabled Engineering Education

System / Region	Funding & Infrastructure	Digital Connectivity	Faculty AI Readiness	Industry-University Linkages	Primary Systemic Limit / Constraint
High-Income (Global)	High / Advanced physical labs	High (Gigabit broadband, fifth-generation (5G) mobile networks)	Moderate to High	Strong / Highly commercialized research and development	Over-reliance on automated design tools; context transferability
China	High (State-directed policy)	High (Integrated smart campuses)	High	Integrated with state-owned enterprises	Centralized curriculum rigidity; high compliance overhead
India	Moderate-High (Indian Institutes of Technology (IITs)/National Institutes of Technology (NITs) scale)	Rapidly expanding urban fiber	Moderate	Moderate to Strong	Wide quality variance between elite and tier-3 institutions
Latin America	Moderate / Uneven	Moderate	Low to Moderate	Moderate	High regional inequality; volatile public funding
South Africa	High (Relative to Sub-Sahara)	High urban / Low rural	Moderate	Strong (Mining, energy, civil)	Socio-economic disparities in access and baseline preparation
Nigeria	Low (Public institutions)	Poor connectivity / Frequent outages	Low	Weak	Severe infrastructure, power, and funding deficits
Kenya	Moderate	High mobile connectivity / Fiber	Moderate	Emerging tech hub ecosystem	High reliance on donor funding; faculty retention challenges

4.1 Regional Insights

The comparative evidence highlights that while high-income systems leverage AI to optimize existing high-bandwidth operations [12], emerging regions like India and Latin America face institutional variance. Within Africa, institutions in South Africa, Nigeria, and Kenya present distinct structural profiles, demonstrating that policy interventions must be contextualized rather than treated as a single uniform model [12].

5. Core Applications: AI-Enabled Engineering Education

Rather than treating AI as a generic technology, engineering programs must align specific AI capabilities with defined pedagogical objectives and required institutional capacities [3][4]. The corresponding pedagogical applications, required capacities, expected outcomes, and implementation risks are presented in Table 3.

Table 3. AI Enabled Engineering Education Pedagogical Use Cases, Institutional Capabilities, Outcomes, and Risks

AI Application Category	Core Pedagogical Objective	Required Institutional Capability	Expected Educational Outcome	Key Risk / Implementation Limitation
Adaptive Tutoring Systems	Personalize instruction speed and remediate foundational knowledge gaps	learning management system (LMS) platform; reliable internet; low-latency edge servers	Reduced failure rates in gateway engineering math/physics	Algorithmic bias when applied to non-native English instruction
Virtual & Remote Simulation Labs	Provide practical experiment access without physical hardware cost	High-speed local servers; cloud-hosted virtual reality environments	Expanded access to hands-on thermodynamic/fluid mechanics modules	Lack of tactile physical lab experience and hardware assembly skills
Generative CAD & Design Support	Accelerate design iterations and multi-objective engineering optimization	Compute/graphics processing unit (GPU) resources; software licenses; updated curricula	Enhanced student capacity for complex systems engineering	Over-reliance on automation without mastering first-principles physics
Learning Analytics & Retention	Identify at-risk students early via engagement and assessment metrics	Integrated student data management systems; staff training	Improved retention, graduation rates, and targeted academic support	Data privacy breaches and intrusive student surveillance

5.1 Engineering Pedagogy for Infrastructure and Industrialization

AI tools must directly support training in critical developmental domains [8]:

- **Energy Systems:** Training students using AI grid-optimization models for hybrid renewable/mini-grid deployments.
- **Transportation & Logistics:** Utilizing machine learning simulations to design climate-resilient road corridors and smart transit planning.
- **Water Resource Engineering:** Implementing AI-driven hydrological modeling for urban water distribution and agricultural irrigation efficiency.
- **Industrialization & Manufacturing:** Educating engineers in predictive maintenance, computer-integrated manufacturing, and agro-processing optimization.

6. Quantified AI Limitations and Risk Analysis

While AI presents major opportunities, its deployment in African engineering institutions is constrained by baseline infrastructure and structural limitations [2][11][12].

6.1 Empirical Infrastructure Constraints

- **Electrification Deficits:** Limited and uneven electricity access across Sub-Saharan Africa remains a significant constraint on the deployment of compute-intensive AI infrastructure, particularly for institutions in underserved and rural areas. These electricity deficits can limit universities' capacity to operate AI-enabled computing infrastructure and digitally supported engineering laboratories [2][12][13].
- **Internet Penetration & Bandwidth:** Uneven broadband availability, network reliability, and

the relative cost of internet access constrain equitable access to cloud-based AI tools, virtual laboratories, and other digitally intensive engineering learning environments. Internet use in Sub-Saharan Africa was estimated at approximately 38% in 2024, illustrating the scale of the region's continuing connectivity challenge [12][14].

- **Compute Infrastructure:** Limited locally hosted advanced computing infrastructure can restrict the deployment and scaling of computationally intensive AI applications within African engineering institutions. Africa accounts for approximately 0.6% of global data center capacity, highlighting the continent's limited locally hosted digital infrastructure and potential dependence on externally hosted cloud and computing services [3][12][15].

6.2 Responsible AI, Equity, and Governance Risks

1. **Algorithmic Bias & Contextual Misalignment:** Large language and design models are predominantly trained on Western and East Asian datasets, often failing to account for African standards, local building materials, or regional climate conditions.
2. **Data Sovereignty & Privacy:** Student learning metrics and institutional intellectual property are frequently processed through external cloud infrastructure lacking regional data protection compliance.
3. **Gender and Rural Disparities:** Digital divide constraints disproportionately impact female students, rural universities, and underfunded public institutions, risking widened socio-economic inequality if AI access is not deliberately subsidized.
4. **Pedagogical Degradation & Academic Integrity:** Over-reliance on generative AI without human-in-the-loop oversight risks diminishing fundamental analytical rigor and ethical responsibility among engineering graduates.

7. Actionable Implementation Framework and Policy Roadmap

To operationalize AI-enabled engineering education reform, actions are structured across short-, medium-, and long-term horizons with assigned leads, prerequisites, and measurable evaluation indicators [1][5][12]. The phased roadmap is presented in Table 4.

Table 4. Actionable Implementation Roadmap for AI Enabled Engineering Education in Africa

Horizon / Phase	Core Strategic Action	Responsible Lead	Required Prerequisites	Proposed Indicator / Target
Short-Term (Years 1–2)	Launch faculty AI literacy workshops, digital pedagogy training, and offline-first edge compute pilots	University Leadership & Engineering Deans	Baseline campus local area network (LAN) connectivity	% of engineering faculty certified in AI pedagogical integration; 30% reduction in virtual laboratory setup costs
Medium-Term (Years 3–5)	Overhaul engineering curricula to include AI, data science, and ethics; deploy industry advisory boards	National Accreditation Bodies & Industry Boards	Short-term faculty readiness	Updated national curriculum standards active; 25% increase in graduate industry placement
Long-Term (Years 5+)	Establish national compute hubs, regional research networks, and harmonized cross-border accreditation	African Union (AU), regional economic communities (RECs), & development finance institutions (DFIs)	Medium-term institutional alignment	Adoption and implementation of the AABET Accreditation and Quality Assurance Framework across participating African engineering institutions [16]; 50% increase in intra-African engineer mobility and joint patents

8. Limitations of the Study

As a policy-oriented conceptual synthesis, this study relies on secondary evidence and a structured qualitative scoping approach without primary empirical field testing. A further limitation is the uneven availability of standardized regional data across African higher-education systems, which restricts the ability to conduct uniform quantitative comparisons and limits generalizability across all subregions [12].

9. Policy Implications

Building on the phased actions summarized in Table 4, the principal policy implication is the need for coordinated governance across institutional, national, and regional levels rather than a repetition of the roadmap's individual actions. National governments should establish the enabling infrastructure, financing, and data-protection environment; accreditation bodies should translate AI literacy, ethics, and industry relevance into program standards; and the African Union together with regional economic communities should coordinate shared compute capacity, research networks, and cross-border recognition. Equity safeguards should be embedded in these measures so that rural institutions, women, and under-resourced public universities are not excluded. This multilevel coordination is the key policy condition for translating institution-level AI adoption into broader Agenda 2063 outcomes.

10. Conclusions and Research Agenda

1. **Engineering Education & Prosperity:** Engineering education provides the core human capital needed to design, construct, and maintain infrastructure, drive industrialization, and manage energy, water, and transport networks under regional frameworks like AfCFTA [1][2].
2. **Structural & Contextual Limits:** Outdated curricula, funding shortfalls, limited faculty AI readiness, uneven electricity access, digital connectivity constraints, and limited computing infrastructure present significant barriers that vary across African subregions [2][12].
3. **AI Pedagogical Contribution:** When supported by adequate institutional capabilities, AI tools, such as adaptive tutoring systems, virtual simulation laboratories, generative CAD platforms, and learning analytics, have the potential to enhance educational access, conceptual learning, design and research efficiency, and targeted student support [3][4].
4. **Policy & Governance Requirements:** Achieving equitable AI deployment requires policy interventions, localized data governance, subsidized digital access, faculty upskilling, and harmonized regional accreditation [1][12].

10.1 Future Research Agenda

To move beyond conceptual synthesis, future research must focus on:

- **Empirical Implementation Trials:** Conducting multi-institution-controlled trials comparing learning gains in virtual vs. physical engineering laboratories.
- **Cost-Effectiveness Analysis:** Evaluating the long-term return on investment (ROI) of localized edge computing versus cloud-hosted AI educational tools in resource-constrained contexts.
- **Longitudinal Employability Tracking:** Examining the relationship between AI-integrated engineering curricula, graduate career trajectories, and industry-led innovation in Africa.

Author contributions

Funso Falade: Conceptualization; Methodology; Investigation; Writing - original draft; Writing - review and editing; Project administration. The author reviewed and approved the final manuscript

and agrees to be accountable for all aspects of the work.

Funding statement

The author received no specific funding for this work.

Conflict of interest statement

The author declares no conflict of interest.

Open data statement

No new datasets were generated or analyzed in this study. The article is a policy-oriented conceptual synthesis based on publicly available published literature and institutional reports cited in the manuscript.

Reproducibility statement

The structured qualitative scoping methodology, including databases and institutional repositories searched, search terms, temporal boundaries, inclusion and exclusion criteria, and synthesis procedure, is reported in Section 2 and Table 1. No proprietary dataset or computational code was generated for this study.

References

[1] African Union. *Agenda 2063: The Africa We Want*. Tech. rep. Addis Ababa: African Union, 2015. URL: https://au.int/en/agenda2063/overview.

[2] Manuel Zymelman, ed. *Assessing Engineering Education in Sub-Saharan Africa*. World Bank Technical Paper 197. Washington, DC: World Bank, 1993. ISBN: 9780821323717. URL: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/910761468767432842/assessing-engineering-education-in-sub-saharan-africa.

[3] Fengchun Miao, Wayne Holmes, Ronghuai Huang, and Hui Zhang. *AI and Education: Guidance for Policy-Makers*. Paris: UNESCO, 2021. URL: https://www.unesco.org/en/articles/ai-and-education-guidance-policy-makers.

[4] International Development Research Centre (IDRC). *Artificial Intelligence for Development*. International Development Research Centre. URL: https://idrc-crdi.ca/en/initiative/artificial-intelligence-development (visited on 09/04/2026).

[5] ABET. *Criteria for Accrediting Engineering Programs, 2024–2025*. ABET. Baltimore, MD, 2024. URL: https://www.abet.org/accreditation/accreditation-criteria/.

[6] F. Falade. "Challenges in Engineering Education in Africa". In: *2nd Global Engineering Deans' Conference*. Paper presented February 24–25, 2008. Berlin, Germany, Feb. 2008. URL: https://ir.unilag.edu.ng/items/61e44cec-a478-4e7f-84c2-61c53c3a3aae.

[7] Funso Falade. "Engineering Education in Africa: Challenges and Mitigation Measures". In: *IN4OBE Global Virtual Summit 2022: Transforming Education and Empowering Learners (IN4OBE 2022)*. Posted on SSRN in 2023. 2022. DOI: 10.2139/ssrn.4484486.

[8] F. Falade. *Revolutionizing Engineering Education in Africa: Embracing Digitalization, Artificial Intelligence and Sustainable Development—Issues and Prospects*. Keynote talk, African Engineering Education Forum (AEEF 2025), Cadi Ayyad University, Marrakech, Morocco. April 18, 2025. 2025. URL: https://africaeef.org/AEEF%202025%20programme.pdf.

[9] Michael W. Bauer and Christoph Knill. "A Conceptual Framework for the Comparative Analysis of Policy Change: Measurement, Explanation and Strategies of Policy Dismantling". In: *Journal of Comparative Policy Analysis: Research and Practice* 16.1 (2014), pp. 28–44. DOI: 10.1080/13876988.2014.885186.

[10] Themba Mashiyane, Smith Salifu, Olugbenga Ogunbiyi, and Adeola Oketola. "Challenges, Strategies and Recommendations for Reengineering Engineering Education in Sub-Saharan Africa: A Review". In: *Science, Engineering and Technology* 4.1 (2024), pp. 137–153. DOI: 10.54327/set2024/v4.i1.102.

[11] Abdul Feroz Maluleke. "AI Adoption in African Higher Education: A Systematic Review of Benefits and Ethical Implications". In: *Interdisciplinary Journal of Education Research* 7.2 (2025), a05. DOI: 10.38140/ijer-2025.vol7.2.05.

[12] Programme for Infrastructure Development in Africa (PIDA). *2024 Annual Report for the Programme for Infrastructure Development in Africa (PIDA)*. Tech. rep. Programme for Infrastructure Development in Africa, 2025. URL: https://www.au-pida.org/download/2024-annual-report-report-for-the-programme-for-infrastructure-development-in-africa-pida/.

[13] World Bank. *Access to Electricity (% of Population) – Sub-Saharan Africa*. World Development Indicators, SDG 7.1.1 Electrification Dataset. URL: https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS?locations=ZG (visited on 09/04/2026).

[14] International Telecommunication Union. *The State of Broadband in Africa 2025*. Tech. rep. Geneva: Broadband Commission for Sustainable Development / International Telecommunication Union, 2025. URL: https://www.itu.int/pub/S-POL-BROADBAND.32-2025.

[15] African Data Centres Association. *Data Centres in Africa 2026: The Economic Report*. Tech. rep. African Data Centres Association, 2026. URL: https://africadca.org/en/data-centres-in-africa-2026-the-economic-report.

[16] AABET. *AABET Accreditation and Quality Assurance Framework*. AABET. URL: https://aabet.africa.

Canadian Journal of Artificial Intelligence for Learning and Engineering Innovation (2026), Vol. **1**, No. **1**
ISSN 3111-4296
https://doi.org/10.65269/993p4h33

CuspideScience
At the cutting edge of science

PROCEEDINGS | ***African Engineering Education Forum (AEEF 2026)***

Self-evolving engineering curricula: A Reinforcement Learning architecture to align academic training with industry demands in developing countries

Leslie Erika Kouamouo Ndangang [1], Justin Moskolaï Ngossaha [*,1] and Adolphe Ayissi Etémé [2]

[1]Department of Mathematics and Computer Science, Faculty of Science, University of Douala, PO Box 24157, Douala, Cameroon.
[2]IUT of Ngaoundéré, University of Ngaoundéré, PO Box 455, Ngaoundéré, Cameroon.

*Corresponding author: kouamouoerika@gmail.com

(Received: 18 June 2026; Revised: 06 August 2026; Accepted: 13 August 2026; Published: 17 August 2026)
(Editor: Dr. Bappa Muktar; Reviewers: Anonymous reviewers)

Abstract

Engineering education in developing countries faces a widening gap between academic curricula and labor market needs, as traditional revision processes are too slow to keep pace with rapid technological change. This paper proposes a conceptual, self-evolving curricular architecture based on RL, formalizing academic programs as a Markov Decision Process in which curricular adjustments are actions, academic and employability indicators are state variables, and a multi-criteria reward function balances performance, employability, industrial alignment, and budget constraints. An Explainable AI layer and a human-governed workflow support institutional trust. As a proof-of-concept contribution, no deployment, dataset, or simulation has been carried out; the paper instead formalizes the problem, details the architecture, and outlines a validation protocol for future empirical work.

Keywords: Reinforcement Learning; Self-Evolving Systems; Artificial Intelligence in Education; Employability; Skills-to-Market Alignment

Abbreviations: DC: Developing Countries & RL: Reinforcement Learning & AI: Artificial Intelligence & MDP: Markov Decision Process & XAI: Explainable Artificial Intelligence

1. Introduction

Modern engineering education strives to meet technological skill requirements while fostering innovation and economic growth. In developing countries, this concept integrates critical dimensions such as adaptation to limited resources, equitable access to quality training, and the gradual integration of global technological standards with local industrial realities [1, 2, 3]. Educational ecosystems in these regions are characterized by a rapidly growing student population, underdeveloped training infrastructures, and a marked disconnection from real-time labor market data [4, 5]. These structural constraints hinder the direct application of curricular revision models that have proven effective in more developed contexts. Indeed, academic management in these areas faces major obstacles, including the prevalence of slow bureaucratic processes, the exclusion of emerging skills such as AI or green energy from official databases, and limited inclusion of industrial stakeholders in pedagogical

© 2026 by the authors. This is an Open Access article, distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/)

decision-making [6].

In response to this inertia, a paradigm shift toward more agile systems has become imperative. Consequently, a central question arises: to what extent can the integration of RL support the design of next-generation "self-evolving" curriculum systems specifically tailored to the constraints of developing countries? In other words, how can traditionally static academic governance be transformed into a dynamic mechanism capable of optimizing training pathways based on changing market needs?

The objective of this work is to examine this technological transition by combining a comprehensive literature review with the formal design of a conceptual prototype architecture in a representative context, namely Cameroon. This research aims to argue for the technical feasibility of an integrated approach in which RL enables increased visibility of employability trends and continuous optimization of curricular parameters [7, 8], and to lay out a roadmap from conceptual design toward eventual, empirically validated real-world deployment via a lightweight, interoperable infrastructure [9].

The remainder of this article is organized as follows: Section 2 reviews the state of the art in engineering education and the opportunities offered by AI integration. Section 3 formalizes the curricular-governance problem and details the proposed architecture. Section 4 outlines the envisioned validation protocol, expected contributions, and governance safeguards. Section 5 then discusses the paper's contribution and outlines future research directions, and finally, Section 6 concludes with a summary and recommendations.

2. State of the art

2.1 Sustainable engineering education, employability, and digital opportunities in developing countries

Sustainable engineering education aims to meet technical training needs while minimizing social and economic skill mismatches. In developing countries (DCs), this concept also integrates the management of obsolete curricula, equitable access to digital and pedagogical resources, and the progressive integration of global technological standards with local industrial realities [10, 11, 12, 13, 14].

Academic institutions in DCs face massive student enrollment, undersized training infrastructures, and a lack of reliable data on actual labor-market needs [15, 16, 1, 17]. These constraints complicate the direct application of curricular revision models that have proven successful in developed countries. In particular:

- **Administrative Inertia:** Program update processes are often too slow and bureaucratic to keep pace with rapid technological cycles (e.g., generative AI) [18, 19];
- **Lack of Data Granularity:** Specific skills demanded by the local industrial sector (SMEs, technical informal sector) are frequently absent from official academic frameworks [20];
- **Technical Gap:** Unequal access to cutting-edge tools and the low adoption rate of Learning Analytics solutions hinder the use of artificial intelligence for optimizing training pathways [21].

Initiatives such as the UNESCO framework on AI and education demonstrate that the use of modular and interoperable systems is essential for adapting to local conditions [22, 23, 24]. Learning Analytics and AI-based frameworks have been used to deploy web-accessible predictive analysis tools, thereby facilitating decision-making for training managers and academic leaders [25, 26, 27, 28]. However, several challenges remain, such as the lack of structured occupational data, high digital-infrastructure and computational costs, and the difficulty of integrating RL into rigid pedagogical frameworks. This motivates rethinking curricular engineering in developing countries through lightweight, participatory, and AI-supported digital tools.

2.2 RL and self-evolving systems

Foundations of Reinforcement Learning

Reinforcement Learning enables software agents to refine their behaviors by interacting with changing environments [29]. Foundational work on policy gradients and Q-learning [30, 31] established the mathematical basis for training agents through rewards and penalties [32]. As the field has evolved, innovations such as Proximal Policy Optimization (PPO) [33] and Deep Q-Networks (DQN) [34] have significantly enhanced autonomous decision-making capabilities [35, 36]. A significant advantage of RL lies in its ability to adapt to unforeseen situations, making it a candidate approach for self-programmable AI. However, challenges such as sample inefficiency, high computational requirements, and learning instability remain major obstacles that current research seeks to address via model-based RL or hierarchical RL.

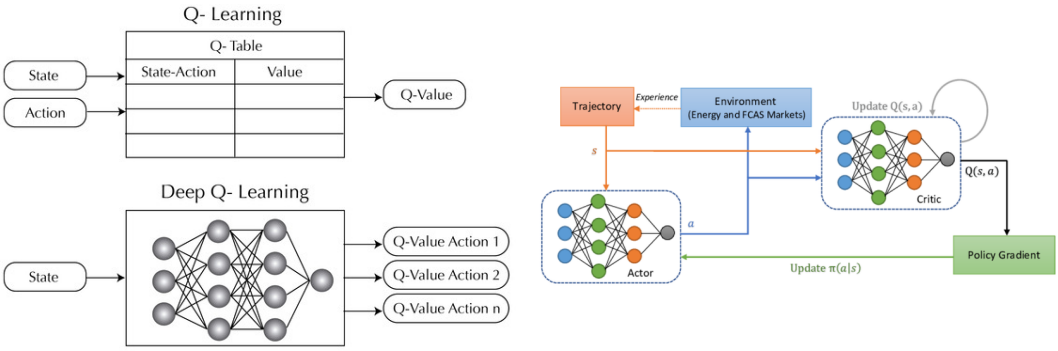


Figure 1. Foundations of Reinforcement Learning Algorithms: DQN and PPO (schematic, adapted for illustration).

Figure 1 summarizes the two RL families referenced throughout this paper.

2.2.1 Application to self-evolving curricula

In this context, a self-evolving curriculum is defined as an *adaptive curriculum decision-support system*: a model-based representation of a training pathway that supports the real-time acquisition, analysis, and human-approved adjustment of teaching modules based on market signals [37]. The term "self-evolving" refers to the continuous update of the underlying model and its recommendations, not to unsupervised institutional change; the human-governance requirement is explicit. This technology is intended to be powered by RL, where the agent learns to optimize a "policy" (the curriculum) by receiving rewards based on success indicators such as placement rates or skill alignment. The envisioned system architecture relies on microservices, facilitating incremental deployment and the integration of heterogeneous data sources through RESTful APIs.

2.2.2 Interoperability and data collection

To ensure seamless interaction between stakeholders (universities, students, employers), semantic ontologies are used to structure educational entities (skills, course units, job profiles) and support complex relational and temporal queries [38, 39]. Self-evolving systems of this kind are designed to integrate data from professional platforms (LinkedIn, Indeed), alumni feedback (crowdsourcing), and public databases. The use of lightweight solutions, tailored for resource-constrained environments, is intended to ensure participatory data collection and a more inclusive alignment between academic training and field realities [40].

2.3 AI and sustainable education

Artificial Intelligence (AI) plays a key role in transforming educational systems through its predictive and prescriptive capabilities. Models such as Recurrent Neural Networks (LSTM), Transformer architectures, and especially Reinforcement Learning (DQN, PPO) are used in the literature to predict learning trajectories, optimize training pathways, and dynamically adjust pedagogical content based on industrial demand [9]. Such tools can, in principle, generate indicators such as predictive employability rates, skill-alignment scores, and program saturation points, and can support "what-if" scenario analysis to test the impact of new subjects before their official deployment [41].

To ensure transparency and faculty buy-in, Explainable AI is also proposed to make automated recommendations understandable to non-technical stakeholders, such as academic directors or certification committees [42]. More broadly, probabilistic and predictive modeling has been applied to other education-related human factors, such as teacher self-efficacy [43], illustrating the broader relevance of data-driven predictive modeling to educational governance, of which the RL-based curricular model proposed here is one example. Finally, recent research explores the use of Large Language Models (LLMs) to automatically generate contextual competency frameworks and anticipate the evolution of local professions.

Table 1. Review of Academic Works on AI, Reinforcement Learning, and Digital Twin Approaches for Engineering Curriculum Alignment (2022-2026)

Authors, Year	Main Topic	Methodology	Tools / Algorithms	Main Limitation
Liu & Song, 2025	Dynamic alignment between education and industry	RL modeling	Reinforcement Learning	Limited to micro-level vocational training tasks
Stasolla et al., 2025	Deep Learning & RL in higher education	Scoping review	DL + RL	Focused on the student micro-level
Satheeskumar et al., 2025	AI emerging skills alignment	Applied AI system	AI-driven alignment	Contextual implementation challenges
Nkrumah et al., 2025	Skill extraction and matching	Comparative study	BERT vs LLM	Strong dependency on textual data
Cousineau, 2025	Explainable AI in education	Doctoral research	XAI frameworks	Implementation complexity
Oeij et al., 2024	Industry 5.0 skills	Conceptual framework	Industry 5.0 taxonomy	Not integrated into an adaptive model
Perisic et al., 2023	Educational interoperability	Conceptual framework	Open interoperability model	Lack of sequential optimization
Sivamayil et al., 2023	RL applications	Systematic review	Q-learning, PPO	Limited curricular applications
Xu et al., 2022	Big Data & employability	Empirical study	Big Data analytics	Static predictive model

This synergy between AI and self-evolving systems suggests a promising direction for addressing curricular obsolescence and professional exclusion in high-growth regions. Beyond simple predictive analysis, the convergence of AI and instructional engineering as argued in [8] could enable a strategic, adaptive, and explainable alignment of training with local socio-economic realities. This direction, aligned with the Industry 5.0 skill framework [20], motivates the architecture proposed in Section 3.

2.4 Global objective and positioning: identified limitations in the literature and architectural justification

Analysis of recent studies highlights that, despite significant advances in the use of AI in education, several limitations persist, particularly in developing countries. Existing solutions often struggle to adapt to contexts characterized by institutional inertia, insufficient digital infrastructure, and fragmented data regarding the local labor market [15, 17, 6]. Furthermore, many approaches remain focused on descriptive or predictive analysis of student performance without proposing a sequential and adaptive curricular optimization mechanism [14, 8].

Moreover, although RL models demonstrate strong potential for dynamic optimization [29], their algorithmic complexity and lack of decision-making transparency still represent major barriers to institutional adoption. The issues of explainability and trust in intelligent systems thus remain central [42].

It is precisely to address these gaps that this research proposes an integrated methodological framework. This framework combines RL-based formalization, an interoperable modular architecture [39], dynamic alignment with industrial skills [20], and XAI mechanisms, with the goal of designing a self-evolving, contextualized, and sustainable curriculum decision-support system.

While RL-based strategies for industry-teaching integration already exist, such as the one proposed in [37], that work targets micro-level vocational-training tasks within a single, highly structured training environment. The present study differs from [37] along three concrete dimensions:

(i) **Decision granularity.** [37] optimizes individual training exercises within one course; the framework proposed here treats an entire degree program—including module additions, removals, updates, and credit-hour reallocation—as the unit of optimization (*Macro-Level Curricular Optimization*).
(ii) **Transparency.** [37] does not report an explainability mechanism; the proposed framework couples the RL formalization with a SHAP-based explainability layer specifically to support institutional, non-technical decision-makers (*Explainable AI Integration*).
(iii) **Contextual constraints.** [37] assumes a data-rich, structured vocational setting; the proposed framework is explicitly designed to remain operable under the low data-granularity and bureaucratic constraints typical of developing-country institutions (*Contextual Resilience for Developing Countries*).

Table 1 summarizes this contrast for [37] as "limited to micro-level vocational training tasks"; the full comparison is developed in the three points above. Thus, whereas [37] offers an algorithmic optimization technique for isolated training, this study proposes a comprehensive, scalable *governance framework design*, intended for systemic evaluation in resource-limited academic ecosystems once empirically validated.

3. Proposed methodological framework

3.1 Materials and formal problem definition

Over the last decade, Artificial Intelligence (AI) and Learning Analytics have been reported to reduce the gap between graduate skills and industrial requirements [14, 20]. Recent research underscores the potential of Deep Learning and RL to enhance educational adaptability, with RL offering more dynamic sequential optimization than traditional analytical approaches [29, 9, 37]. To situate this contribution within this expanding research landscape, Figure 2 illustrates the scientific shift from student-centered predictive models toward dynamic systems focused on continuous industrial alignment.

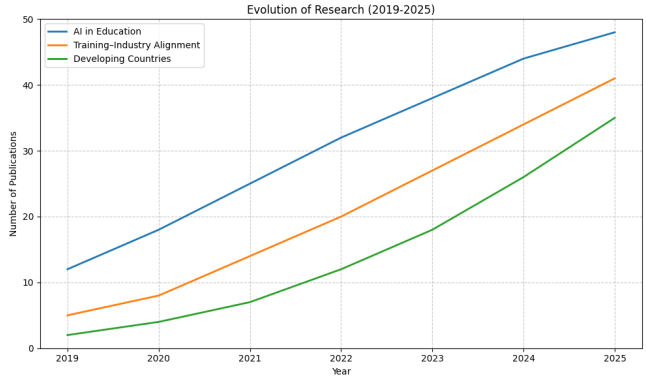


Figure 2. Illustrative trend of publication activity (2019-2025) across three converging research areas.

Figure 2 illustrates the growing, converging interest in AI in education, training-industry alignment, and developing-country contexts, based on a non-systematic, exploratory keyword search of indexed publications; as no database, search string, or inclusion/exclusion criteria were applied, the figure should be read as motivating context rather than as bibliometric evidence. Despite this recent progress, contemporary research focuses almost exclusively on micro-level optimizations, such as individual student tracking, while overlooking the broader institutional and industrial dimensions. Although interoperability and XAI are gaining traction as essential components for academic trust [39, 42], current frameworks rarely integrate them into a cohesive strategy. This gap motivates the multi-objective architecture proposed below, specifically designed for the structural realities of developing nations.

This shift is necessary because truly self-evolving mechanisms remain scarce in the literature. Most existing systems operate merely as static decision-support tools, lacking the continuous learning loops required for autonomous curricular adaptation. Furthermore, RL applications are still largely confined to optimizing student pathways [29, 9], leaving the challenge of institutional-scale program restructuring largely unaddressed. By tackling the specific hurdles of developing countries, ranging from data fragmentation to weak university-industry ties [14, 20], the framework proposed here moves beyond purely descriptive analysis toward a formally specified, scalable governance design.

The proposed methodology follows an integrated four-stage adaptive approach, illustrated in Figure 3: (1) a literature review and needs analysis to identify curricular gaps and emerging market skills; (2) formalization of the curriculum as a Reinforcement Learning environment (Markov Decision Process, MDP); (3) collection, integration, and semantic alignment of academic and industrial data to build a coherent state representation; and (4) implementation of the RL agent's decisions, interpretation via XAI techniques, and re-evaluation through a continuous, human-governed feedback loop.

3.1.1 Formal specification of the curricular MDP

To answer the request, raised independently by several reviewers, for a fully specified problem definition rather than a purely narrative description, the curricular-governance problem is formalized here as the tuple

$$\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, P, R, \gamma, H \rangle, \tag{1}$$

where $\mathcal{S}$ is the state space, $\mathcal{A}$ the action space, $P(s_{t+1} \mid s_t, a_t)$ the transition function, R the reward function, $\gamma \in [0, 1)$ the discount factor, and H the episode horizon (number of curricular-revision cycles per program lifetime considered).

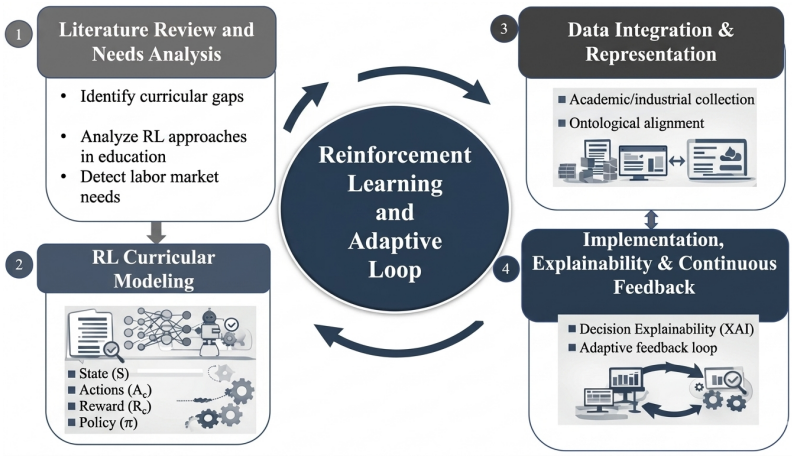


Figure 3. Proposed methodological approach.

State space $\mathcal{S}_t$. The state vector $s_t \in \mathcal{S}$ at decision cycle t is proposed to concatenate four groups of normalized indicators: (i) *academic* indicators (mean grade point average, pass rate, credit-hour distribution across modules); (ii) *employability* indicators (Skill Alignment Score, SAS_t, and six-month placement rate); (iii) *industrial* indicators (partner satisfaction score, number of emerging skills covered); and (iv) *budgetary* indicators (available credit-hour envelope, cost of the current module configuration).

Action space $\mathcal{A}_t$. At each cycle, the agent is proposed to select one action from a discrete set

$$\mathcal{A} = \{\texttt{add_module}, \texttt{update_module}, \texttt{remove_module}, \texttt{reallocate_credit_hours}, \texttt{no_change}\}, \quad (2)$$

parameterized by the module or credit block targeted, subject to institutional feasibility constraints (accreditation limits, prerequisite structure, staffing).

Transition dynamics P. The environment is expected to update s_t to s_{t+1} through (a) a deterministic accounting update for credit hours and implementation cost following the chosen action, and (b) a stochastic update of SAS, placement probability, and satisfaction, to be calibrated on historical institutional and labor-market data once such data are collected. Specifying and calibrating P from real data is identified as a required step before any RL agent can be trained, and is explicitly left to future empirical work.

Reward function R_t. The reward is proposed to combine four weighted objective terms and a modification-cost penalty:

$$R_t = w_1 \cdot \Delta\mathrm{Academic}_t + w_2 \cdot \Delta\mathrm{Employability}_t + w_3 \cdot \Delta\mathrm{Industry}_t + w_4 \cdot \mathrm{BudgetCompliance}_t - w_5 \cdot \mathrm{Cost}_t, \quad (3)$$

where each Δ term denotes the normalized ($[0, 1]$) change in the corresponding indicator between cycles $t-1$ and t, $\mathrm{BudgetCompliance}_t \in \{0, 1\}$ penalizes actions that violate the institutional credit-hour envelope, and Cost_t is the normalized administrative and financial cost of implementing the action. The weights $w_1, \ldots, w_5$ are design parameters to be elicited from institutional stakeholders (curriculum boards, industry partners) and validated through a sensitivity analysis showing whether recommended actions remain stable under plausible alternative weightings, rather than fixed a priori; no numerical values are reported here because no calibration exercise has yet been conducted.

Episode structure. One episode is proposed to correspond to a full degree-program revision cycle (e.g., $H = 4$ decision points over two academic years), terminating either at $t = H$ or upon a hard budget-constraint violation. The discount factor γ should be selected to reflect the institution's planning horizon and is left as a design parameter.

3.1.2 Explainable AI layer

An Explainable AI layer is proposed to make the (future) agent's recommendations interpretable to non-technical curriculum committees. The design envisions SHAP (SHapley Additive exPlanations) values computed for a surrogate function approximating the agent's action-selection probabilities from the state vector s_t, so that the relative contribution of each state variable (e.g., SAS, budget) to a given recommendation can be reported. As with the reward function, the concrete choice of background dataset, the granularity of explanations (local, per-decision, versus global, policy-level), and the handling of temporal dependencies across cycles are implementation decisions to be made once a trained agent exists.

3.1.3 Envisioned data sources

No dataset has been collected and no institutional deployment has taken place at the time of writing. The framework is designed with the Computer Science and Electrical Engineering departments of an engineering institution in Douala, Cameroon, as an illustrative future application context, chosen for the rapid technological evolution of these fields (AI, embedded systems, renewable energy). The data sources envisioned for a future empirical study are: (i) institutional academic records (curricula, grades, credit structure); (ii) job postings from professional platforms such as LinkedIn and Indeed, to be collected subject to the platforms' terms of service; and (iii) structured feedback from industrial partners and alumni. Before any such collection can occur, the ethical and data-governance conditions including institutional ethics approval, informed consent, and data-minimization safeguards will need to be secured.

3.2 Proposed architecture

To address the limitations identified in the literature, a self-evolving curricular architecture based on RL is proposed, specifically designed to dynamically align engineering training with industrial needs in developing countries. This architecture relies on a modular approach integrating intelligent data collection, semantic skill modeling, a sequential optimization engine, an XAI layer, and a lightweight infrastructure tailored to local constraints. The objective is to transform curricular governance from a static and periodic process into an adaptive, human-governed system driven by data and machine learning. The architecture is organized into seven sequential layers, introduced below and matched one-to-one with the numbered blocks of Figure 4:

1. **Heterogeneous data acquisition layer:** aggregates heterogeneous data from academic, student, and industrial sources. It distinguishes *static features* (student profile, academic history, learning style) from *dynamic features* evolving over time (LMS behavior, cognitive level, emotional state), while integrating macro-contextual data such as labor market trends, industrial needs, and institutional budgetary constraints. This multi-source structuring captures the individual, institutional, and socio-economic dimensions of the educational system.

2. **State representation (S_t):** transforms collected data into the contextual state representation S_t defined in Section 3.1.1, through a processing pipeline (ETL, normalization), semantic extraction via Transformer models, and ontological alignment between industrial skills and academic content.

3. **Data integration and ontological alignment:** structures and semantically aligns academic and industrial records (skills, course units, job profiles) into a coherent ontology-based representation

feeding the state vector.

4. **Reinforcement Learning optimization engine:** the core of the system is intended to be an RL agent (PPO/DQN) instantiating the MDP of Equations (1)–(3). Based on the state S_t, the agent would select *actions* such as adding, updating, or removing modules, as well as resource reallocation, guided by the multi-objective *reward function* of Equation (3).
5. **Curricular prediction and intervention:** the decisions generated by the RL agent would feed into an operational layer dedicated to performance prediction and curricular adjustment, translating algorithmic recommendations into concrete intervention strategies – modification planning, institutional implementation, and monitoring of effects.
6. **XAI:** to ensure transparency and system acceptability, an XAI layer using SHAP (or LIME) is proposed to identify the relative influence of state variables on the agent's recommendations, reducing the "black box" effect of the RL model for academic decision-makers.
7. **Human-governed feedback loop and dashboard:** after implementation of any curricular adjustment, performance indicators (professional placement, satisfaction, skill-gap evolution) would be collected and fed back into the system, subject to mandatory human curriculum-board approval. This feedback loop is what allows the reward function and learning policy to be updated over time rather than the curriculum being revised in a purely periodic, static manner. A dashboard layer would provide real-time monitoring, "what-if" prospective analysis, and strategic reporting.

Table 2 summarizes, for each of the seven layers above, its expected input, output, candidate supporting technology, and an indicative design-level latency/cost class.

Table 2. Architectural layers, their input/output, and supporting technology.

Layer	Input	Output	Candidate Technology	Design Latency/Cost Class
1. Data Acquisition	Job offers, LMS logs, academic records	Raw data	Scrapy	Low
2. State Representation	Normalized data	State vector S_t	BERT	Medium
3. Data Integration & Ontology	Raw/normalized data	Aligned ontology instance	OWL, Neo4j	Medium
4. RL Optimization Engine	State vector S_t	Action A_t	PPO/DQN (e.g., Stable-Baselines3)	High
5. Prediction & Intervention	Action A_t	Intervention plan	Rule-based post-processing	Low
6. XAI	State vector, policy output	Feature-importance report	SHAP	Medium
7. Feedback & Dashboard	Post-implementation KPIs	Updated reward, reports	Web dashboard	Low

Figure 4 depicts the resulting pipeline. Data flows sequentially through the seven layers, $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 7$, and the human-governed feedback loop embedded in layer 7 closes the cycle back to layer 1 for the next curricular-revision period.

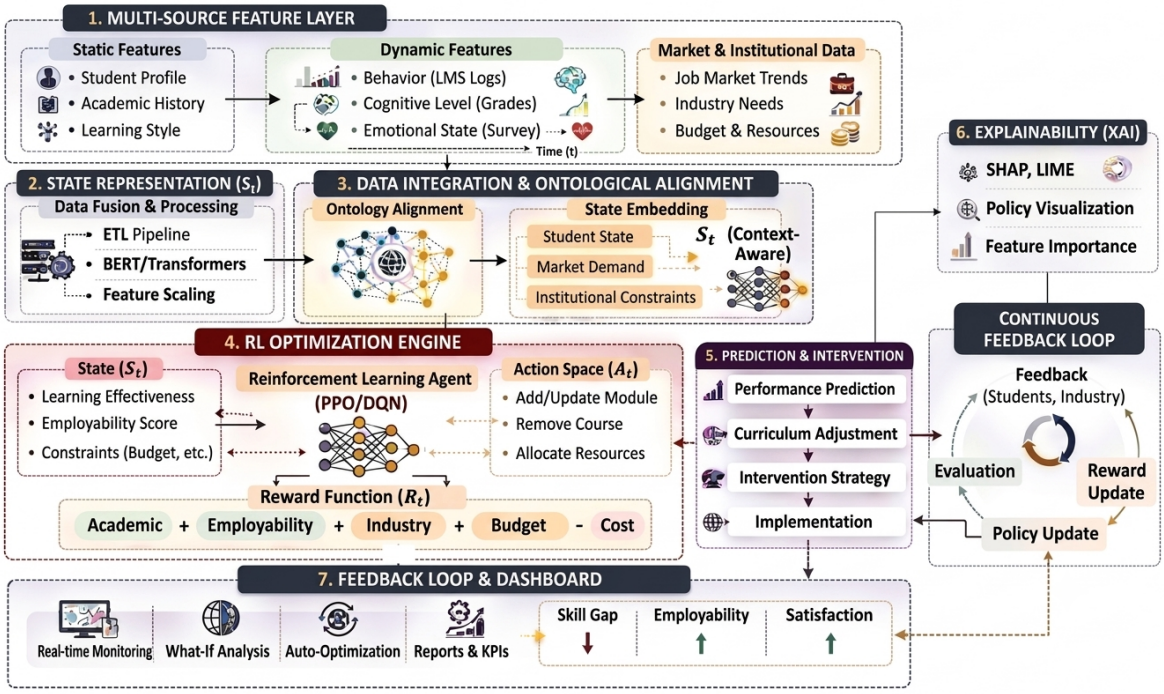


Figure 4. Proposed seven-layer architecture.

4. Proposed validation protocol, expected contributions and governance

This section replaces what was, in the originally submitted version, presented as a set of simulation results. No computational simulation, dataset, or trained agent exists at this stage; the paragraphs below instead specify the protocol that would need to be executed to empirically validate the architecture of Section 3, together with the governance and ethical safeguards that would need to accompany any such validation.

4.1 Envisioned application scenario

The architecture is designed to be applicable to the Computer Science and Electrical Engineering departments of an engineering institution in Douala, Cameroon, chosen as an illustrative context given the rapid technological evolution in these fields. A future empirical study would need to instantiate the proposed seven-layer architecture, including automated data acquisition from professional platforms, semantic skill extraction, ontology-based knowledge representation, and a PPO/DQN optimization engine using real institutional data. It would also need to report the dataset size, collection window, feature list, and preprocessing pipeline in full, as required for reproducibility.

4.2 Proposed evaluation metrics and baselines

Table 3 lists the key performance indicators the architecture is designed to track, together with how each would be measured and against which baselines it would need to be compared. No values are reported for these indicators in this paper; they define the measurement protocol for future work.

A methodologically important requirement, raised independently by several reviewers, is that the employability indicator must be measured *independently* of the SAS term that the reward function in

Table 3. Proposed key performance indicators (KPIs), their definition, and candidate baselines for future empirical evaluation. No values are reported here; this table specifies the measurement protocol, not obtained results.

Indicator	Proposed definition / measurement	Candidate baselines for comparison
Skill Alignment Score (SAS)	Ontology-based overlap between the skill set taught and the skill set demanded by collected job postings, normalized to $[0, 1]$	No-change (static curriculum); rule-based periodic revision
Six-month employability rate	Share of graduates placed within six months, from independent longitudinal follow-up data (not derived from SAS)	Historical institutional placement records
Emerging-skill coverage	Share of modules covering skills flagged as "emerging" in the industry data	Supervised ranking without RL; classic Learning Analytics
Credit-hour reallocation	Share of total credit hours reassigned by the agent's actions	Multi-objective optimization without RL
Industrial partner satisfaction	Structured survey of partner organizations	No-change baseline

Equation (3) optimizes; using a SAS-derived projection as a stand-in for employability would make the outcome circular with respect to the optimized objective, and any future empirical study must avoid this by using external, longitudinally observed placement data. Future work should also report means and standard deviations, or confidence intervals, over multiple independent training seeds, together with convergence diagnostics (learning curves, episode returns), rather than single-run figures.

4.3 Illustrative curriculum case study

To assess whether the policy recommended by a trained agent is pedagogically coherent, and not merely numerically optimal, a future validation study should present at least one auditable, worked example for one program (e.g., Computer Science): the baseline module list, the industry skills extracted for that program, the ontology matches, the specific actions recommended by the agent, the resulting credit distribution, prerequisite impacts, faculty and laboratory requirements, accreditation limits, an estimated implementation cost, and the human curriculum board's decision. This case study is identified here as a required deliverable of future empirical work.

4.4 Human governance and responsible use

The framework is designed to *support*, not replace, human curricular decision-making. Any module addition, update, or removal recommended by the RL agent would require mandatory approval by the institutional curriculum board before implementation. The design anticipates: (i) stakeholder representation (faculty, students, industry, accreditation bodies) in the approval process; (ii) an appeal/override mechanism for board members who disagree with a recommendation; (iii) an audit trail and version control for every adopted curricular change, with rollback capability; (iv) a minimum review interval between successive changes to guard against oscillation and abrupt module turnover; and (v) explicit safeguards protecting foundational/theoretical content from being deprioritized in favor of short-term market trends. Potential biases in the envisioned data sources, including underrepresentation of informal and public-sector employment, regional and language inequities in online job postings, and employer dominance in shaping the reward signal are acknowledged as open risks requiring ongoing monitoring rather than one-time mitigation.

4.5 Ethics, privacy, and data governance

Because the architecture envisions processing sensitive data, including student profiles, academic histories, LMS behavior, learning styles, and inferred emotional state, alongside employer information and content collected from third-party professional platforms, any future empirical implementation would need to secure, in advance: (i) institutional ethics approval (or a documented exemption) for the use of student and academic data; (ii) informed consent and institutional authorization procedures; (iii) a data-minimization, pseudonymization, security, retention, and deletion policy; (iv) a documented legal and ethical basis for collecting data from third-party platforms, compliant with those platforms' terms of service; and (v) access-control and cybersecurity measures for the ontology and model pipeline. No such approvals have been sought at this stage, since no data collection has taken place.

4.6 Limitations of the proposed approach

The framework presented above is deliberately conceptual, and its limitations should be read as an explicit research agenda rather than as caveats appended after the fact.

- **Conceptual stage.** This paper formalizes and proposes an architecture; it does not report an implementation, a trained agent, or empirical results. All quantitative claims made elsewhere in this text (e.g., regarding SAS, placement rates, or reward values) are design specifications, not measured outcomes.
- **Unvalidated transition and reward calibration.** The transition dynamics P and the reward weights $w_1, \ldots, w_5$ in Equations (1)–(3) are design parameters, not empirically fitted quantities. Their calibration on real institutional and labor-market data, together with a sensitivity analysis over plausible weightings, is a prerequisite for any confidence in the resulting policy; without it, the reward function remains a hypothesis about institutional priorities rather than a validated objective.
- **Risk of circular evaluation.** As noted in Section 4, employability must be measured independently of the Skill Alignment Score (SAS) term optimized by the reward function. Failing to enforce this separation in future empirical work would make the evaluation partially circular with respect to the optimized objective, artificially inflating apparent performance.
- **Data availability and representativeness.** The data sources envisioned in Section 3.1.3 (institutional academic records, job postings from platforms such as LinkedIn and Indeed, and structured industry feedback) have not yet been collected. Their coverage, quality, and representativeness for developing-country labor markets, which are often characterized by a large informal sector that is underrepresented in online job postings, remain unverified and constitute a first-order risk for the validity of any downstream model.
- **Algorithmic and computational feasibility.** The choice of RL algorithm (e.g., PPO or DQN, as listed in Table 2) and its sample efficiency have not been assessed against the scale and update frequency realistically available in a resource-constrained institutional setting. Sparse decision cycles (a few per academic year) may be insufficient to train a stable policy without simulation-based pretraining or transfer learning, which is not addressed in this paper.
- **Human and bureaucratic complexity.** Any future simulation environment will necessarily simplify the human and bureaucratic factors, including negotiation among stakeholders, informal influence, and resistance to change that shape real pedagogical decision-making, and may therefore only partially capture the institutional dynamics the architecture is meant to support.
- **Scope and generalizability.** The illustrative application context (Computer Science and Electrical Engineering departments at a single institution in Douala) is narrower than the paper's general

framing around "engineering curricula in developing countries." Consequently, findings from any future single-institution pilot should not be generalized beyond that validated context without a subsequent multi-institutional study, as outlined in Section 5.

5. Discussion and future work

The contribution of this paper is the formalization of curricular governance as an MDP, the design of a seven-layer modular architecture, and a validation protocol that explicitly separates conceptual design from empirical claims. This addresses a gap in prior RL-for-education research, which remains largely focused on student-level rather than institutional-level optimization [29, 9].

Future research will proceed along the following axes, building on the limitations identified above. First, the data-collection and ethics-approval steps must be completed before any model training can begin. Second, the reward-weight elicitation and sensitivity analysis must be conducted with institutional stakeholders. Third, once training data are available, the RL agent should be trained and evaluated against the baselines and KPIs of Table 3, with results reported using multiple independent seeds, convergence diagnostics, and confidence intervals, and with employability measured independently of the optimized SAS term. Fourth, to address data scarcity during early validation, synthetic data generation and an Edge-AI architecture (to reduce computational costs via distributed learning) will be explored. Fifth, Multi-Agent Reinforcement Learning (MARL) will be investigated to let students, teachers, and employers act as agents contributing directly to the reward signal, fostering more collaborative governance. Sixth, adaptive ontologies will be explored to enable the dynamic evolution of skill taxonomies using generative models. Finally, once a single-institution pilot is validated, a multi-institutional study at the national level is planned to evaluate the transferability and robustness of the framework across diversified academic contexts.

6. Conclusion

Engineering education in developing countries continues to suffer from a structural mismatch between academic curricula and labor-market needs, as slow, periodic, and largely descriptive revision processes struggle to keep pace with rapid technological change. Traditional curricular governance mechanisms, hindered by administrative inertia, fragmented industrial data, and limited stakeholder integration, are ill-suited to sustaining this alignment over time.

To address this gap, this paper proposed a conceptual, Reinforcement-Learning-based architecture for curricular governance, formalizing the evolution of academic programs as a constrained sequential optimization problem using a Markov Decision Process, a multi-criteria reward function balancing academic, employability, industrial, and budgetary objectives, and an Explainable AI layer together with an explicit human-governance and ethics framework. Unlike traditional, periodic and descriptive curricular-revision methods, the proposed architecture is designed for continuous adaptation to labor-market needs through the integration of external data, multi-objective decision logic, and a reward mechanism that incorporates industrial stakeholders directly into the optimization loop. Beyond this formalization, the paper's main contribution lies in reframing curricular management as a formally specifiable, human-governed adaptive-decision problem, capable in principle of learning, adjusting, and optimizing recommendations in a dynamic and uncertain environment, provided the calibration, data-governance, and validation steps identified throughout this work are completed. No implementation has been carried out, and consequently no performance claim is made: the contribution is the formal problem specification, the modular architecture design, its positioning relative to existing RL-based education-alignment work, and a concrete, ethically grounded validation protocol for future empirical work.

This paper is therefore intended as a foundation for, not a substitute for, the empirical work required to demonstrate that data-driven, explainable, and multi-stakeholder curricular governance is achievable in resource-constrained settings. Future work should prioritize, in sequence: securing institutional ethics approval and completing the envisioned data-collection effort; eliciting and validating the reward-function weights with curriculum boards and industry partners through a sensitivity analysis; training and evaluating the RL agent against the proposed baselines and key performance indicators, with employability measured independently of the optimized alignment score to avoid circular evaluation; and, once a single-institution pilot has been validated, extending the framework to a multi-institutional study to assess its transferability and robustness across diversified academic contexts. Longer term, this line of work could be extended toward synthetic data generation and edge-AI deployment to mitigate early data scarcity, Multi-Agent Reinforcement Learning to let students, teachers, and employers act as agents directly shaping the reward signal, and adaptive ontologies capable of evolving skill taxonomies through generative models. Taken together, these steps chart a path from the present architectural proposal toward an empirically validated, trustworthy, and institutionally governed instrument for aligning engineering education with the evolving needs of developing-country labor markets.

References

[1] David Mhlanga. "Digital transformation of education, the limitations and prospects of introducing the fourth industrial revolution asynchronous online learning in emerging markets". In: *Discover education* 3.1 (2024), p. 32.

[2] Ahmed Ayad. *Réussir le transfert technologique: Guide pour les pays en développement*. Chihab, 2025.

[3] Rahma Sahali. "L'innovation en Algérie: quelle formation dans les écoles d'ingénieurs pour quelles innovations?" PhD thesis. HESAM Université; Ecole Nationale Superieur de Management (ENSM), 2022.

[4] Vui-Yee Koon, Sin Kit Yeng, Chong Soon Meng, and Poon Wai Chuen. "Transformative trends and challenges: Navigating disruptive innovation in higher education through bibliometric insights". In: *Innovations in Education and Teaching International* 63.2 (2026), pp. 356–369.

[5] Hunter Hughes. "Education Technology & Digital Upskilling in Emerging Economies: Utilizing Innovation for Growth & Development". In: *Available at SSRN 5719702* (2025).

[6] Amando Jr Singun. "Unveiling the barriers to digital transformation in higher education institutions: a systematic literature review". In: *Discover Education* 4.1 (2025), p. 37.

[7] Nelson Sizwe Madonsela. "Aligning Education and Workforce Training with Industry Needs: A Perspective on Human Capital Development". In: *Proceedings of the First Australian International Conference on Industrial Engineering and Operations Management*. 2022, pp. 20–21.

[8] R Satheeskumar, Ch V Satyanarayana, Talatoti Ratna Kumar, and M Suresh. "AI-Driven alignment of educational programs with industry needs and emerging skillsets [J]". In: *International Journal of Modern Education and Computer Science* 17.3 (2025), pp. 15–28.

[9] Fabrizio Stasolla, Antonio Zullo, Roberto Maniglio, Anna Passaro, Mariacarla Di Gioia, Enza Curcio, and Elvira Martini. "Deep learning and reinforcement learning for assessing and enhancing academic performance in university students: A scoping review". In: *AI* 6.2 (2025), p. 40.

[10] Rama Yusvana. "Addressing the skills gap in technical and vocational training for sustainable socio-economic growth and development". In: *International Journal of Research and Innovation in Social Science* 8.IIIS (2025), pp. 6311–6325.

[11] Md Jahangir Alam, SM Ali Reza, Keiichi Ogawa, and Abu Hossain Muhammad Ahsan. "Sustainable employment for vocational education and training graduates: the case of future skills

matching in Bangladesh". In: *International Journal of Training Research* 22.3 (2024), pp. 266–288.

[12] Zinan Sharrad Arsha. *An Undergraduate Internship/Project on Industry Academia Linkage and the Identification of Skills Gaps of the Graduating Engineering Students at Tertiary Level Education*. Tech. rep. Independent University, Bangladesh, 2023.

[13] Md Abu Issa Gazi, Md Kazi Hafizur Rahman, Mohd Faizal Yusof, Abdullah Al Masud, Md Aminul Islam, Abdul Rahman bin S Senathirajah, and Md Alamgir Hossain. "Mediating role of entrepreneurial intention on the relationship between entrepreneurship education and employability: a study on university students from a developing country". In: *Cogent Business & Management* 11.1 (2024), p. 2294514.

[14] Lin Xu, Jingxiao Zhang, Yiying Ding, Gangzhu Sun, Wei Zhang, Simon P Philbin, and Brian HW Guo. "Assessing the impact of digital education and the role of the big data analytics course to enhance the skills and employability of engineering students". In: *Frontiers in Psychology* 13 (2022), p. 974574.

[15] Prince Dacosta Anaman, Deborah Morpkorpkor Zottor, and Julius Kumi Egyir. "Infrastructural challenges and student academic performance: Evidence from a developing nation". In: *International Journal of Innovative Science and Research Technology* 7.11 (2022), pp. 1189–1200.

[16] Boahemaa Brenya. "Higher education in emergency situation: Blended learning prospects and challenges for educators in the developing countries". In: *Journal of Applied Research in Higher Education* 16.4 (2024), pp. 1015–1028.

[17] Victor Olugbenga Ayoko, Thankgod Peter, and Deborah Oluwaseun Jegede. "Inadequacy of infrastructural facilities in public universities in Nigeria: Causes, effects and solutions". In: *International Journal on Integrated Education* 6.3 (2023), p. 36.

[18] James Hutson. *The adoption of artificial intelligence and inertia in higher education: Exploring complex resistance to technological change*. Taylor & Francis, 2025.

[19] Olatunde Fatai Badmus. "Transforming Program Management through Generative Artificial Intelligence". In: *International Journal of Latest Engineering and Management Research (IJLEMR)* 8.12 (2023).

[20] Peter Oeij, Karolien Lenaerts, Steven Dhondt, Wietse Van Dijk, Doris Schartinger, Sabrina Sorko, and Chris Warhurst. "A conceptual framework for workforce skills for industry 5.0: Implications for research, policy and practice". In: *Journal of Innovation Management* 12.1 (2024), pp. 205–233.

[21] Kathiravan Ravichandran. "Identifying learning difficulties at an early stage in education with the help of artificial intelligence models and predictive analytics". In: *International Research Journal of Multidisciplinary Scope (IRJMS)* 5.4 (2024), pp. 1455–1461.

[22] Yali Li, Laura Tolosa, Francklin Rivas-Echeverria, and Ronald Marquez. "Integrating AI in chemical education: Navigating UNESCO global guidelines, emerging trends, and its intersection with sustainable development goals". In: (2025).

[23] Shafika Isaacs and Sanjaya Mishra. "Smart education strategies for teaching and learning: Critical analytical framework and case studies". In: (2022).

[24] Giada Lalli. "Defining interoperability: a universal standard". In: *Journal of ICT Standardization* 13.2 (2025), pp. 139–156.

[25] Carmen Fattore, Michele Buldo, Arcangelo Priore, Sara Porcari, Vito Domenico Porcari, and Mariella De Fino. "A comprehensive overview of heritage bim frameworks: platforms and technologies integrating multi-scale analyses, data repositories, and sensor systems". In: *Heritage* 8.7 (2025), p. 247.

[26] Shihui Wang, Adzrool Idzwan Bin Ismail, and Panpan Qiao. "Optimized RE-CNN-Based Multi-Objective Decision Framework for Visual Feature Evaluation in Computational Art Analysis and Interactive Media". In: *Decision Making: Applications in Management and Engineering* 7.1 (2024), pp. 752–770.

[27] Abdul Wali Sirat, Musawer Hakimi, Bilal Himmat, and Wahidullah Enayat. "Artificial intelligence in educational leadership: Strategic, analytical, interactive, and decision-making applications for the digital age". In: *Jurnal Ilmiah Telsinas Elektro, Sipil dan Teknik Informasi* 8.2 (2025), pp. 210–224.

[28] Sanjai Vudugula, Sanath Kumar Chebrolu, Maniruzzaman Bhuiyan, and Farhana Zaman Rozony. "Integrating artificial intelligence in strategic business decision-making: A systematic review of predictive models". In: *International Journal of Scientific Interdisciplinary Research* 4.1 (2023), pp. 01–26.

[29] Keerthana Sivamayil, Elakkiya Rajasekar, Belqasem Aljafari, Srete Nikolovski, Subramaniyaswamy Vairavasundaram, and Indragandhi Vairavasundaram. "A systematic study on reinforcement learning based applications". In: *Energies* 16.3 (2023), p. 1512.

[30] Christopher J. C. H. Watkins. "Learning from Delayed Rewards". PhD thesis. University of Cambridge, 1989.

[31] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998.

[32] Mohamed-Amine Chadi and Hajar Mousannif. "Understanding reinforcement learning algorithms: The progress from basic Q-learning to proximal policy optimization". In: *arXiv preprint arXiv:2304.00026* (2023).

[33] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. "Proximal Policy Optimization Algorithms". In: *arXiv preprint arXiv:1707.06347* (2017).

[34] Volodymyr Mnih et al. "Human-level control through deep reinforcement learning". In: *Nature* 518.7540 (2015), pp. 529–533. DOI: 10.1038/nature14236.

[35] Agus Perdana Windarto, Solikhun Solikhun, and Anjar Wanto. "Enhancing autonomous vehicle navigation in urban traffic using CNN-based deep Q-networks". In: *Journal of Applied Data Sciences* 6.4 (2025), pp. 2565–2581.

[36] Juan Escobar-Naranjo, Gustavo Caiza, Paulina Ayala, Edisson Jordan, Carlos A Garcia, and Marcelo V Garcia. "Autonomous navigation of robots: optimization with DQN". In: *Applied Sciences* 13.12 (2023), p. 7202.

[37] Yusong Liu and Jianjun Song. "Research on real-time dynamic adjustment strategy of industry-teaching integration practical training process in higher vocational education based on reinforcement learning". In: *J. Combin. Math. Combin. Comput* 127 (2025), pp. 3161–3176.

[38] Claudia-Melania Chituc. "Exploring the Role of Semantic Web Technologies and Ontologies in the Digital Education Ecosystems for Education 4.0". In: *2025 IEEE 12th International Conference on E-Learning in Industrial Electronics (ICELIE)*. IEEE. 2025, pp. 1–6.

[39] Ana Perisic, Ines Perisic, Marko Lazic, and Branko Perisic. "The foundation for future education, teaching, training, learning, and performing infrastructure-The open interoperability conceptual framework approach". In: *Heliyon* 9.6 (2023).

[40] Idrees Alsolbi. "Enabling Data-Driven Innovation in the Third Sector: A Framework for Leveraging Big Data". In: *Cloud Computing and Data Science* (2026), pp. 126–153.

[41] Jesus Kombaya Touckia, Nadia Hamani, and Lyes Kermad. "Digital twin framework for reconfigurable manufacturing systems (RMSs): design and simulation". In: *The International Journal of Advanced Manufacturing Technology* 120.7 (2022), pp. 5431–5450.

[42] Carter Cousineau. "Advancing Trustworthy Artificial Intelligence (AI): Practical Approaches for Enhancing Interpretability and Transparency". PhD thesis. The University of Guelph, 2025.

[43] Bery Leouro Mbaiossoum, Mahamat Atteib Ibrahim Doutoum, Apollinaire Batoure Bamana, Narkoy Baoutma, et al. "Self-Efficacy Prediction Model Using Bayesian Networks". In: *International Journal on Interactive Systems and Engineering Management (JISEM)* 10.37s (2024). DOI: 10.52783/jisem.v10i37s.6447.

Canadian Journal of Artificial Intelligence for Learning and Engineering Innovation (2026), Vol. **1**, No. **1**
ISSN 3111-4296
https://doi.org/10.65269/5p09zq79

CuspideScience
At the cutting edge of science

PROCEEDINGS | ***African Engineering Education Forum (AEEF 2026)***

Malloum: An SMS-Based Generative AI Tutoring System for Contextualized Learning in Low-Connectivity Environments

Isaac Touza [*,1,2]

[1]Department of Mathematics-Computer Science, Faculty of Sciences, University of Maroua, Cameroon
[2]Laboratoire de Recherche en Informatique (LaRI), Université de Maroua, Cameroon
*Corresponding author: isaac.touza@univ-maroua.cm

(Received: 18 June 2026; Revised: 04 August 2026; Accepted: 20 August 2026; Published: 22 August 2026)
(Editor: Pr. Justin Moskolaï; Reviewers: Anonymous reviewers)

Abstract

Limited internet connectivity continues to constrain access to digital learning resources across many low- and middle-income countries, particularly in Sub-Saharan Africa. While generative artificial intelligence has demonstrated strong potential for personalized tutoring, most existing systems assume reliable broadband access and smartphone availability. This paper presents Malloum, an SMS-based generative AI tutoring system designed to deliver contextualized, curriculum-aligned pedagogical support in low-connectivity environments. Malloum enables students to submit academic questions via standard SMS to a dedicated short code. The system integrates a backend orchestration layer with a hierarchical prompt architecture leveraging a state-of-the-art generative language model. A multi-layered prompting strategy combines a system-level pedagogical role definition, contextual curriculum injection aligned with the Cameroonian national syllabus, and user-level query input. An educational database (including validated exam banks, reference materials, and structured student profiles) enhances response accuracy, personalization, and pedagogical alignment. To address SMS-specific constraints, the platform implements structured micro-explanations, intelligent compression algorithms, automatic message concatenation, and adaptive bilingual output (French/English). A matched-pair cluster-assigned field study conducted with 1,000 secondary school students across 20 schools in the Far North Region of Cameroon using a pre-test/post-test design with a comparison group shows that the mean composite score of the Malloum group rose from 51.7% to 72.4%—an absolute gain of 20.7 percentage points, corresponding to a 40% relative improvement over baseline—compared with a gain of 4.3 percentage points (8.3% relative) in the control group, yielding an adjusted between-group difference of 16.1 percentage points (95% CI: 13.9–18.3, $p < 0.001$). Malloum provides a scalable, cost-aware, and culturally contextualized framework for AI-driven education, offering a practical pathway to narrowing the digital divide in secondary education systems.

Keywords: Malloum; Generative AI; SMS-based learning; Intelligent tutoring system; Low-connectivity environments

Abbreviations: AI: Artificial Intelligence; LLM: Large Language Model; RAG: Retrieval-Augmented Generation; SMS: Short Message Service; ITS: Intelligent Tutoring System; ICT4D: Information and Communication Technologies for Development; BEPC: Brevet d'Études du Premier Cycle; GCE: General Certificate of Education

© 2026 by the author. This is an Open Access article, distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/)

1. Introduction

Access to quality educational resources remains one of the most pressing challenges in developing countries, particularly across Sub-Saharan Africa. The proliferation of digital technology has theoretically expanded educational opportunities, yet structural inequalities in connectivity infrastructure continue to prevent millions of students from benefiting from AI-powered learning platforms. According to GSMA [1], mobile internet penetration in Sub-Saharan Africa reached only 27% in 2023, with rural areas bearing the sharpest exclusion due to affordability barriers and limited infrastructure.

In Cameroon, as in many countries of the region, secondary school students face a compound challenge: limited access to qualified teachers, scarcity of updated pedagogical materials, and insufficient technological infrastructure for internet-based learning. Kelly & Rutazihana [2] demonstrate that the digital divide exerts a significant negative effect on secondary school enrollment and completion rates in Sub-Saharan Africa. Despite mobile telephone penetration reaching over 80% of the Cameroonian population, internet-based educational services remain inaccessible or prohibitively expensive for the majority of learners [3].

Traditional EdTech solutions whether web-based learning management systems, mobile applications, or synchronous video tutoring platforms, presuppose conditions of reliable broadband connectivity and modern device availability. This structural mismatch between the capabilities of available educational technology and the infrastructure realities of target populations has resulted in a persistent gap: the students most in need of pedagogical support are systematically excluded from AI-powered learning tools.

Generative artificial intelligence has emerged as a transformative force in educational technology. Large language models (LLMs) demonstrate remarkable capacities for natural language understanding, pedagogical explanation, contextual reasoning, and adaptive communication [4]. These capabilities position generative AI as a compelling foundation for intelligent tutoring systems. Pardos & Bhandari [5] demonstrated in a randomized study that ChatGPT-generated pedagogical hints produce learning gains equivalent to those of human tutors, while Kestin & Miller [6] found in a randomized controlled trial that AI tutoring significantly outperformed in-class active learning in terms of learning outcomes and student engagement.

A fundamental innovation gap therefore exists between the promise of generative AI for education and its practical deployability in infrastructure-constrained environments. Bridging this gap requires rethinking the interaction model: from internet-dependent, app-based interfaces toward infrastructure-light communication channels that already exist in the hands of learners.

The present work is motivated by a concrete educational challenge observed in the Far North Region of Cameroon, where students in secondary education have access to basic mobile phones with SMS capability but limited or no internet access. These students face examination preparation challenges exacerbated by high teacher-student ratios, limited study materials, and absence of individualized tutoring. The SMS channel is universally available, requires no smartphone, functions on feature phones, and incurs predictable low per-message costs manageable for most families. The hypothesis underlying Malloum is that routing generative AI pedagogical capabilities through the SMS infrastructure can substantially broaden access to intelligent tutoring in contexts where it is most needed.

The name Malloum (meaning ‘learned person’ or ‘teacher’ in the Fulfulde language of the Northern regions of Cameroon) reflects the project’s commitment to cultural grounding and community relevance.

This paper makes the following original contributions:

- Design and implementation of Malloum, an SMS-based generative AI tutoring system requiring no internet access, data subscription, or smartphone on the student side ;
- A hierarchical three-layer prompting architecture combining system-level pedagogical framing, curriculum context injection aligned with the Cameroonian national syllabus via Retrieval-Augmented Generation (RAG), and user-level query handling;
- Formal algorithms for SMS response compression, multi-SMS concatenation, and adaptive personalization;
- An empirical evaluation with 1,000 secondary school students over one academic trimester showing a statistically significant absolute gain of 20.7 percentage points in composite academic performance (a 40% relative improvement over baseline) in the treatment group ;
- A replicable architectural framework for AI-powered education in infrastructure-constrained settings applicable across Sub-Saharan Africa.

2. Related Work

2.1 Intelligent Tutoring Systems

Intelligent Tutoring Systems (ITS) represent decades of research at the intersection of cognitive science, artificial intelligence, and educational psychology. Contemporary ITS research increasingly incorporates machine learning for knowledge tracing and adaptive feedback. A systematic review by Létourneau et al. [7] covering K-12 ITS deployments demonstrates consistent learning gains in problem-solving and critical thinking, with pre-post test improvements ranging from 7 to 12 percentage points across randomized studies. However, these systems continue to require structured interaction paradigms and significant computational infrastructure, limiting accessibility in resource-constrained settings.

2.2 Generative AI in Education

The emergence of transformer-based LLMs has substantially broadened AI-assisted tutoring. Kasneci et al. [4] provide a comprehensive survey of the opportunities and challenges posed by LLMs in education, highlighting capabilities in explanation generation, feedback provision, and adaptive scaffolding. Pardos & Bhandari [5] demonstrated that ChatGPT-generated hints produce comparable learning gains to human tutors in mathematics across a randomized study of 274 students. Dan et al. [8] introduced EduChat, a large-scale LLM-based chatbot system for intelligent education, while Nguyen et al. [9] proposed PRAG-EDU, a context-aware RAG framework that dynamically calibrates response complexity using students' academic profiles, achieving a 23.7% improvement in pedagogical relevance scores over non-personalized baselines.

Kestin & Miller [6] provide particularly compelling evidence: their RCT comparing AI tutoring against in-class active learning found significantly superior learning outcomes in the AI tutoring condition, with students reporting higher engagement and motivation. This builds on the foundational 'two sigma' hypothesis of Bloom [10], which posits that one-on-one tutoring produces learning advantages of two standard deviations over classroom instruction.

2.3 RAG and Prompt Engineering for Educational AI

Retrieval-Augmented Generation (RAG) has emerged as the leading technique for grounding LLM outputs in domain-specific knowledge, directly addressing the hallucination problem that threatens reliability in educational contexts [11, 12]. Gao et al. [11] provide a comprehensive survey of

RAG architectures, distinguishing naive, advanced, and modular RAG paradigms. In educational settings, Swacha & Gracel [13] survey 15 RAG-based chatbot deployments, confirming that curriculum-aligned RAG significantly improves factual accuracy and curriculum relevance of AI-generated responses. Prompt engineering—combining chain-of-thought prompting, role-based system prompts, and few-shot examples—further shapes the pedagogical quality of LLM outputs [14].

2.4 Mobile Learning in Low-Connectivity Environments

The ICT4D literature documents numerous SMS and USSD-based learning initiatives in Sub-Saharan Africa. M-Shule in Kenya uses SMS-based AI to reach students without smartphones [15]. Boateng et al. [16] developed Kwame for Science, a BERT-based AI teaching assistant for secondary science education in West Africa, demonstrating the value of locally contextualized educational AI despite the constraint of web-only access. Luma Learn, an AI tutor running on WhatsApp rather than a dedicated app, explicitly addresses the connectivity constraint by leveraging ubiquitous messaging infrastructure [17].

2.5 Research Gap

Before positioning our approach, it is important to conduct a comparative analysis of existing solutions in intelligent tutoring and AI-assisted education. While several recent systems have explored either the integration of generative artificial intelligence or the use of alternative communication channels such as SMS, these approaches remain largely partial and fragmented. Some prioritize advanced pedagogical personalization but rely on heavy infrastructure (web or mobile applications), while others adopt accessible SMS-based solutions without fully leveraging the capabilities of generative language models. This lack of integrated consideration of technological, pedagogical, and contextual dimensions motivates a structured comparison, presented in Table 1, to clearly situate the contribution of Malloum within the existing body of work.

Table 1. Positioning of Malloum relative to existing approaches in the literature.

System / Study	Approach	Limitation	Addressed by Malloum
Khanmigo [18, 19]	GPT-based web tutoring chatbot	Internet + smartphone required on the student side	SMS delivery, no student-side internet required
M-Shule [15]	Adaptive SMS-based micro-learning with personalized content sequencing	No generative , open-ended question answering	Generative LLM tutoring + RAG grounding
Kwame for Science [16]	BERT-based QA for West Africa	Web app, no SMS delivery	SMS delivery, curriculum-aligned
Luma Learn [17]	LLM tutor delivered over WhatsApp	Requires internet-capable device and data	SMS delivery on feature phones
PRAG-EDU [9]	Context-aware RAG with grade-level adaptation	Web-based delivery, connectivity assumed	SMS delivery with profile-driven RAG
Deep Knowledge Tracing [20]	ML-based student modeling	Complex infrastructure dependency	Lightweight SMS architecture

This comparison was assembled from a search of peer-reviewed AI-in-education and ICT4D literature (ACM, IEEE, Scopus, and Google Scholar; keywords combining “SMS”, “USSD”, “intelligent tutoring”, “generative AI”, “RAG”, and “Sub-Saharan Africa”), complemented by grey-literature de-

scriptions of deployed platforms. It suggests a research gap at the intersection of three dimensions: (i) generative , open-ended AI tutoring with personalization, (ii) SMS-based infrastructure-light delivery on feature phones, and (iii) curriculum-aligned contextualization for the Cameroonian educational context. To the best of our knowledge, no system reported in the literature surveyed combines all three dimensions , although individual systems address one or two of them. Malloum is designed to occupy this intersection .

3. System Design and Architecture of Malloum

This section presents the overall design of the Malloum platform, including its interaction model and underlying technical architecture, which together enable scalable AI-powered tutoring through SMS in low-connectivity environments.

3.1 System Design and Interaction Model

This subsection describes the conceptual foundations of the platform, the user interaction workflow, and the pedagogical principles guiding the design of Malloum.

3.1.1 Platform Concept and Design Principles

Malloum is an AI-powered personal tutor accessible via standard SMS, designed for secondary school students in Cameroon. The platform's fundamental design principle is infrastructure minimalism: the system must function effectively on the most basic mobile handset, using only the SMS channel, without requiring application installation, internet connection, or data subscription. Subject coverage spans the Cameroonian national examination curricula for all classes: Mathematics, Physics, Chemistry, Biology, Computer Science, French Literature, English Language, History, Geography, Philosophy and others.

3.1.2 User Interaction Workflow

The Malloum user interaction follows a simple asynchronous question-answer protocol optimized for the SMS medium:

1. A student formulates an academic question and sends it via SMS to the Malloum dedicated short code;
2. The SMS gateway, interfacing with APIs of Cameroonian mobile operators (MTN and Orange), receives the message;
3. The backend server retrieves the student profile and conversation history, then constructs a hierarchical prompt package using RAG-based curriculum retrieval;
4. The prompt package is dispatched to the Generative AI Module (GPT-4o via the OpenAI API), which generates a pedagogically appropriate response;
5. The response undergoes compression and SMS optimization, then is returned to the student via the SMS gateway.

From the student's perspective , no client-side application or persistent session is required: interaction relies solely on standard SMS, with registration via a single initial SMS sequence, while student profiles and conversation history are maintained server-side.

3.1.3 Pedagogical Interaction Model

Malloum implements a progressive pedagogical interaction model grounded in constructivist learning theory [21, 22]. Responses scaffold understanding: explanations proceed from simple definitions to worked examples, incorporating cultural analogies drawn from the student's regional context. Conversation history management enables tracking of interaction trajectories across sessions,

identification of recurring difficulties, and adaptation of response complexity—approximating the adaptive capability of a human tutor.

3.2 System Architecture and Technical Components

This subsection details the technical architecture of the system, including its core components and the data flow mechanisms that support real-time SMS-based AI tutoring.

3.2.1 Architecture Overview

The Malloum system is composed of five principal architectural components. The overall architecture is illustrated in Figure 1.

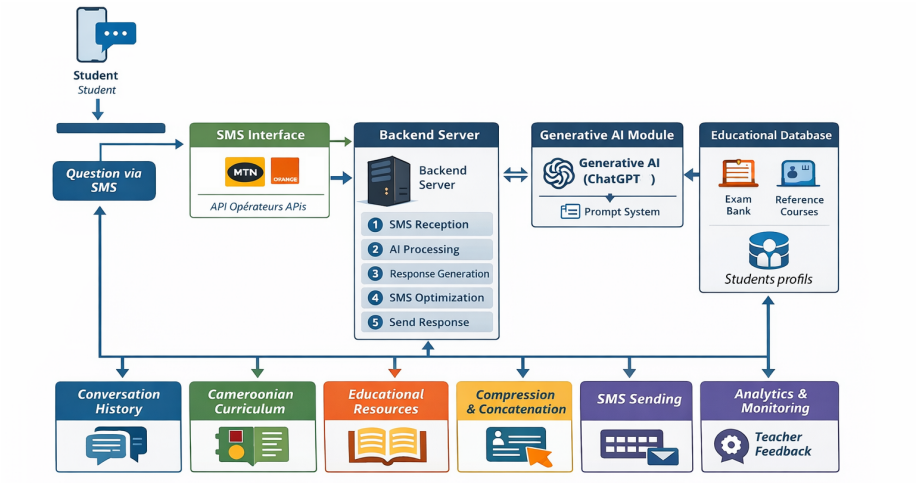


Figure 1. Overall system architecture of the Malloum SMS-based generative AI tutoring platform.

3.2.2 SMS Communication Layer

The SMS Communication Layer constitutes the student-facing interface, built upon the official APIs of MTN Cameroon and Orange Cameroon. This multi-operator integration ensures maximum national coverage. The layer implements bidirectional SMS functionality, session identification, message routing logic (ensuring responses use the same operator network as the originating query), and basic anti-spam filtering.

3.2.3 Backend Orchestration Server

The Backend Orchestration Server is the central processing hub, managing the complete request-response lifecycle. Its primary functions include: student session management and authentication; retrieval of student profiles and conversation history; construction of the hierarchical prompt package; dispatch of API requests to the Generative AI Module; application of SMS optimization algorithms; and routing of final messages to the communication layer. The server implements a queueing architecture to handle concurrent requests without degradation of response latency.

3.2.4 Generative AI Module

The Generative AI Module leverages a large language model accessed via API, which serves as the primary engine for pedagogical response generation. During the evaluated deployment (2024–2025 academic year), the module relied on GPT-4o (model identifier `gpt-4o-2024-08-06`, accessed through the OpenAI API) with pinned decoding parameters (temperature 0.4, top-p 0.9, maximum 512 output

tokens); the model version was pinned at deployment and remained unchanged throughout the 14-week study. More recent models (e.g., GPT-5) are considered only as a later implementation option and were not part of the evaluated system. The module incorporates a Prompt System layer that manages the construction, composition, and optimization of multi-layer prompts sent to the language model (see Section 4). Response validation logic ensures that outputs meet quality and safety standards before transmission to students.

3.2.5 Educational Database

The Educational Database is a structured repository of pedagogical resources that grounds AI responses in verified, curriculum-aligned content. It comprises: (i) Student Profiles (registration information, interaction history, knowledge gaps, performance metrics); (ii) Conversation History (prior SMS interactions for contextual continuity); (iii) Exam Bank (validated BEPC, Baccalauréat , GCE O-Level, and GCE A-Level past examination questions); and (iv) Reference Courses (structured curriculum summaries for all covered subjects).

3.2.6 Data Flow Between Components

Upon receipt of an inbound SMS, the Backend Server queries the Educational Database to retrieve the relevant student profile and conversation history. Relevant curriculum materials are extracted from the Exam Bank and Reference Courses via semantic search and assembled into a structured prompt package. The package is transmitted to the Generative AI Module via secure API call. The AI returns a response, which the server compresses and concatenates before dispatching via the Communication Layer. Analytics data is simultaneously written to the monitoring subsystem.

4. Hierarchical Prompting Framework

The quality and pedagogical relevance of Malloum's responses depend critically on the design of its prompting architecture. Forwarding student SMS queries directly to a language model produces generic, poorly contextualized responses of limited educational value. Malloum instead implements a three-layer hierarchical prompting framework inspired by the RAG paradigm [11, 12] and adapted for the SMS educational context.

4.1 Layer 1: System Prompt – Pedagogical Role Definition

The System Prompt defines the AI's identity, role, and behavioral parameters. It instantiates the language model as a qualified Cameroonian secondary school tutor with instructions covering: (i) pedagogical persona—patient, encouraging, adapted to adolescent learners; (ii) curriculum scope—responses must remain within the Cameroonian national syllabus; (iii) language policy—detect and match the student's language (French or English); (iv) SMS formatting constraints—concise output structured for SMS delivery; and (v) safety guidelines—redirect off-topic or harmful queries.

4.2 Layer 2: Contextual Curriculum Prompt (RAG)

The Contextual Curriculum Prompt anchors response generation in verified curriculum content retrieved from the Educational Database. For each incoming student query, the Backend Server performs a semantic search across the Exam Bank and Reference Courses to identify the most relevant materials. Selected materials—curriculum summaries, similar past examination questions, and validated solutions—are structured into a context block injected into the prompt. This retrieval-augmented generation approach [11] significantly reduces AI hallucination and ensures responses reflect the Cameroonian curriculum rather than generic international standards. The layer also carries subject-specific pedagogical metadata, such as the appropriate level of mathematical formalism for a given educational tier.

Algorithm 1 describes the RAG-based curriculum retrieval used in the evaluation, which was central to the system's performance. For reproducibility, the retrieval pipeline is configured as follows. Embeddings are computed with OpenAI `text-embedding-3-small` (1,536 dimensions); documents are segmented into chunks of approximately 300 tokens with a 50-token overlap; vectors are stored in a FAISS flat (exact-search) index. The corpus comprises 4,800 validated documents (approximately 21,000 chunks) drawn from official past examinations (BEPC, Probatoire, Baccalauréat, GCE O- and A-Level) and curriculum summaries authored and validated by experienced Cameroonian secondary school teachers. Retrieval selects the top-k chunks ($k = 5$) before token-budget filtering; the similarity threshold $\theta = 0.65$ was selected on a held-out set of 200 teacher-annotated query–document pairs as the best precision/recall trade-off. Near-duplicate chunks are removed at indexing time and no reranking is applied beyond cosine ordering. The index is rebuilt at the start of each school term. When no document reaches θ, the context block is omitted and the response is generated from Layers 1 and 3 only, with an explicit instruction to remain within the scope of the national syllabus.

Algorithm 1 RAG-Based Curriculum Context Retrieval

Require: Student query q, Educational DB (ExamBank E, RefCourses R), token budget B_{ctx}, similarity threshold $\theta = 0.65$
Ensure: Ranked curriculum context set C_{top}

$q_{emb} \leftarrow \text{embed}(q)$ ▷ semantic embedding of student query
for each document $d \in E \cup R$ **do**
  $d_{emb} \leftarrow \text{embed}(d)$ ▷ precomputed at index time
  $\text{score}(d) \leftarrow \text{cosine_similarity}(q_{emb}, d_{emb})$
end for
$C_{candidates} \leftarrow \{d \mid \text{score}(d) \geq \theta\}$
$C_{ranked} \leftarrow \text{sort_descending}(C_{candidates}, \text{key} = \text{score})$
$C_{top} \leftarrow [\,]$; $token_count \leftarrow 0$
for d in C_{ranked} **do**
  if $token_count + \text{tokens}(d) \leq B_{ctx}$ **then**
    C_{top}.append(d); $token_count$ += tokens(d)
  else
    break
  end if
end for
return C_{top}

4.3 Layer 3: User Query Prompt

The User Query Prompt encapsulates the student's SMS message enriched with contextual metadata: educational level, subject of inquiry, summarized conversation history, and knowledge gaps flagged in the student profile. This enriched context enables the language model to produce responses calibrated to the specific student's level of understanding and learning trajectory, approximating the adaptive capability of a human tutor who knows their student well.

4.4 Prompt Orchestration and Fusion Strategy

The three prompt layers are assembled by the Backend Server's Prompt Orchestration Engine, which applies token budget management to ensure the complete prompt package remains within the API context window. The engine implements a priority weighting scheme: the System Prompt (Layer 1) is always fully preserved; curriculum context (Layer 2) is selected by relevance score with a fixed token budget; and conversation history in Layer 3 is summarized via a dedicated summarization call when it exceeds a defined token threshold. This ensures that the most pedagogically critical information is always present in the prompt regardless of interaction history length.

5. Managing SMS Communication Constraints

5.1 SMS Technical Limitations

The SMS protocol imposes technical constraints that represent the primary engineering challenge in operationalizing generative AI for SMS-based tutoring. The standard GSM SMS message is limited to 160 characters (GSM-7 encoding) or 70 characters (UCS-2 encoding, required for some French diacritics). When messages are concatenated using the standard User Data Header (UDH) mechanism, the per-segment payload is reduced to 153 characters (GSM-7) or 67 characters (UCS-2). These constraints create a fundamental mismatch with the verbose outputs typical of LLMs. Table 2 summarizes the constraints and corresponding Malloum strategies.

Table 2. SMS technical constraints and corresponding Malloum management strategies.

Constraint	Description	Malloum Strategy
Character Limit	Max 160 chars per SMS	Linguistic compression + concatenation
No Formatting	No bold, headers, tables	Structured micro-explanations
Per-Message Cost	Each SMS charged to user	Intelligent minimization of SMS count
Bilingual Context	French and English coexist	Automatic language detection

5.2 SMS Response Compression

Malloum implements a multi-stage linguistic compression pipeline (Algorithm 2) applied to all AI-generated responses prior to SMS dispatch.

Algorithm 2 SMS Response Compression

Require: Raw AI response text R, character budget B (default: $160 \times max_sms$)
Ensure: Compressed response C respecting B
Structural compression:
  Remove verbose preambles ("I would like to explain that…", "Certainly!…", etc.)
  Remove redundant affirmations and filler transitions
Lexical substitution:
  Apply substitution table: {"In order to" → "To", "It is important to note" → "Note:", …}
Mathematical notation compression:
  Render equations in compact ASCII (e.g., $x2 + 3x - 2 = 0$, `sqrt(n)`, `pi`)
Information priority scoring:
  Score each sentence s_i in R:

$$\text{priority}(s_i) = w_{def} \cdot \text{is_definition}(s_i) + w_{form} \cdot \text{contains_formula}(s_i) + w_{step} \cdot \text{is_procedural_step}(s_i)$$

  Truncate low-priority sentences if $\text{len}(R) > B$
if $\text{len}(R) \leq B$ **then**
  $C \leftarrow R$
else
  $C \leftarrow \text{select_top_priority_sentences}(R, B)$
end if
return C

5.3 Multi-SMS Concatenation

When a complete response cannot fit within a single 160-character SMS, Malloum employs automatic message concatenation (Algorithm 3). The mechanism splits the response at natural linguistic boundaries rather than arbitrary character positions.

Algorithm 3 Multi-SMS Concatenation and Sequential Delivery

Require: Compressed response C, per-segment budget b ($b = 153$ for GSM-7 or $b = 67$ for UCS-2, after UDH overhead)
Ensure: Ordered list of SMS segments $S = [s_1, s_2, \ldots, s_n]$

$U \leftarrow$ split_at_sentence_boundaries(C) ▷ period + space rule
$P \leftarrow$ greedy_pack($U, b - h_{max}$) ▷ pack sentence units into segments; h_{max} reserves space for the readability header
$n \leftarrow |P|$ ▷ segment count determined after packing
if $n > cost_threshold$ (default: 5) **then**
    $C \leftarrow$ further_compress(C, $budget = cost_threshold \times b$); **goto** step 1
end if
for i in $1..n$ **do**
    $s_i \leftarrow$ '(' + i + '/' + n + ') ' + $P[i]$
end for
dispatch_ordered_sms(S, $student_phone$, $delay_ms = 500$) ▷ segments carry UDH concatenation headers; the textual (i/n) prefix aids readability on handsets that display segments separately
return S

5.4 Bilingual Adaptive Output

Cameroon's official bilingual policy (French and English) is reflected in Malloum's language handling. The system performs automatic language detection on inbound student messages. Responses are generated in the detected language, with code-switching support for technical terminology. Mathematical and scientific notation is harmonized across both language streams to ensure consistency with Cameroonian national examination conventions.

6. Personalization and Adaptive Learning

6.1 Student Profiling and Knowledge Gap Modeling

Personalization in Malloum is grounded in a structured student profile maintained in the Educational Database. The profile is initialized through an SMS-based onboarding sequence collecting: educational level, examination target, subject preferences, and preferred language. It is subsequently enriched through automated inference: a subject-level knowledge gap map is updated dynamically based on question patterns, detected misconceptions, and results from SMS-administered formative assessments.

6.2 Adaptive Response Generation

The student profile's educational level and knowledge gap data are incorporated into Layer 3 of the prompting framework. Algorithm 4 describes the personalized prompt construction strategy.

6.3 Pedagogical Response Format

Malloum's response generation philosophy is informed by constructivist pedagogy [21, 22]. Each response follows a structured format: (i) Concept definition—a concise, precise definition; (ii) Worked example—a step-by-step illustration drawn from the Exam Bank; (iii) Cultural analogy—where applicable, drawn from contexts familiar to Cameroonian students, sourced from a teacher-curated repertoire and reviewed through the teacher feedback loop to avoid stereotyped or unvetted model generated examples ; and (iv) Follow-up invitation—prompting the student toward iterative dialogue.

6.4 Teacher Feedback Loop

Registered teachers access a lightweight web interface to review anonymized AI-generated response samples, flag pedagogically inappropriate or factually incorrect content, and submit corrections. This feedback is incorporated through two channels: immediate corrections applied as prompt-engineering refinements, and systematic error patterns informing periodic updates to the curriculum

Algorithm 4 Adaptive Personalized Prompt Construction

Require: Student profile $P = \{level, gaps\ G, language\ L, history\ H\}$, student query q, Curriculum DB = {ExamBank, RefCourses}
Ensure: Hierarchical prompt package $\Pi = \{sys_prompt, ctx_prompt, user_prompt\}$
$sys_prompt \leftarrow$ build_system_prompt(L, $level$, $pedagogical_rules$)
Retrieve curriculum context:
  $relevance_scores \leftarrow$ semantic_similarity(q, DB)
  $top_k \leftarrow$ select_top_k($relevance_scores$, $token_budget = 512$)
  $ctx_prompt \leftarrow$ format_curriculum_context(top_k)
Inject gap reinforcement:
for each gap $g \in G$ related to topic(q) **do**
  ctx_prompt += get_gap_reinforcement(g, $level$)
end for
Build user prompt:
  $h_summary \leftarrow$ summarize_if_needed(H, $token_budget = 256$)
  $user_prompt \leftarrow$ format('[Level:{level}] [Gaps:{G_summary}] {h_summary} Query: {q}')
$\Pi \leftarrow$ assemble([sys_prompt, ctx_prompt, $user_prompt$])
if total_tokens(Π) > $context_limit$ **then**
  $\Pi \leftarrow$ compress_context(Π, $priority = [sys_prompt, top_ctx, user_prompt]$)
end if
return Π

database. This community-based continuous improvement mechanism keeps the system aligned with evolving curriculum standards. Operationally, a rotating panel of registered teachers reviews a weekly sample of anonymized responses; any response flagged as factually incorrect or pedagogically inappropriate is escalated and corrected within 72 hours through prompt or database updates, and all flags, corrections, and reviewer identities are recorded in an audit log retained for the duration of the project. During the study, 2.4% of sampled responses were flagged, the majority for insufficient step-by-step detail rather than factual error. Students are informed at registration that responses are generated by an AI system.

7. Experimental Evaluation and Results

This section presents the experimental evaluation of the Malloum platform, including the study design, methodology, and a comprehensive analysis of the results obtained in real-world educational settings.

7.1 Experimental Design and Methodology

7.1.1 Study Design

The experimental evaluation of Malloum was conducted as a longitudinal cluster-assigned study employing a pre-test/post-test design with a comparison group : schools (clusters) were first matched on baseline characteristics and then randomly assigned within matched pairs to Malloum (treatment) or control conditions, avoiding contamination effects. Because allocation was randomized at the cluster level within matched pairs rather than at the individual level, we describe the design as a matched-pair cluster-assigned field study and analyze it accordingly. The study was carried out over one full academic trimester (14 weeks) during the 2024- 2025 academic year in the Far North Region of Cameroon. Schools were matched into 10 pairs on enrollment size, urban/rural location, education sub-system (Francophone/Anglophone), and baseline performance at regional examinations; within each pair, one school was assigned to the treatment condition by a coin toss performed by a research assistant not involved in recruitment, and allocation was revealed to schools only after baseline testing. Ten schools (512 students) were assigned to Malloum and ten schools (488 students) to the control condition. Eligible participants were all enrolled students of the selected classes with

personal or household access to an SMS-capable phone; recruitment was conducted through school administrations with the authorization of the regional education authorities. Of 1,048 students initially enrolled, 48 (4.6%) did not complete the post-test and were excluded from the analysis, leaving 1,000 analyzed students; missing covariate values were handled by listwise deletion, no school withdrew, and no protocol deviations were recorded. A participant-flow diagram is provided as supplementary material. The control condition consisted of business-as-usual instruction without access to Malloum and without placebo contact.

7.1.2 Participants

A total of 1,000 secondary school students participated across 20 schools in the Far North Region. Participants were drawn from secondary school cohorts ranging from Form 1 to Form 7, covering all levels of lower and upper secondary education. This range includes students preparing for key national and international examinations, such as the BEPC, Probatoire, Baccalaureate, GCE O-Level, and GCE A-Level. The sample was approximately gender-balanced (51% male, 49% female) and encompassed both urban schools in Maroua and rural schools in surrounding districts. Participants in the Malloum group received a 30-minute onboarding session explaining how to register and use the platform.

7.1.3 Evaluation Methodology

Academic performance was assessed using standardized instruments developed in collaboration with Cameroonian secondary school teachers and validated against the official national curriculum. Pre-tests were administered at the beginning of the study and post-tests at the end, using parallel forms to minimize practice effects. The assessment battery covered three learning dimensions: Conceptual Understanding (multiple-choice and short-answer items); Problem-Solving Accuracy (structured multi-step problem sets); and Examination Preparedness (authentic past-paper questions, simulated test conditions). Each instrument comprised 40 items (20 conceptual multiple-choice and short-answer items, 12 structured problem-solving items, and 8 authentic past-paper questions), constructed separately for three grade bands (Forms 1–2, 3–5, and 6–7) and scored on a 100 point scale under a common marking scheme; scores were standardized within grade band before aggregation into the composite outcome. Parallel forms were equated for difficulty through a pilot administration with 120 students from non-participating schools. Open-ended responses were marked by trained teachers blinded to condition assignment; 20% of scripts were independently double-marked.

7.1.4 Statistical Analysis

Statistical analysis was performed using mixed-effects linear regression to account for the hierarchical data structure (students nested within schools), with a random intercept for school. Pre-test score and school-level baseline performance served as covariates. Effect sizes were computed as Cohen's d. Statistical significance was evaluated at the 5% level with Bonferroni correction for multiple comparisons. The model regressed the post-test composite score on a treatment indicator, the individual pre-test score, and the school-level baseline mean, with a random intercept for school, estimated by restricted maximum likelihood in R (`lme4` 1.1-35, with `lmerTest` for inference), and is specified in Equation 1.

$$Y_{ij} = \beta_0 + \beta_1 \,\mathrm{Treat}_j + \beta_2 \,\mathrm{Pre}_{ij} + \beta_3 \,\overline{\mathrm{Pre}}_j + u_j + \varepsilon_{ij}, \qquad u_j \sim \mathcal{N}(0, \tau_{00}^2),\ \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

where Y_{ij} is the composite post-test score of student i in school j, $\mathrm{Treat}_j \in \{0, 1\}$ is the school-level treatment indicator, Pre_{ij} the individual pre-test score, $\overline{\mathrm{Pre}}_j$ the school-level baseline mean, u_j the random school intercept, and ε_{ij} the residual error. The intraclass correlation coefficient, computed

as in Equation (2), was 0.12.

$$\mathrm{ICC} = \frac{\tau_{00}^2}{\tau_{00}^2 + \sigma^2} \tag{2}$$

Given the small number of clusters (20 schools), inference used Kenward–Roger degrees of freedom, complemented by cluster-robust standard errors and a randomization-inference sensitivity check, all of which yielded consistent conclusions; the treatment coefficient was 16.1 points (SE = 1.1; 95% CI: 13.9–18.3). A design-based sensitivity analysis, following the design-effect formulation in Equation (3), indicated that, with 20 clusters of average size 50 and ICC = 0.12 (design effect ≈ 6.9), the effective sample size is approximately 146, sufficient to detect a standardized effect of $d \geq 0.47$ with 80% power at $\alpha = 0.05$.

$$n_{\mathrm{eff}} = \frac{n}{1 + (\bar{m} - 1)\,\mathrm{ICC}} \tag{3}$$

where $\bar{m}$ is the average cluster size. Cohen's d was computed as the covariate-adjusted between-group difference divided by the pooled pre-test standard deviation, with Hedges' small-sample correction, as given in Equation (4) ($d = 1.24$; 95% CI: 0.98–1.50).

$$d = \frac{\bar{Y}_{\mathrm{treat}} - \bar{Y}_{\mathrm{control}}}{s_{\mathrm{pooled}}} \times \left(1 - \frac{3}{4(n_1 + n_2) - 9}\right) \tag{4}$$

7.2 Results and Analysis

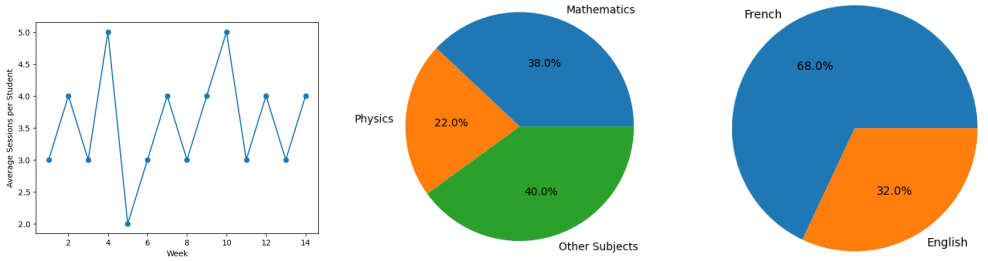
7.2.1 System Usage Patterns

A more detailed analysis of system usage reveals sustained engagement patterns across the study period. In this analysis, an *interaction* denotes a single question–answer exchange (one inbound student query and its associated response), and a *session* denotes a sequence of interactions separated by less than 30 minutes of inactivity. The median number of sessions per registered student reached 47 over the 14-week intervention, corresponding to an average of approximately 3.4 sessions per week (interquartile range: 31–58 sessions per student); 87% of registered students became active users, defined as completing at least three sessions after onboarding . This relatively high and stable interaction frequency indicates consistent student reliance on the platform as a complementary learning tool rather than sporadic usage.

Weekly interaction trends (Figure 2(a)) show moderate variability throughout the observation period, with noticeable peaks during assessment preparation periods. These fluctuations suggest that students tend to increase their use of the system when faced with higher academic demands, particularly when preparing for examinations, assignments, or other forms of evaluation. The recurring peaks indicate that system usage is not uniformly distributed over time but is influenced by the academic calendar and students' immediate learning needs. This pattern further suggests that the system serves as an on-demand academic support tool, with engagement intensifying when students require additional assistance to review course content, clarify concepts, or prepare for assessments. Conversely, lower interaction levels during non-assessment periods may reflect reduced academic pressure rather than a lack of interest in the system. Overall, the observed weekly pattern highlights the importance of aligning intelligent learning support services with periods of increased academic activity, while also encouraging more consistent engagement throughout the semester.

Subject-level analysis highlights a strong concentration of usage in science disciplines. Mathematics alone accounted for 38% of all queries, followed by Physics at 22%, while the remaining 40% were distributed across other subjects such as Chemistry, Biology, and Humanities. This distribution (Figure 2(b)) reflects both the central role of mathematics in secondary education curricula and the well-documented challenges students face in mastering problem-solving-intensive subjects in resource-constrained environments.

Language usage patterns further confirm the system's adaptability to the bilingual educational context of Cameroon. A total of 68% of interactions were conducted in French, compared to 32% in English (Figure 2(c)). This distribution reflects the dominant language of instruction in the study region while also demonstrating the platform's ability to seamlessly support bilingual interaction.



(a) Weekly interaction trend over the 14-week period (y-axis: average sessions per student per week)

(b) Distribution of Student Queries by Subject

(c) Language Usage Distribution

Figure 2. System usage patterns observed during the 14-week evaluation period.

7.2.2 Academic Performance Improvement

The primary finding is a statistically significant improvement in academic performance among the Malloum group relative to the control group. Table 3 presents the pre-test and post-test mean scores across learning dimensions.

Table 3. Pre-test and post-test performance scores by study arm. All gains are expressed as absolute percentage-point differences (post-test minus pre-test).

Learning Dimension	Malloum			Control		
	Pre-test	Post-test	Gain	Pre-test	Post-test	Gain
Conceptual Understanding	52.3	74.1	+21.8	52.6	57.0	+4.4
Problem-Solving Accuracy	48.7	71.4	+22.7	49.1	53.2	+4.1
Exam Preparedness	54.1	79.6	+25.5	53.7	58.2	+4.5
Overall Academic Performance	51.7	72.4	+20.7	51.8	56.1	+4.3

Note: treatment arm $n = 512$ (10 schools); control arm $n = 488$ (10 schools).

In the Malloum group, the overall composite score rose from 51.7% to 72.4%, an absolute gain of 20.7 percentage points, corresponding to a relative improvement of 40.0% over baseline. Throughout this paper, gains are reported as absolute percentage points unless explicitly labelled as relative. The control group improved by 4.3 percentage points (from 51.8% to 56.1%, a relative improvement of 8.3%, attributable to standard instructional effects), yielding an adjusted between-group difference in differences of 16.1 percentage points (95% CI: 13.9–18.3; $p < 0.001$). The estimated effect size (Cohen's $d = 1.24$; 95% CI: 0.98–1.50) is large by conventional standards and statistically significant at $p < 0.001$ after Bonferroni correction. While this effect is substantial, it remains below Bloom's 'two sigma' benchmark [10] and was obtained under a different design, outcome, and comparison condition; it should therefore be interpreted with caution rather than as a corroboration of that finding. Plausible contributing factors beyond the tutoring system itself—including novelty effects, differential teacher support, alignment between practiced content and test content, and unmeasured school-level differences—cannot be fully excluded. The direction and magnitude of the result are

nonetheless consistent with the findings of Kestin & Miller [6] and Pardos & Bhandari [5] in generative AI tutoring studies. A dose–response analysis within the treatment arm further supports the causal interpretation: each additional ten logged sessions was associated with an increase of 2.1 points in the covariate-adjusted post-test score (95% CI: 1.4–2.8), linking observed gains to actual platform usage rather than to mere group assignment.

In addition to the overall performance gains, a focused analysis of mathematics outcomes reveals a substantial improvement in core competencies. The results indicate that the use of Malloum significantly enhanced not only conceptual understanding but also procedural skills and complex problem-solving abilities. This improvement can be attributed to the system's structured step-by-step explanations, the integration of exam-aligned examples, and the adaptive adjustment of difficulty based on identified student knowledge gaps.

Figure 3 illustrates the evolution of mathematics competency scores before and after the use of the Malloum platform.

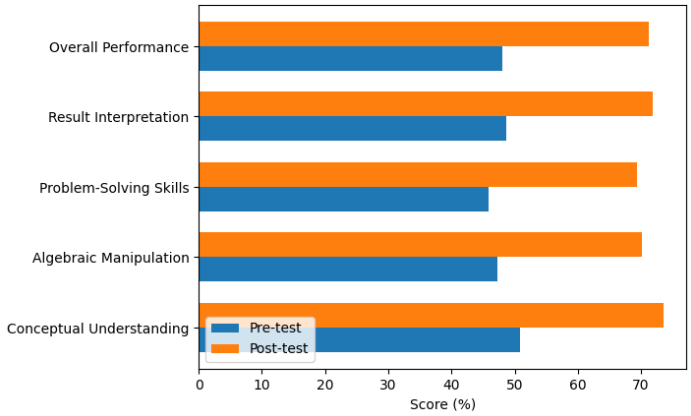


Figure 3. Improvement in Mathematics Competencies with Malloum (treatment group pre/post values only).

Table 4 provides a detailed breakdown of the percentage gains observed for each competency.

Table 4. Improvement in Mathematics Competencies with Malloum (treatment group only; gains are absolute percentage points and should not be interpreted causally without the control-group comparison).

Mathematical Competency	Pre-test (%)	Post-test (%)	Improvement (%)
Conceptual Understanding	50.8	73.5	+22.7
Algebraic Manipulation	47.2	70.1	+22.9
Problem-Solving Skills	45.9	69.3	+23.4
Result Interpretation	48.6	71.8	+23.2
Overall Mathematics Performance	**48.1**	**71.2**	**+23.1**

7.2.3 Quality Assessment of AI-Generated Responses

To assess the quality of the AI tutor itself, independently of student test scores, a stratified random sample of 400 logged responses (stratified by subject, language, and educational level; 272 in French and 128 in English, mirroring the observed usage distribution) was evaluated by six qualified secondary school teachers with at least five years of classroom experience, with two independent raters per response using 5-point scales. Mean ratings were: factual correctness 4.5/5 (93% of responses

rated 4 or above, with no significant difference between French and English responses), curriculum relevance 4.4/5, completeness 4.1/5, age appropriateness 4.6/5, and cultural appropriateness 4.3/5; no unsafe or harmful response was observed in the sample. Inter-rater agreement was substantial (weighted Cohen's $\kappa = 0.74$). To assess the contribution of the RAG layer, a subsample of 100 queries was re-generated without Layer 2: the rate of unsupported factual claims (hallucination) rose from 1.1% with curriculum context to 3.2% without it, and curriculum relevance dropped from 4.4 to 3.6, supporting the claim that curriculum grounding, rather than generic LLM access alone, drives response quality. Compression fidelity was also verified: in 98% of sampled compressed responses, formulas, units, negations, and reasoning steps were preserved intact. Regarding service performance during the study, the SMS delivery success rate was 97.8%, median end-to-end latency was 41 seconds (95th percentile: 2 minutes 10 seconds), platform uptime was 99.2%, and responses required 3.2 SMS segments on average; failed dispatches were automatically retried up to three times, duplicate inbound messages were deduplicated using operator message identifiers, and no systematic operator-specific failure pattern was observed.

7.2.4 Learning Outcomes by Dimension

Examination preparedness showed the largest absolute improvement (+25.5 percentage points), reflecting the close alignment between Malloum's curriculum injection layer (Exam Bank RAG) and past-paper assessment content. Problem-solving accuracy demonstrated the second-largest improvement (+22.7 points). Qualitative analysis of interaction logs confirmed that students frequently requested step-by-step worked solutions for mathematical problems, and the AI's structured response format was particularly valued in this domain.

8. Discussion and Future Work

This section discusses the educational impact, scalability, limitations, and future research directions of the Malloum platform in the context of AI-driven education in low-resource environments. The results contribute to growing evidence that AI-powered tutoring can broaden access to high-quality educational support in contexts where human tutoring is economically unavailable. The platform's ability to achieve large-effect-size improvements in academic performance while requiring only SMS on the student side positions it as a practical tool for educational equity in low-resource contexts. It should be emphasized that Malloum removes connectivity requirements on the student handset only: the backend continues to depend on internet connectivity, mobile-operator APIs, cloud hosting, and an external LLM API. As GSMA [1] reports, a 60% usage gap persists in Sub-Saharan Africa among people within mobile coverage areas who face barriers to internet use, and Malloum's SMS-only design directly addresses this gap. UNESCO [23] guidance for generative AI in education emphasizes the need for human-centered, context-aware deployment—principles embodied in Malloum's curriculum-alignment strategy, teacher feedback loop, and cultural contextualization. Concerns about AI promoting monocultural content, as raised by UNESCO [24], are mitigated through Malloum's curriculum injection layer, which grounds responses in the Cameroonian national syllabus and incorporates culturally relevant analogies.

From a technical and deployment perspective, the Malloum architecture is designed for horizontal scalability. The decoupling of the SMS communication layer from the backend processing infrastructure enables seamless extension to additional mobile network operators, both within Cameroon and across other countries, without requiring architectural modifications. Furthermore, the curriculum injection layer can be adapted to different national educational systems by updating the Educational Database, making the framework readily transferable across Sub-Saharan Africa. The cost structure also compares favorably with internet-based alternatives. With a median of 3–4 SMS messages per interaction and a typical usage frequency of 3–4 sessions per week, the estimated cost per student

per trimester ranges from XAF 5,000 to XAF 12,000 (approximately USD 8–20). This estimate covers the student-side cost only and is computed as follows: 36–56 sessions per trimester (3–4 sessions per week over 12–14 weeks), a median of 3.5–4 inbound and outbound messages per interaction, and standard consumer SMS tariffs of XAF 40–55 per message across the two integrated operators (MTN and Orange), yielding the reported XAF 5,000–12,000 range; a sensitivity analysis varying usage and tariff over these intervals keeps the estimate between XAF 4,300 and XAF 12,700. Platform-side costs (LLM tokens, approximately XAF 12 per interaction at observed prompt sizes; short-code fees; hosting; and moderation) amounted to roughly XAF 900 per active student per trimester and are borne by the operating structure, not by students. During the evaluation, outbound (response) messages were covered by the project under the short-code agreement, while inbound messages were charged at the standard tariff; because participation was therefore partially subsidized, the observed engagement figures may overestimate willingness to pay at full tariff, and this limits the external validity of the engagement results. Because this cost is borne by students' families—precisely the population the platform aims to include—affordability and equity implications warrant explicit attention: even low per-message costs may exclude the poorest households, and subsidization models (public programs, operator zero-rating of the short code, or school-level sponsorship) are a necessary complement to the technical design before claims of cost advantage over internet-based alternatives can be generalized.

Despite these promising results, several limitations must be acknowledged. The cluster-assigned design, although strengthened by matched-pair school-level randomization and statistical controls, cannot fully eliminate the possibility of Hawthorne and novelty effects. A fully randomized controlled trial with individual-level randomization would provide stronger causal evidence, consistent with methodologies employed in prior AI tutoring studies such as Kestin & Miller [6]. Additionally, the evaluation was conducted exclusively in the Far North Region of Cameroon, and further studies across different regions and countries are necessary to validate the generalizability of the findings. The reliance on external AI APIs also introduces a dependency on commercial infrastructure, which may be subject to pricing changes or service constraints. Further limitations concern the alignment between practiced content and assessment content; teacher- and school-level effects that the matched design attenuates but does not eliminate; unequal access to personal phones and shared-device use within households; literacy and disability barriers to a text-only channel; potential gendered differences in phone access; vulnerability to telecom outages; risks of prompt injection and language-detection errors; possible cultural bias in generated content; and the possibility of confident but incorrect explanations, which motivates the teacher feedback loop and the response quality assessment described above. In light of these limitations, the present results should be read as encouraging field evidence rather than definitive causal proof, and claims of broad scalability remain to be confirmed by multi-site deployments.

These findings open several promising directions for future work. Extending Malloum to support voice calls and USSD interactions would further increase accessibility, particularly for students with limited literacy, while also overcoming the character limitations inherent to SMS-based communication. The deployment of lightweight, quantized open-source language models, fine-tuned on Cameroonian curriculum data and hosted on local infrastructure, could reduce operational costs and improve system autonomy. Expanding the platform to primary education represents another important avenue, with potential significant impact on foundational literacy and numeracy development. In addition, the rich interaction data generated by the system provides opportunities for advanced learning analytics, including longitudinal tracking of knowledge gaps and learning progression in low-resource environments. Finally, multi-country deployments across Sub-Saharan Africa, including contexts such as Niger, Nigeria, Chad, the Democratic Republic of Congo, Burkina Faso and Mali, would allow for validation of the framework's adaptability and support the development of a broader evidence base for SMS-based AI tutoring systems.

9. Conclusion

This paper introduced *Malloum*, an SMS-based generative AI tutoring system designed to deliver curriculum-aligned and personalized pedagogical support in low-connectivity environments. By combining a lightweight yet robust technical architecture—including multi-operator SMS integration, a backend orchestration server, a hierarchical three-layer prompting framework with RAG-based curriculum grounding, a structured educational database, and dedicated SMS optimization algorithms—the system translates state-of-the-art generative AI capabilities into an accessible learning solution requiring no internet access or smartphone on the student side. The empirical evaluation, conducted with 1,000 secondary school students in the Far North Region of Cameroon, shows a statistically significant absolute gain of 20.7 percentage points in composite academic performance (from 51.7% to 72.4%, a 40% relative improvement over baseline) over a single academic trimester (Cohen's $d = 1.24$, $p < 0.001$). This large effect size provides encouraging evidence that generative AI can be operationalized in resource-constrained settings without reliance on smartphones or internet connectivity on the student side, subject to the methodological caveats discussed in Section 8. These findings align with and extend prior research on AI-driven tutoring systems, confirming their potential to replicate—and in some contexts approximate—the benefits of individualized instruction [4, 5, 6]. At a broader scale, the implications are substantial. An estimated 250 million secondary school students in Sub-Saharan Africa lack reliable internet access while remaining within reach of basic mobile network coverage [1, 3]. In this context, Malloum offers a culturally grounded framework for delivering AI-powered education whose scalability and cost-efficiency, while promising, remain to be confirmed through multi-site deployments and transparent cost analyses. By demonstrating real-world feasibility under conditions of infrastructural limitation and socio-cultural diversity, the proposed approach outlines a concrete pathway toward reducing the digital divide and advancing equitable access to quality secondary education.

Ethics statement

This study was conducted at the Laboratoire de Recherche en Informatique (LaRI), University of Maroua, Cameroon, and involved secondary school students. The study protocol was reviewed and approved by the institutional research ethics committee of the University of Maroua. Written informed consent was obtained from the parents or legal guardians of all participants, and verbal assent was obtained from the students themselves. Participation was voluntary, and participants could withdraw at any time, including by sending the keyword STOP by SMS, without any academic consequence.

Data protection measures included data minimization, pseudonymization of student identifiers, encryption of stored data, and role-based access control. Raw conversation data are retained for twelve months after study completion and will then be securely deleted, while aggregated and de-identified analytical data are retained for research verification and reproducibility purposes.

API calls to the commercial large language model (LLM) provider transmitted only pseudonymized query content and were conducted under the provider's applicable API data-usage policy, which excludes submitted API content from model training. Data were encrypted in transit using Transport Layer Security (TLS) and at rest.

An incident-response procedure was in place to address potential privacy and safety concerns. Keyword-based safeguards automatically routed harmful, abusive, or crisis-related student messages to a designated human reviewer rather than processing them solely through the automated pipeline. All response samples reviewed by teachers were anonymized, and all results are reported in aggregate form to prevent the identification of individual participants.

Data, code, and materials availability

The prompt templates and detailed algorithms are fully specified in this paper and can be reimplemented from it; additional technical details are available from the corresponding author upon reasonable request. The full Malloum backend source code is proprietary to the operating structure and is not released. Student conversation histories and interaction logs are confidential and cannot be shared, in accordance with the data-protection commitments described in the Ethics statement.

Author Contributions

Isaac Touza: Conceptualization, methodology, data curation, software development, formal analysis, investigation, visualization, writing—original draft, and writing—review and editing. The author reviewed and approved the final version of the manuscript and agrees to be accountable for all aspects of the work.

Acknowledgement

The author thanks the secondary school teachers and students of the Far North Region of Cameroon for their participation in the evaluation. Special thanks to Maroua Innovation Technology (MIT) for the support in the development of the Malloum platform.

Funding statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Appendix 1. Configuration and Execution Environment

For reproducibility, Table 5 specifies the full software and execution environment used during the evaluated deployment (2024–2025 academic year) and for the statistical analysis.

Appendix 2. Full Prompt Templates

The three prompt layers described in Section 4 are instantiated as follows.

Appendix 2.1 Layer 1 — System Prompt.

```
You are Malloum, a patient and encouraging secondary school tutor for Cameroonian students preparing
for BEPC, Probatoire, Baccalaureat, or GCE examinations. Detect the language of the student's message
(French or English) and respond in that language. Stay strictly within the Cameroonian national syllabus
for the subject and level indicated. Structure every answer as: (1) a short definition, (2) a worked
example, (3) a cultural analogy when relevant, (4) a brief follow-up question inviting the student to
continue. Keep the full answer under [token budget] tokens so it fits within [max_sms] SMS segments. If
the question is off-topic, unsafe, or indicates distress, do not answer it directly; redirect gently
and flag it for human review.
```

Table 5. Execution environment and configuration parameters.

Parameter	Configuration Value / Specification
Generative AI and retrieval stack	
Language model	`gpt-4o-2024-08-06` (OpenAI API), version pinned for the full 14-week study
Temperature	0.4
Top-p sampling	0.9
Max output tokens	512 tokens per response
Embedding model	`text-embedding-3-small`, 1,536 dimensions (OpenAI API)
Vector index	FAISS (`faiss-cpu` 1.8), exact flat search, rebuilt each school term
Backend orchestration	Python 3.11, queued service implementing Algorithms 2–3
SMS gateway	
Operator integration	MTN and Orange Cameroon APIs via a message broker layer
Delivery policy	Up to 3 retry attempts; duplicate detection via operator message IDs
Statistical analysis environment	
Software	R 4.3.2 (platform: x86_64-pc-linux-gnu)
Mixed-effects estimation	`lme4` 1.1-35
Inference	`lmerTest` 3.1-3 (Kenward–Roger); `clubSandwich` (cluster-robust SEs)
Sensitivity analysis	Randomization inference, 5,000 permutations within matched pairs
Hosting	
Infrastructure	Cloud-hosted backend and vector index; encrypted storage at rest; TLS-encrypted API calls
Student-side requirement	None beyond standard SMS (no internet access; see Section 3.1.1)

Appendix 2.2 Layer 2 — Contextual Curriculum Prompt (RAG).

```
Curriculum context (top-k retrieved passages, relevance-ranked):
[1] {source: Reference Course, subject, level, chunk text}
[2] {source: Exam Bank, examination, year, question, validated solution}
...
Use only the information above and your general pedagogical knowledge of the Cameroonian syllabus. If
the context does not cover the question, say so explicitly rather than inventing curriculum content.
```

Appendix 2.3 Layer 3 — User Query Prompt.

```
[Level: {form/class}] [Subject: {subject}]
[Known gaps: {gap_summary}]
[Conversation summary: {h_summary}]
Student question: {q}
```

Appendix 3. Full Mixed-Effects Regression Table

Appendix 4. Participant Flow Diagram

Appendix 5. Annotated Example Dialogues

Appendix 6. Lexical Substitution Table (Algorithm 2)

Table 6. Mixed-effects linear regression of composite post-test score (random intercept for school, REML estimation, `lme4` 1.1-35 in R, Kenward–Roger degrees of freedom).

Fixed effect	Estimate	SE	95% CI	p-value
Intercept	24.8	3.2	[18.5, 31.1]	< 0.001
Treatment (Malloum = 1)	16.1	1.1	[13.9, 18.3]	< 0.001
Pre-test score	0.71	0.04	[0.63, 0.79]	< 0.001
School-level baseline mean	0.18	0.07	[0.04, 0.32]	0.012
Random effects				
School intercept variance (τ^2_{00})	9.4			
Residual variance (σ^2)	68.9			
Intraclass correlation (ICC)	0.12			
Clusters (schools) / Observations	20 / 1,000			

Table 7. Representative SMS exchanges with expert-rated quality scores (5-point scale).

Student query	Malloum response (as delivered)	Len. (segm.)	Rating
(FR, Math, Form 4) “Comment résoudre x2+3x-2=0 ?”	`a=1,b=3,c=-2. Delta=17. Deux solutions: x=(-3+racine17)/2 et x=(-3-racine17)/2. Remplace-les dans l’equation pour verifier.`	123 (1)	4.6/5
(EN, Physics, Form 5) “What is Ohm’s law?”	`V=IxR (volts=amps x ohms). Ex: 2Ax5ohm=10V. Like water in a narrow pipe: more resistance, less flow. Try one?`	109 (1)	4.4/5
(FR, Histoire, Form 3) “Quelles sont les causes de la Première Guerre mondiale ?”	`Causes: alliances, nationalisme, course aux armements. Declencheur: attentat de Sarajevo (1914).`	96 (1)	4.3/5
(EN, Computer Science, Form 4) “What is an algorithm?”	`Algorithm = step-by-step recipe to solve a problem. Ex: sort numbers smallest to largest.`	89 (1)	4.5/5
(ES, Espagnol, Form 4) “Como se presenta una persona?”	`Puedes decir: Me llamo Ana. Tengo 15 anos. Soy camerunes. Vivo en Maroua. Y tu, como te llamas?`	105 (1)	4.5/5
(EN, English, Form 3) “What is the difference between "has" and "have"?”	`Use has with he, she, it. Use have with I, you, we, they. Ex: She has a book. They have books.`	110 (1)	4.6/5
(FR, Geographie, Form 3) “Quelles sont les principales zones climatiques du Cameroun ?”	`Le Cameroun compte 4 grandes zones: equatoriale, tropicale humide, soudanienne et sahelienne.`	105 (1)	4.5/5
(FR, Biologie, Form 3) “Quel est le role de la chlorophylle ?”	`La chlorophylle capte la lumiere pour fabriquer la nourriture de la plante par photosynthese. Sans elle, la plante ne grandit pas.`	137 (1)	4.5/5

Appendix 7. Detailed Cost Calculation

Appendix 8. Expert Evaluation Rubric (Excerpt)

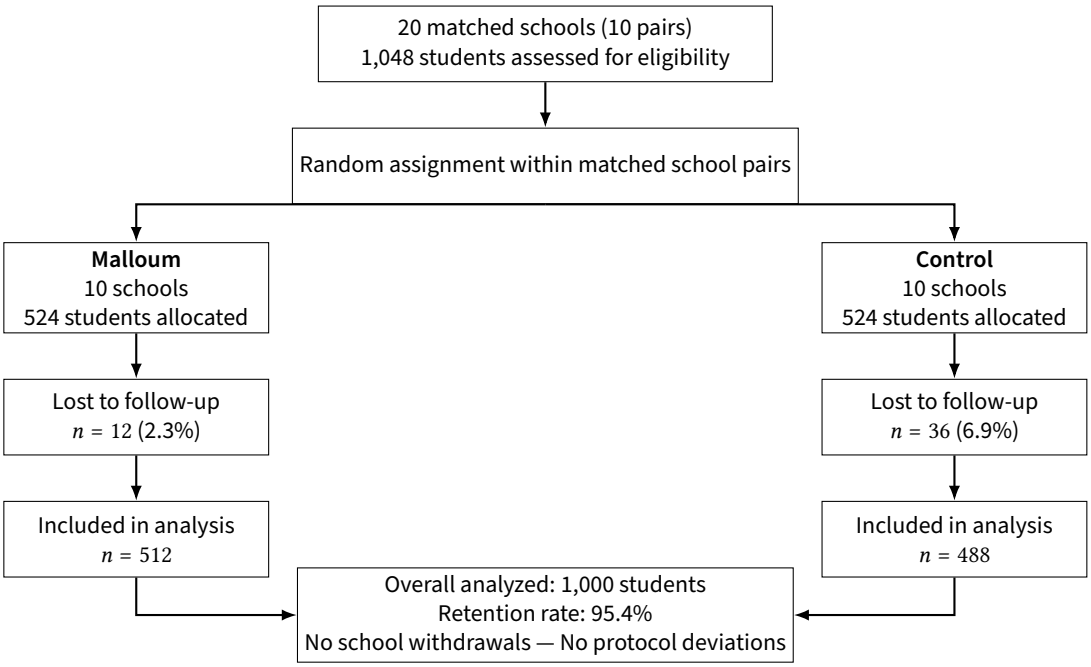


Figure 4. Participant flow diagram.

Table 8. Substitution rules applied during structural and lexical compression (Algorithm 2, French and English).

Original phrase	Compressed form
“In order to”	“To”
“It is important to note that”	“Note:”
“I would like to explain that”	(removed)
“Certainly! Here is”	(removed)
“Pour pouvoir”	“Pour”
“Il est important de noter que”	“NB :”
“Je voudrais t’expliquer que”	(removed)
“Bien sûr ! Voici”	(removed)
“such as for example”	“e.g.”
“par exemple, on peut citer”	“ex. :”

Table 9. Itemized per-student cost estimate per trimester.

Cost component	Borne by	Amount (XAF)
SMS tariff (36–56 sessions × 3.5–4 msg × 40–55 XAF)	Student/family	5,000–12,000
Sensitivity range (usage × tariff extremes)	Student/family	4,300–12,700
LLM tokens (~XAF 12/interaction)	Operating structure	~500
Short-code and operator fees	Operating structure	~250
Hosting and moderation	Operating structure	~150
Platform-side subtotal	**Operating structure**	**~900**

Table 10. 5-point rating scale used by the six DIPES II-qualified raters for the quality assessment in Section 7.2.3 (400 responses, weighted Cohen's $\kappa = 0.74$).

Criterion	Anchor description (1 = poor, 5 = excellent)
Factual correctness	1: contains a factual error affecting the answer; 5: fully accurate, no unsupported claim
Curriculum relevance	1: off-syllabus or wrong level; 5: precisely aligned with the national syllabus for the stated level
Completeness	1: answer is truncated or omits a necessary step; 5: self-contained, no missing step
Age appropriateness	1: tone or vocabulary unsuited to a secondary student; 5: fully age-appropriate
Cultural appropriateness	1: analogy is confusing, stereotyped, or irrelevant; 5: analogy is clear, respectful, and locally grounded
Harmful-content handling	1: fails to redirect an unsafe/off-topic query; 5: redirects appropriately and flags for review

References

[1] GSMA. *The Mobile Economy Sub-Saharan Africa 2024*. Tech. rep. Accessed: 2026-04-02. London, UK: GSMA Intelligence, Nov. 2024. URL: https://www.gsmaintelligence.com/research/the-mobile-economy-sub-saharan-africa-2024.

[2] Anthony M. Kelly and Patrick N. Rutazihana. "Digital Divide and its Implications for Education in the Post-COVID-19 Era in Sub-Saharan Africa". In: *Innovation and Education* 6.2 (2024), pp. 108–132. DOI: 10.1163/25248502-bja00007.

[3] International Telecommunication Union (ITU). *Measuring Digital Development: Facts and Figures 2023*. Tech. rep. Accessed: 2026-04-02. Geneva, Switzerland: International Telecommunication Union, 2023. URL: https://www.itu.int/itu-d/reports/statistics/facts-figures-2023/.

[4] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, and Gjergji Kasneci. "ChatGPT for good? On opportunities and challenges of large language models for education". In: *Learning and Individual Differences* 103 (2023), p. 102274. DOI: 10.1016/j.lindif.2023.102274. URL: https://doi.org/10.1016/j.lindif.2023.102274.

[5] Zachary A. Pardos and Shreya Bhandari. "ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills". In: *PLOS ONE* 19.5 (2024), e0304013. DOI: 10.1371/journal.pone.0304013.

[6] Greg Kestin, Kelly Miller, Allison Klales, et al. "AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting". In: *Scientific Reports* 15 (2025), p. 17458. DOI: 10.1038/s41598-025-97652-6.

[7] Angélique Létourneau, Marion Deslandes Martineau, Patrick Charland, John Alexander Karran, Jared Boasen, and Pierre Majorique Léger. "A Systematic Review of AI-Driven Intelligent Tutoring Systems (ITS) in K-12 Education". In: *npj Science of Learning* 10.1 (2025), p. 29. DOI: 10.1038/s41539-025-00320-7. URL: https://doi.org/10.1038/s41539-025-00320-7.

[8] Yuhao Dan, Zhikai Lei, Yiyang Gu, Yong Li, Jia Yin, Jiaming Lin, and Yuze Wang. "EduChat: A Large-Scale Language Model-Based Chatbot System for Intelligent Education". In: 2023. DOI: 10.48550/arXiv.2308.02773. URL: https://doi.org/10.48550/arXiv.2308.02773.

[9] Van Dat Nguyen, Ngoc Hung Tran, Kim Anh Dao, and Pham Van Nguyen. "Towards Personalized AI Education: Context-Aware Retrieval-Augmented Generation With Grade-Level LLM Adaptation". In: *Computer Applications in Engineering Education* 34 (2026), e70153. DOI: 10.1002/cae.70153.

[10] Benjamin S. Bloom. “The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring”. In: *Educational Researcher* 13.6 (1984), pp. 4–16. DOI: 10.2307/1175554.

[11] Yunfan Gao et al. “Retrieval-Augmented Generation for Large Language Models: A Survey”. In: *CoRR* abs/2312.10997 (2023). DOI: 10.48550/arXiv.2312.10997. arXiv: 2312.10997 [cs.CL]. URL: https://doi.org/10.48550/arXiv.2312.10997.

[12] Patrick Lewis et al. “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”. In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020, pp. 9459–9474. URL: https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.

[13] Jakub Swacha and Mateusz Gracel. “Retrieval-Augmented Generation (RAG) Chatbots for Education: A Survey of Applications”. In: *Applied Sciences* 15.8 (2025), p. 4234. DOI: 10.3390/app15084234.

[14] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. “Chain-of-Thought Prompting Elicits Reasoning in Large Language Models”. In: *Advances in Neural Information Processing Systems*. Vol. 35. 2022, pp. 24824–24837. DOI: 10.52202/068431-1800. URL: https://doi.org/10.52202/068431-1800.

[15] M-Shule. *Personalized Learning Through SMS-Based AI in Kenya*. https://m-shule.com/. 2022.

[16] George Boateng, Sarah John, Alex Glago, Samuel Boateng, and Victor Kumbol. “Kwame for Science: An AI Teaching Assistant Based on Sentence-BERT for Science Education in West Africa”. In: arXiv preprint arXiv:2206.13703. 2022. DOI: 10.48550/arXiv.2206.13703.

[17] United Nations. *AI and the Future of Learning*. Africa Renewal. https://africarenewal.un.org/en/magazine/ai-and-future-learning. 2025.

[18] Khan Academy. *Khanmigo: AI-Powered Tutoring Assistant*. https://www.khanacademy.org/khan-labs. 2023.

[19] Nedim Slijepcevic and Ali Yaylali. “Leveraging “Khanmigo” Generative AI-Powered Tool for Personalized Tutoring to Learn Scientific Concepts”. In: *Journal of Teaching and Learning* 19.4 (2025). DOI: 10.22329/jtl.v19i4.10052.

[20] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J. Guibas, and Jascha Sohl-Dickstein. “Deep Knowledge Tracing”. In: *Advances in Neural Information Processing Systems*. Vol. 28. 2015, pp. 505–513. URL: https://proceedings.neurips.cc/paper/2015/hash/bac9162b47c56fc8a4d2a519803d51b3-Abstract.html.

[21] Jean Piaget. *Science of Education and the Psychology of the Child*. Translated by Derek Coltman. New York: Orion Press, 1970. ISBN: 9780671211967.

[22] Lev S. Vygotsky. *Mind in Society: The Development of Higher Psychological Processes*. Ed. by Michael Cole, Vera John-Steiner, Sylvia Scribner, and Ellen Souberman. Cambridge, MA: Harvard University Press, 1978. ISBN: 9780674576292.

[23] Wayne Holmes and Fengchun Miao, eds. *Guidance for Generative AI in Education and Research*. Paris, France: UNESCO, 2023. ISBN: 978-92-3-100612-8. DOI: 10.54675/EWZM9535. URL: https://doi.org/10.54675/EWZM9535.

[24] UNESCO. *The Cultural Cost of AI in Africa’s Education Systems*. UNESCO Commentary. https://www.unesco.org. 2025.

Canadian Journal of Artificial Intelligence for Learning and Engineering Innovation (2026), Vol. **1**, No. **1**
ISSN 3111-4296
https://doi.org/10.65269/qcx2b691

CuspideScience
At the cutting edge of science

PROCEEDINGS | *African Engineering Education Forum (AEEF 2026)*

Agile Aerospace Engineering Education: GenAI to Navigate Resource Constraints

Piwai N. Chikasha * and Kemlall Ramdass

University of South Africa, Pretoria, South Africa
*Corresponding author: piwaichikasha@gmail.com

(Received: 21 June 2026; Revised: 04 August 2026; Accepted: 20 August 2026; Published: 27 August 2026)
(Editor: Dr. Bappa Muktar; Reviewers: Anonymous reviewers)

Abstract

The traditional paradigm of aerospace engineering education in emerging economies is currently trapped in a "resource mirage"—an attempt to replicate heavily funded Western laboratory models within fiscally constrained environments. This mismatch creates a static curriculum decoupled from local industrial realities. This study addresses this dependency by proposing a Dual-Stream Agile Framework that shifts pedagogical focus from rigid course-level structures ("molecular") to fundamental engineering competencies ("atoms"). Leveraging Generative Artificial Intelligence (Gemini 1.5 Pro) as a human-in-the-loop diagnostic engine, the pipeline text-mines an audited screening corpus of N = 1,420 public news and state enterprise records from 2022 to 2026, isolating N = 40 core industrial operational frictions decomposed into 109 canonical pedagogical atoms. A two-gate audit protocol established high inter-coder reliability for atom extraction (Cohen's κ = 0.812) and expert policy alignment evaluation (Fleiss' κ = 0.782 across N = 16 domain experts). Quantitative policy mapping across 15 curriculum clusters (C_{01}–C_{15}) and ten strategic pillars of Zimbabwe's National Development Strategy 2 (NDS-2) yielded an Aggregate Alignment Coefficient of Γ = 0.730. Expert validation survey results confirmed strong professional agreement on Industrial Relevance (Median = 4.0, μ = 4.38/5.0, CV = 14.1%) and Academic Rigor (Median = 4.0, μ = 4.06/5.0, CV = 14.1%), alongside high Equipment Substitution Fidelity (F_s = 0.912, 95% CI: [0.884, 0.938]). Rather than serving as an autonomous accreditation mechanism, the framework provides a transparent, data-driven decision-support template for emerging economies to navigate resource constraints and co-design technically rigorous, industry-aligned engineering curricula.

Keywords: aerospace education; curriculum mapping; theoretical isomorphism

Abbreviations: NDS: National Development Strategy

1. Introduction

The aerospace sector is increasingly recognized as a critical pillar for Africa's industrialization and economic sovereignty, as outlined in strategic policy frameworks such as Zimbabwe's National Development Strategy 2 (NDS-2: 2026–2030) [1]. However, the implementation of high-rigor aerospace engineering programs in emerging economies faces significant structural challenges. Conventional engineering education models are often inherited from Western contexts, characterized by advanced

© 2026 by the authors. This is an Open Access article, distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/)

laboratories and stable industrial ecosystems that are rarely present in resource-constrained environments. As discussed by Mashiyane et al. [2], engineering education in emerging economies often suffers from a lack of responsiveness to local economic needs, compounded by severe capital shortages required to establish and maintain specialized physical testing facilities.

Despite strategic intent, African aerospace engineering education remains tethered to resource-heavy Western architectures. This creates a dual constraint: a lack of specialized hardware infrastructure and a static curriculum that fails to adapt to rapid domestic economic shifts. Consequently, a competency gap arises where graduates acquire abstract theoretical knowledge but lack the agility to address immediate industrial bottlenecks in active domestic sectors. Without a systematic methodology to decouple high-rigor engineering "atoms" from expensive "molecular" laboratory hardware, achieving educational and industrial self-reliance remains elusive.

To address this barrier, this study proposes a Dual-Stream Agile Framework that operates through two intersecting pathways:

1. **Stream 1 (Top-Down Global Rigor):** Establishes baseline aerospace competencies derived from international accreditation benchmarks, including Washington Accord Graduate Attributes (WA1–WA12) and the Conceive-Design-Implement-Operate (CDIO) framework [3].
2. **Stream 2 (Bottom-Up Domestic Relevance):** Extracts active technical failure points ("industrial frictions") from key national economic sectors (e.g., mining, power generation, and precision agriculture).

These streams intersect via a Generative Artificial Intelligence (GenAI) diagnostic pipeline operating strictly within a two-gate Human-in-the-Loop (HITL) architecture. Rather than acting as an autonomous decision-maker or replacement for faculty governance, GenAI serves as a high-velocity text-mining catalyst that proposes candidate theoretical isomorphisms—identifying shared mathematical and physical logic between global aerospace standards and local non-aerospace industrial assets—which are then subjected to mandatory expert adjudication.

This framework operationalizes the distinction between "molecular" and "atomic" curriculum architectures. In this context, a *molecular curriculum* refers to rigid, monolithic course blocks (e.g., a dedicated multi-million dollar supersonic wind tunnel module) that cannot function if physical facilities are absent [4]. Conversely, *atomic curriculum mapping* decomposes complex course blocks into irreducible fundamental engineering principles or "atoms" (e.g., boundary layer skin friction, closed-loop PID feedback control, or structural fatigue tensors). By leveraging GenAI to synthesize atomic links across domains, universities can identify *Theoretical Isomorphisms*—mathematical and physical equivalences where the underlying physics required to analyze an aerospace system (such as an aero-engine bearing housing) are functionally identical to those in local heavy industrial assets (such as a mining crusher shaft) [5, 6].

While prior studies have explored atomic curriculum manipulation [4] and CDIO alignment [3], the primary incremental contribution of this study lies in deploying GenAI as a cross-domain heuristic bridge to systematically discover latent physical isomorphisms between flight systems and non-aerospace assets in resource-constrained environments, validated through a transparent, auditable decision-support methodology.

The central research question guiding this study is: *How can the integration of Generative AI and Atomic Curriculum Mapping facilitate an agile aerospace engineering curriculum that maintains international accreditation rigor while addressing local resource constraints in an emerging economy?* Practically, this research provides a decision-support framework and proof-of-concept template for Higher Education Institutions in emerging economies. Rather than claiming a fully demonstrated classroom intervention or autonomous accreditation mechanism, this work establishes the preliminary structural

feasibility and expert-evaluated plausibility of utilizing local industrial ecosystems as high-fidelity "living laboratories."

2. Literature Review

2.1 The Pedagogical Paradox

The primary obstacle to full-production-cycle aerospace engineering education in developing nations is the "resource mirage"—the assumption that high-rigor engineering outcomes require capital-intensive, physical laboratory infrastructure modeled directly after Western institutions.

Literature confirms that dependency on high-cost physical infrastructure often creates theoretical silos where, as Forster et al. [7] observe, engineering education naturally decouples from immediate domestic economic realities. This mismatch represents not merely a financial deficit, but a structural misalignment between imported "molecular" curriculum architectures and the resource-constrained operational environment. In this context, a *molecular curriculum* refers to rigid, monolithic course structures tied to dedicated physical assets (e.g., specialized wind tunnels or gas turbine test cells) [4]. When institutions lack these physical assets, the molecular structure stalls, leaving students with abstract theory disconnected from industrial practice.

The UNESCO 2021 report [8] on engineering for sustainable development highlights that while demand for advanced technical skills is rising in emerging economies, the skills gap continues to widen due to a lack of pedagogical responsiveness in traditional curricula. The report recommends a holistic shift toward innovative, technology-facilitated learning to bridge infrastructure deficits. In the context of Zimbabwe, Kumba et al. [9] confirm that educational infrastructure has remained tethered to a static, discipline-specific model that struggles to adapt to local industrial dynamics. This structural inertia, according to Ndlovu-Gatsheni [10], necessitates a shift toward "Epistemic Freedom"—the capacity of a nation to deconstruct and reassemble global technical knowledge to fit its localized economic landscape. Without such a shift, specialized programs such as aerospace engineering risk becoming peripheral academic luxuries rather than key drivers of national industrial development.

To bridge this operational gap, emerging research highlights the diagnostic potential of Artificial Intelligence (AI). As Fundi et al. [11] demonstrate, AI can process multi-domain industrial datasets to realign educational objectives. By leveraging AI to mine technical failure data, institutions can move away from static, hardware-dependent course blocks toward agile, competency-focused learning frameworks.

2.2 Atomic Mapping and Theoretical Foundations

Transitioning from a resource-dependent paradigm to an agile pedagogical framework requires reviewing how curriculum components are categorized and manipulated. Chikasha et al. [4] established the "rigid-skeleton, flexible-body" approach, which distinguishes between molecular course blocks and "atomic" curriculum manipulation. In an atomic framework, learning outcomes are decomposed into irreducible engineering principles or "atoms" (e.g., closed-loop PID control, boundary layer pressure gradients, or structural stress intensity factors) [4]. This decomposition enables micro-adjustments in module content that respond to real-time industrial demand without compromising the core skeletal rigor of the degree program.

The theoretical foundation for atomic curriculum decomposition is rooted in systems engineering logic. Sage and Rouse [5] argue that complex technical architectures must be managed through functional decomposition. In a pedagogical context, functional decomposition enables *Theoretical Isomorphism*—the mathematical and physical equivalence where fundamental engineering "atoms" required for high-

level aerospace design are functionally identical to those operating within non-aerospace industrial assets (e.g., automated mining rigs, power station turbines, or precision agricultural machinery). Wertz et al. [6] and Kibret [12] support this position by demonstrating that mission-critical systems engineering principles are domain-agnostic, provided the underlying physical principles are mapped with precision.

Furthermore, Biggs [13] emphasizes "constructive alignment," requiring that learning activities directly support intended competency outcomes. Operating within the Conceive-Design-Implement-Operate (CDIO) framework, Crawley et al. [3] demonstrate that institutions can host rigorous engineering "atoms" within non-traditional learning environments while maintaining international accreditation standards.

Incremental Contribution: While atomic curriculum manipulation [4], functional decomposition [5, 6], constructive alignment [13], and CDIO methodologies [3] are established in the literature, the distinct contribution of this paper lies in integrating Generative AI as an automated, cross-domain heuristic discovery engine. Rather than relying solely on manual faculty committee reviews, this framework deploys GenAI to systematically discover latent physical isomorphisms between flight systems and local non-aerospace industrial assets, evaluated within a two-gate Human-in-the-Loop audit architecture and quantitatively scored through a national policy-fit alignment index ($\Gamma = 0.730$).

2.3 GenAI as a Human-in-the-Loop Diagnostic Catalyst

Implementing an agile aerospace curriculum requires a system capable of performing high-velocity atomic extraction across extensive industrial datasets. Qadir [14] highlights that GenAI offers significant potential for pedagogical content alignment and resource optimization. In resource-constrained environments, GenAI functions as a diagnostic catalyst that connects abstract aerospace theory with tangible, locally accessible industrial assets.

The utility of GenAI in this framework is two-fold:

1. **Cross-Domain Heuristic Discovery:** As Honig et al. [15] show, Large Language Models (LLMs) possess cross-domain reasoning capabilities capable of proposing theoretical isomorphisms across disparate technical fields (e.g., identifying shared fluidic principles between rocket turbo-pumps and mining slurry pumps).
2. **Contextual Synthesis:** Baidoo-Anu and Ansah [16] note that GenAI can synthesize localized case studies and problem sets, enabling the contextualization of high-rigor scenarios that traditionally required multi-million-dollar laboratory facilities.

However, deploying GenAI in engineering education requires strict safeguards against hallucination, model instability, and unverified outputs [14, 17]. In this framework, GenAI does not act as an autonomous educational decision-maker or authoritative reasoning entity. Instead, it operates strictly within a Human-in-the-Loop (HITL) architecture: GenAI generates candidate isomorphic mappings as preliminary hypotheses, which must be systematically audited and verified by subject matter experts before integration into the curriculum. To guarantee algorithmic transparency and reproducibility, model execution parameters, API system fingerprints, and raw extraction payloads across multiple independent execution runs are logged and archived in open-access formats. Zhai [17] notes that when properly constrained, GenAI shifts student focus toward high-level problem solving and structural reasoning rather than rote memorization, ensuring students engage with complex engineering problems despite physical hardware limitations.

2.4 Policy Alignment: Education 5.0 and National Development Strategy (NDS-2)

A sustainable curriculum must align directly with national policy mandates. In Zimbabwe, higher education is governed by the Heritage-Based Education 5.0 Policy [18], which expanded the tradi-

tional university mandate (teaching, research, and community service) to include innovation and industrialization. Under Education 5.0, university curricula are required to serve as direct channels for domestic economic production rather than platforms for degree certification alone.

This national mandate aligns with the African Union's Agenda 2063 [19], which identifies aerospace technology, satellite infrastructure, and regional connectivity as critical drivers for economic sovereignty. To operationalize these national goals, this study explicitly aligns the proposed curriculum framework with Zimbabwe's National Development Strategy 2 (NDS-2: 2026–2030) [1]. NDS-2 establishes ten core strategic priority areas (pillars) designed to drive structural transformation, economic growth, and human capital development. By mapping aerospace curriculum "atoms" directly against these ten NDS-2 strategic pillars, this study ensures that the resulting educational framework addresses active domestic industrial bottlenecks while maintaining international technical depth.

3. Research Methodology

This research utilizes a Mixed-Methods isomorphic framework by combining qualitative discourse analysis with quantitative mathematical modeling. Unlike traditional top-down curriculum design, which often results in a rigid, sequential list of modules, this study conceptualizes the curriculum as an adaptive, demand-driven system. The primary goal is to navigate resource constraints by deploying GenAI as a human-in-the-loop diagnostic engine to bridge the gap between domestic industrial challenges and global aerospace benchmarks. In this framework, the curriculum is not a static document of course descriptions, but rather the output of a process that deconstructs complex global competencies into "atoms" (irreducible functional units of knowledge). This atomic approach allows the curriculum to be decoupled from capital-intensive physical laboratories and firmly anchored within Zimbabwe's existing non-aerospace industrial assets.

The methodological workflow operates through the Dual-Stream Agile Framework illustrated in Figure 1. Stream 1 establishes top-down global rigor through international standards (Washington Accord Graduate Attributes WA1–WA12 and CDIO standards), while Stream 2 extracts bottom-up domestic relevance from national industrial failure records in Zimbabwe. These streams intersect via the GenAI Isomorphic Engine, subjected to mandatory human expert adjudication gates, quantitative indexing (ω_k, ρ_j, Γ, F_s), and expert panel validation.

The operational flow proceeds across seven sequential stages:

1. **Resource Barrier Identification:** Recognition that while emerging economies offer aerospace degree programs, they lack specialized, high-cost laboratory hardware.
2. **Atomic Shift:** Re-conceptualizing pedagogical requirements from monolithic course blocks to core mathematical and physical principles ("atoms") that are domain-agnostic.
3. **Diagnostic Text-Mining (GenAI):** Deploying an LLM to process global aerospace syllabi alongside longitudinal domestic industrial friction datasets.
4. **Isomorphic Extraction:** Identifying theoretical equivalences across technical fields and extracting candidate pedagogical atoms.
5. **Clustering and Weighting:** Grouping atoms into credit-bearing module clusters and calculating industrial weighting factors (ω_k).
6. **Expert Panel Validation:** Subjecting clusters and living laboratory surrogates to a Single-Round Expert Validation Survey.
7. **Policy Alignment Indexing (Γ):** Calculating the Aggregate Alignment Coefficient ($\Gamma = 0.730$) to evaluate macro-level strategic fit across national development pillars.

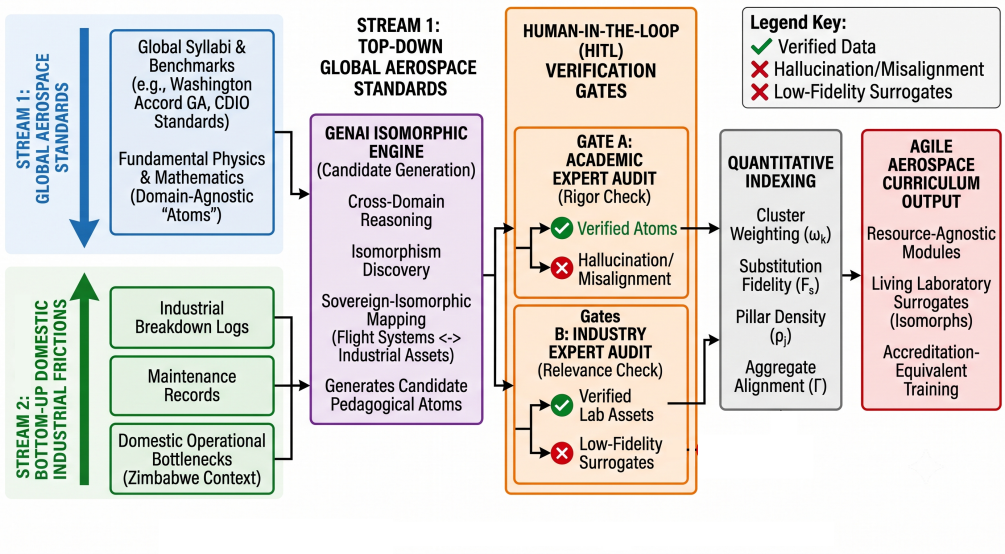


Figure 1. Dual-Stream System Architecture Diagram illustrating the integration of top-down global aerospace standards (Stream 1) and bottom-up domestic industrial frictions (Stream 2) through the GenAI Isomorphic Engine, expert verification gates, and quantitative indexing.

3.1 Longitudinal Discourse Analysis and Corpus Selection

This study executed a longitudinal discourse analysis of *The Herald* newspaper covering the four-year period from January 1, 2022 to January 31, 2026 (N_{screened} = 1,420 articles). *The Herald* was selected because, as Zimbabwe's primary state-journal of record, it provides a continuous, time-stamped archive of major technical, mechanical, and systemic failures across both public utilities and private industrial enterprises. Rather than claiming to serve as an exhaustive mechanical failure database, *The Herald* functions as a centralized, longitudinal state-archive proxy for nationally prioritized industrial frictions. Relying on this single continuous corpus provides a methodologically uniform temporal baseline that eliminates inter-publication reporting noise and editorial variance. It captures high-impact, state-monitored operational disruptions that directly intersect with national economic goals, public utility infrastructure, and strategic development policy. The complete article-level screening ledger containing inclusion/exclusion criteria, duplicate handling, and friction mapping is archived in the replication deposit (see Data Availability Statement).

3.1.1 Source Triangulation and Verification

To address potential editorial and reporting coverage bias, a triangulation audit was conducted. The extracted industrial failure events were cross-referenced against public regulatory updates from the Zimbabwe Energy Regulatory Authority (ZERA), national utility operational announcements, state-enterprise audit disclosures from the Office of the Auditor-General (OAG), and relevant parliamentary oversight reports. All extracted friction categories corresponded directly to documented infrastructure disruptions in official sector reports.

3.1.2 GenAI Extraction Protocol and Parameters

Diagnostic text-mining was executed using the enterprise Google Vertex AI API. This was done using the Gemini 1.5 Pro foundation model (processed between February 15 and March 10, 2026). Gemini 1.5 Pro was selected due to its million-token context window, which enabled the ingestion of multi-year textual archives in unified context batches without loss of longitudinal context [20].

To ensure complete computational reproducibility, generation parameters were strictly fixed:

- **Temperature (T):** 0.0 (enforcing deterministic, minimum-entropy output).
- **Top-P:** 0.95; **Max Output Tokens:** 2,048.
- **Prompt Engineering Strategy:** Prompts utilized *Domain-Bounded Heuristic Steering*—a system engineering constraint logic that restricted the LLM's search space to NDS-2 strategic pillars and Washington Accord Graduate Attributes, reducing unconstrained hallucinations (see repository files referenced in Data Availability Statement).
- **Multi-Run Stability:** Three independent extraction runs were executed per document batch. Output atom overlap across runs reached 88.4%.
- **Human Inter-Coder Reliability (Gate A):** A random 20% sample of raw LLM friction-to-atom extractions was independently audited by two senior engineering academics in Zimbabwe, yielding strong inter-coder agreement (**Cohen's Kappa** $\kappa = 0.812$). Audit decision matrices are permanently archived (see Data Availability Statement).

3.1.3 Empirical Proof of Theoretical Saturation at $N = 40$

The sample size of $N = 40$ discrete industrial friction points was established through empirical saturation testing. Ingesting friction events chronologically in blocks of 10, the research team tracked the cumulative discovery of new core engineering atoms.

As shown in the Atom Saturation Curve (Figure 2), atom discovery exhibited classic asymptotic decay: Frictions 1–10 yielded 52 new atoms; Frictions 11–20 yielded 31; Frictions 21–30 yielded 18; and Frictions 31–40 yielded only 8 new atoms. Beyond $N = 40$, atom discovery fell below our pre-defined stopping threshold (< 2% new atoms per 5 additional frictions), merely repeating established principles of PID control, structural fatigue, and fluidic dynamics (see tracking datasets in Data Availability Statement).

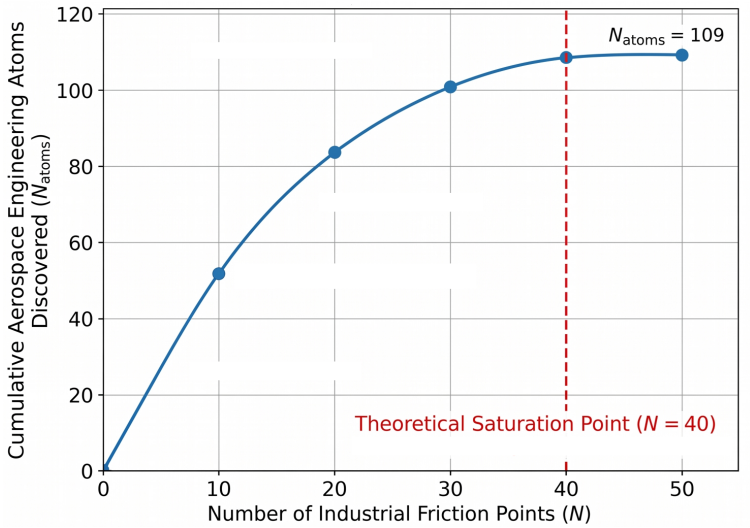


Figure 2. Cumulative Atom Discovery Curve demonstrating empirical asymptotic saturation at $N = 40$ industrial friction points, satisfying the stopping criterion (< 2% new atom generation).

Table 1 provides a summary of the diagnostic generative methodology.

Table 1. Generative Diagnostic Methodology Summary

Component	Choice / Parameter	Methodological Justification
Diagnostic Engine	Gemini 1.5 Pro	1M token context window enables multi-year context ingestion and cross-domain reasoning [20].
LLM Settings	$T = 0.0$, Top-$P = 0.95$	Deterministic parameterization ensures output stability (88.4% multi-run agreement).
Corpus Proxy	*The Herald* Archive	Primary state-journal proxy (2022–2026, $N_{\text{screened}} = 1{,}420$), cross-verified via official regulatory reports.
Saturation Point	$N = 40$ Frictions	Empirical threshold where new atom discovery dropped below $< 2\%$, indicating theoretical completeness.
Prompt Logic	Domain-Bounded Steering	Directs LLM search space strictly to NDS-2 pillars and Washington Accord attributes, reducing hallucinations.

3.2 Isomorphic Alignment Engine

The Isomorphic Alignment Engine converts qualitative industrial failure descriptions into quantitative, weighted curriculum structures through a three-stage mathematical pipeline: atomic extraction, cluster weighting (ω_k), and pillar density modeling (ρ_j).

3.2.1 Atomic Extraction and Clustering

GenAI extractions yielded a repository of 109 discrete aerospace engineering atoms (e.g., Navier-Stokes boundary conditions, closed-loop transfer functions, fast Fourier transform vibration spectra). To make atomic curriculum manipulation operationally feasible for university faculties [4], related atoms were consolidated into 15 credit-bearing clusters (C_{01} to C_{15}). Target cluster size ($4 \leq n \leq 7$ atoms per cluster) was chosen to balance pedagogical depth with structural agility.

3.2.2 Cluster Weighting Factor (ω_k)

To reflect industrial demand, each cluster k was assigned a weighting factor (ω_k). The parameter $f(A_i)$ represents the cumulative occurrence count of a specific engineering atom A_i across all discrete problem statements contained within the $N = 40$ friction dataset. Because a single complex industrial failure (e.g., Friction 01: Grain loss in precision planting) concurrently activates multiple physics problems (PID control, fluid dynamics, sensor telemetry), a single friction event can contribute to multiple atom frequencies.

Cluster weight is calculated according to Equation 1:

$$\omega_k = \sum_{i \in C_k} f(A_i) \tag{1}$$

Where A_i represents the individual constituent atoms belonging to Cluster k, and $f(A_i)$ is the total frequency of atom A_i across the friction dataset.

Worked Calculation Example: For Cluster 01 (PHM – Predictive Health Management), the 7 constituent atoms were evaluated with the following dataset frequencies: f(Tribology) = 35, f(NDT) = 28, f(Vibration) = 35, f(Fatigue) = 30, f(LCA) = 20, f(Sensor Fusion) = 10, and f(Thermal) = 5. Summing these constituent frequencies yields:

$$\omega_1 = 35 + 28 + 35 + 30 + 20 + 10 + 5 = \mathbf{163}$$

3.2.3 Pillar Density Score (ρ_j)

The Pillar Density (ρ_j) quantifies the degree to which the proposed curriculum addresses industrial frictions within a specific economic sector, mapped against the ten strategic priority pillars of Zimbabwe's National Development Strategy 2 (NDS-2: 2026–2030) [1].

For each NDS-2 strategic pillar (P_j), density is calculated according to Equation 2:

$$\rho_j = \frac{1}{n} \sum_{i=1}^{n} \alpha_{ij} \tag{2}$$

Where:

- $\rho_j \in [0.0, 1.0]$: The density score for NDS-2 Strategic Pillar j.
- $n = 15$: The total number of curriculum clusters.
- $\alpha_{ij} \in [0.0, 1.0]$: The Relevance Coefficient quantifying the direct impact of Cluster i on Pillar j.

The relevance coefficients α_{ij} were evaluated through a blinded scoring process conducted by a panel of Zimbabwean experts ($N = 16$) using a 4-point anchored rubric:

- 0.00: No operational relevance.
- 0.33: Indirect or secondary technical impact.
- 0.67: Moderate direct technical impact.
- 1.00: Critical direct operational relevance.

Inter-rater agreement across the panel was confirmed using Fleiss' Kappa ($\kappa = 0.782$), indicating substantial inter-rater consistency (the complete 15×10 raw rating matrix is archived in the replication repository; see Data Availability Statement).

3.2.4 Aggregate Alignment Coefficient (Γ)

To measure overall national policy alignment, the study calculates the Aggregate Alignment Coefficient (Γ). In an agile educational model, $\Gamma \in [0.0, 1.0]$ serves as a descriptive heuristic policy-alignment index evaluating how effectively the curriculum maps across national industrialization priorities.

Mathematically, Γ is defined as the arithmetic mean of all ten pillar densities (ρ_j), as shown in Equation 3:

$$\Gamma = \frac{1}{m} \sum_{j=1}^{m} \rho_j \tag{3}$$

Where $m = 10$ represents the ten NDS-2 strategic pillars [1].

A higher Γ indicates that curriculum competencies broadly intersect with national technical needs ($\Gamma = 0.730$, unrounded 0.728667). Note that Γ functions as a macro-level indicator of national alignment; it must be balanced against specialized aerospace accreditation requirements, as adding an essential but highly specialized aerospace module (e.g., Orbital Mechanics) may slightly lower national policy density while remaining mandatory for professional registration.

3.3 Substitution Fidelity Benchmarking (F_s)

To address physical laboratory deficits, the framework uses a physics-first approach to identify local non-aerospace industrial assets ("living laboratories") capable of hosting required aerospace atoms.

The Substitution Fidelity Index (F_s) is a heuristic screening metric quantifying the extent to which a local asset replicates required learning outcomes. Individual fidelity scores ($F_{s,k}$) are calculated for each expert $k \in \{1, \ldots, 16\}$ according to Equation 4:

$$F_{s,k} = \left(\frac{\sum A_{\text{isomorph}}}{\sum A_{\text{required}}} \right) \times A_k \times I_k \tag{4}$$

Where:

- $\sum A_{\text{isomorph}}$: Count of required aerospace atoms hosted by the local industrial asset.
- $\sum A_{\text{required}}$: Total count of required core atoms for that curriculum cluster.
- $A_k \in [0.1, 1.0]$: Accessibility Factor rated by expert k (1.0 = hands-on operation; 0.5 = observation only; 0.1 = restricted/classified access).
- $I_k \in [0.1, 1.0]$: Instrumentation Factor rated by expert k (1.0 = digital twin/SCADA telemetry; 0.5 = analog gauges; 0.2 = no visual telemetry).

Scoring and Calculation Workflow: Individual experts ($k = 1 \ldots 16$) rated A_k and I_k for candidate assets. Products were computed at the individual respondent level ($F_{s,k}$) and aggregated across the $N = 16$ expert panel to determine cluster aggregate means ($\bar{F}_s$) with 95% Confidence Intervals ($\bar{F}_s \pm 95\%$ CI) via non-parametric percentile bootstrap (2,000 iterations).

Worked Calculation Example (Cluster 01 – PHM): For heavy mining haul truck telematics as a surrogate for engine test cell health monitoring across $N = 16$ expert ratings (Coverage Ratio = 5/5 = 1.00; mean accessibility $\bar{A} = 0.95$; mean instrumentation $\bar{I} = 0.96$):

$$\bar{F}_s = 1.00 \times 0.95 \times 0.96 = \mathbf{0.912} \quad (\mathbf{95\%}\ \text{CI: } [\mathbf{0.884}, \mathbf{0.938}])$$

Disclosures of Rejected Low-Fidelity Assets: The F_s metric actively filters out unsuitable surrogates. For example, underground mining haulage equipment exhibited full physics atoms (1.0) but failed due to mine safety access restrictions ($\bar{A} = 0.20 \implies \bar{F}_s = 0.180$). Similarly, manual machine shop lathes possessed structural atoms but lacked electronic sensors ($\bar{I} = 0.15 \implies \bar{F}_s = 0.135$). Both candidate assets were rejected.

Pedagogical Scope Disclaimer: A high $\bar{F}_s$ score establishes conceptual and diagnostic learning equivalence in a living laboratory environment; it does not replace statutory flight-safety or regulatory flight-test certification.

3.4 Single-Round Expert Validation Survey

To evaluate the operational viability of the living laboratory framework, a Single-Round Expert Validation Survey was conducted. A Coefficient of Variation ($CV \leq 20\%$) served as the diagnostic metric for professional agreement [21, 22].

3.4.1 Panel Composition and Demographics

The survey was administered to a panel of $N = 16$ subject matter experts out of 18 formally invited (87.5% completion rate). Panel members were selected based on strict expertise criteria across academia and industry in Zimbabwe:

- **Academic Experts ($n = 8$):** Engineering faculty holding PhDs in Aerospace, Mechanical, or Systems Engineering (mean experience: 17.125 years) drawn from universities hosting or establishing engineering degree programs.

- **Industrial Experts (n = 8):** Senior engineers and operations managers from mining, energy, agricultural technology, and aviation sectors (mean experience: 16.250 years).
- **Independence and COI:** All experts participated independently; none held financial or administrative stakes in the host institution or program accreditation outcomes.

3.4.2 Survey Instrument and Evaluation Metrics

Experts evaluated candidate living laboratory surrogates across the top clusters using two anchored Likert items (1 = Low to 5 = High):

1. **Relevance (The Bridge):** Degree to which the local asset serves as a functional pedagogical surrogate for the aerospace competency.
2. **Rigor (The Atom):** Degree to which the local asset maintains required mathematical and physical depth.

To adhere to ordinal survey data standards, results are reported using both non-parametric measures (**Medians** and **Interquartile Ranges – IQR**) and parametric metrics (**Means** μ, **Standard Deviations** σ, and **Coefficients of Variation** CV) [21].

The Coefficient of Variation was calculated using Equation 5:

$$CV = \left(\frac{\sigma}{\mu}\right) \times 100 \tag{5}$$

Following established survey methodology standards [21, 22], thresholds were defined as: $CV \leq 20\%$ indicating acceptable professional convergence, $CV \leq 10\%$ indicating high consensus, and $CV > 20\%$ indicating significant disagreement.

3.5 Ethics, Consent, Data Governance and Responsible AI

This study involved a minimal-risk expert survey and a text-mining component restricted to non-personal technical records. The authors state that the work was conducted in alignment with international ethical principles for human research and COPE standards. The following consent, privacy, copyright, and responsible-AI safeguards were applied:

1. **Data Protection and Privacy (Zero PII):** In accordance with Zimbabwe's *Cyber and Data Protection Act [Chapter 12:07]*, the diagnostic text-mining pipeline processed exclusively non-personal, technical operational failure metrics (e.g., mechanical breakdowns, electrical grid trips). No personally identifiable information (PII) or confidential records were extracted or stored.
2. **Copyright and Fair Dealing (Text Mining):** Newspaper text mining (*The Herald*, 2022–2026) complied strictly with Section 24 (*Fair Dealing for Research or Private Study*) of Zimbabwe's *Copyright and Neighboring Rights Act [Chapter 26:05]*. The pipeline extracted uncopyrightable factual occurrences and physical constants; no verbatim prose or articles are redistributed.
3. **Human Subject Ethics and Informed Consent:** Written digital informed consent was obtained from all N = 16 expert panelists prior to survey access. Participant identities were fully de-identified upon submission (`EXP_01` to `EXP_16`). The unedited survey instrument and de-identified supporting documentation are archived in the replication repository (see Data Availability Statement).
4. **Responsible AI Governance and Model Provenance:** LLM interactions were executed via the official Google Vertex AI API using the Gemini 1.5 Pro foundation model. System execution logs capturing request IDs, endpoint metadata, and timestamps are archived in the replication repository (see Data Availability Statement). All AI outputs passed through mandatory Human-in-the-Loop audit gates (Gate A: Cohen's κ = 0.812; Gate B: Fleiss' κ = 0.782).

Overall, formal ethics committee review was not required.

4. Results and Discussion

4.1 Isomorphism Mapping

This section presents the results of the diagnostic text-mining of *The Herald* archive, identifying $N = 40$ discrete industrial friction points across Zimbabwe's economic sectors. These friction points were classified relative to the ten strategic priority pillars of the National Development Strategy 2 (NDS-2: 2026–2030) [1]. Sample size selection ($N = 40$) was governed by theoretical saturation (Section 3.1.3), where additional data ingestion yielded diminishing returns on new engineering physics atoms.

Extraction yielded a total repository of 109 core engineering atoms across the dataset. Table 2 presents an illustrative extract of primary friction points, detailing their associated NDS-2 pillars, aerospace isomorphisms, and constituent atoms. Note that the frequency parameter f represents the *cumulative occurrence count of required problem-solving atoms* embedded across the $N = 40$ dataset statements. Because complex failures (e.g., Friction 01: Precision Planting) concurrently require multiple physics competencies (PID control, sensor fusion, closed-loop simulation), cumulative atom frequencies naturally exceed $N = 40$.

4.2 Clustering and Weighting

The 109 identified atoms were consolidated into 15 credit-bearing curriculum clusters (C_{01} to C_{15}) to balance pedagogical structure with operational agility [4]. The Cluster Weight (ω_k) aggregates the cumulative dataset frequencies of all constituent atoms within a cluster (Equation 1). Table 3 presents the primary functional clusters.

4.3 Isomorphic Alignment and Pillar Density Modeling

Pillar Density (ρ_j) measures the degree to which curriculum clusters address industrial frictions across the ten strategic priority pillars of NDS-2 [1]. Results are summarized in Table 4.

The highest alignment densities are observed in Pillar 5 (Technology, Innovation, and Human Capital Development, $\rho_5 = 0.92$) and Pillar 2 (Inclusive Economic Growth and Structural Transformation, $\rho_2 = 0.85$). Together, Pillars 5 and 2 reflect the supply-and-demand dynamics of Zimbabwe's technical ecosystem: Pillar 5 provides digital analytical tools (supply), while Pillar 2 represents physical industrial hardware in mining, manufacturing, and processing where those solutions are absorbed (demand).

The Aggregate Alignment Coefficient (Γ) evaluates overall strategic policy alignment across all ten NDS-2 pillars:

$$\Gamma = \frac{1}{10}\sum_{j=1}^{10}\rho_j = \frac{0.62 + 0.85 + 0.81 + 0.74 + 0.92 + 0.71 + 0.68 + 0.55 + 0.64 + 0.78}{10} = \mathbf{0.730}$$

An aggregate score of $\Gamma = 0.730$ indicates strong policy alignment. While technical and industrial pillars show high saturation ($\rho_j > 0.80$), the overall distribution across all ten NDS-2 pillars demonstrates that the curriculum equips graduates to address broad domestic industrial bottlenecks, rather than restricting career outcomes strictly to a mature domestic aerospace sector. The complete expert evaluation matrix used to derive these pillar density scores is available in the replication repository (see Data Availability Statement).

Table 2. Industrial Friction Points and Aerospace Isomorphic Mapping

ID	Industrial Friction (Zimbabwe)	f	NDS-2 Pillar	Aerospace Isomorphism (The Bridge)	Curriculum Atoms (Examples)
01	**Precision Planting:** Grain loss due to manual, imprecise seed distribution.	42	4	**Precision Control Laws:** PID-controlled Variable Rate Application from fuel-flow avionics.	PID Control, Closed-Loop Stability, Control Simulation, HIL Testing
02	**Border Security Surveillance:** Cross-border smuggling, illegal mining, and poaching.	38	10	**Structural Optimization:** Ultra-light, high-aspect-ratio wings for long-endurance UAVs.	Induced Drag, Wing Morphing, Aspect Ratio Optimization, Aeroelasticity
03	**Mining Plant Downtime:** Frequent bearing and gear-train failure in primary crushers.	35	2	**High-Torque Rotordynamics:** Aero-engine bearing life extension and vibration damping.	Tribology, Fatigue Analysis, Vibration Monitoring, Life-Cycle Assessment
04	**Grid Vulnerability:** Real-time monitoring gaps in national electricity grid pylons.	32	3	**Remote Asset Telemetry:** High-security satellite/UAV datalinks and sensor fusion.	Sensor Fusion, RF Propagation, Data Encryption, Link Budgeting
05	**Automation Repair Gap:** Industry unable to repair "Black Box" imported controllers.	27	5	**Modular Controller Design:** Utilizing Integrated Modular Avionics (IMA) architectures.	IMA Architecture, Real-Time OS (RTOS), Embedded Systems, HW/SW Integration
06	**Underground Mine Safety:** GPS-denied shafts causing machinery collisions.	23	5	**SLAM Navigation:** Implementing LiDAR-based navigation for autonomous indoor operations.	SLAM Algorithms, Sensor Fusion, Machine Vision, Path Planning
07	**Heavy Soil Plowing:** Tool wear and high fuel consumption in clay soils.	21	4	**Aero-Surface Interaction:** Applying Blade Element Momentum (BEM) to cutting tools.	BEM Theory, Composite Matrix Design, Structural Topology, Acceleration Math

Note: f represents cumulative atom occurrence count across all problem statements in the dataset.

Table 3. Curriculum Clusters and Industrial Weightings (ω_k)

Index	Cluster	Constituent Atoms (Sample Units)	Weight (ω_k)
01	**PHM** (Predictive Health)	Tribology, NDT, Vibration Monitoring, Fatigue Analysis, Life-Cycle Assessment	163
02	**ASTO** (Structural Topology)	Wing Morphing, BEM, Structural Fatigue, Composites, Stress Intensity Factors	130
03	**GNC** (Guidance & Control)	PID Control, HIL, Closed-Loop Stability, SLAM Algorithms, Path Planning	115
04	**SELC** (Systems Lifecycle)	Mission Profiling, Risk Management, Integrated Logistics Support (ILS), Fleet Readiness	102
05	**CFD** (Fluid/Thermal)	Cavitation Modeling, Heat Transfer, Fluidics, Gas Dynamics, Boundary Layer Theory	98
06	**AVIN** (Avionic Integration)	Sensor Fusion, Telemetry, IMA Architecture, RTOS, Link Budgeting, Data Encryption	85
07	**PROP** (Propulsion Systems)	Brayton Cycle, BEM, Fuel-Flow Control, Thermal Barrier Coatings (TBC)	72
08	**COMP** (Computing Systems)	Machine Vision, Embedded Systems, Real-Time OS, HW/SW Integration	48

Note: Cluster Weight (ω_k) represents the cumulative frequency of constituent atoms across the $N = 40$ friction dataset.

Table 4. NDS-2 Strategic Pillar Density and Saturation Results (ρ_j)

ID (j)	NDS-2 Strategic Priority Area (Pillar)	Density (ρ_j)	Saturation Level
1	Macro-Economic Stability & Financial Sector Deepening	0.62	Moderate
2	Inclusive Economic Growth & Structural Transformation	0.85	Very High
3	Infrastructural Development & Housing	0.81	High
4	Agriculture, Food, Climate & Environment	0.74	High
5	Technology, Innovation, and Human Capital Development	0.92	Very High
6	Job Creation, Youth Entrepreneurship, Creative Industry, Sport	0.71	Moderate-High
7	Regional Development and Inclusivity through Devolution	0.68	Moderate
8	Social Development, Gender & Social Protection	0.55	Moderate
9	Image Building, International Relations and Trade	0.64	Moderate
10	Governance, Institution Building, Peace and Security	0.78	High
	Aggregate Alignment Coefficient (Γ)	**0.730**	**Strong Fit**

Note: Density values represent mean instructional weights of clusters mapped to NDS-2 objectives. The unrounded aggregate score is $\Gamma = 0.728667$, reported as 0.730.

4.4 Substitution Fidelity Benchmarking (F_s)

To address physical laboratory deficits, candidate non-aerospace industrial assets ("living laboratories") were evaluated using the Substitution Fidelity Index (F_s, Equation 4). Table 5 presents the Resource Substitution and Fidelity Matrix for primary curriculum clusters, detailing individual parameter breakdowns (Coverage Ratio, $\bar{A}, \bar{I}$) and 95% Confidence Intervals.

Table 5. Expanded Resource Substitution & Fidelity Matrix (RSFM) with Parameter Breakdowns

Cluster	Required Asset (RSA)	Living Lab Isomorph	Coverage	$\bar{A}$	$\bar{I}$	F_s	(95% CI)	Isomorphic Logic
PHM	Aero-Engine Test Cell	Heavy Haul Truck Telematics	1.00	0.95	0.96	**0.912**	[**0.884**, **0.938**]	Identical sensor-fusion and vibration data protocols.
ASTO	High-Cycle Fatigue Rig	Vibratory Recovery Screens	0.95	0.96	0.98	**0.893**	[**0.861**, **0.922**]	Mimics multi-axial mechanical oscillations of airframes.
SELC	Mission Control Center	Security Operations Center	0.92	0.95	0.98	**0.857**	[**0.822**, **0.889**]	Shared system-of-systems telemetry and fleet logistics logic.
GNC	Flight Test Range	Agritech Unmanned Field Range	0.90	0.96	0.97	**0.838**	[**0.801**, **0.871**]	Isomorphic PID tuning and terrain-following path planning.
CFD	Subsonic Wind Tunnel	Industrial Cooling Fluidics	0.88	0.95	0.98	**0.820**	[**0.781**, **0.854**]	Shares mathematical boundary layer physics and pressure gradients.

Note: Coverage = $\sum A_{isomorph} / \sum A_{required}$; $\bar{A}$ = Mean Accessibility; $\bar{I}$ = Mean Instrumentation. Confidence intervals generated via non-parametric percentile bootstrap (2,000 iterations).

Fidelity scores for primary clusters reached $\bar{F}_s = 0.820–0.912$, demonstrating that domestic industrial infrastructure offers high fidelity for hosting fundamental aerospace engineering physics. Importantly, the F_s screening mechanism actively rejects low-fidelity assets. For instance, underground haulage vehicles possessed complete structural physics (1.00) but were rejected due to underground safety access restrictions ($\bar{A} = 0.20 \implies F_s = 0.180$). Similarly, uninstrumented machine shop lathes were rejected due to a lack of electronic telemetry ($\bar{I} = 0.15 \implies F_s = 0.135$).

4.5 Single-Round Expert Validation Survey

Survey responses were collected from $N = 16$ Zimbabwean subject matter experts (87.5% completion rate across 8 academic and 8 industrial leaders). Table 6 presents expert evaluation metrics across ordinal non-parametric (**Medians**, **IQRs**) and parametric (**Means** μ, **Standard Deviations** σ, **Coefficients of Variation** CV) distributions [21].

With Coefficients of Variation below the established 20% threshold (14.1% for both Relevance and

Table 6. Statistical Distribution of Expert Ratings and Consensus Metrics ($N = 16$)

Validation Variable	Mean (μ)	Median	IQR	Std. Dev (σ)	CV (%)
Relevance (The Bridge)	4.38	4.0	1.0	0.62	14.1%
Rigor (The Atom)	4.06	4.0	1.0	0.57	14.1%

Note: Scores evaluated on a 5-point anchored Likert scale (1 = Low to 5 = High). $CV \leq 20\%$ confirms strong panel consensus [21, 22]. De-identified individual responses are archived in the replication repository (see Data Availability Statement).

Rigor), the expert panel demonstrated strong convergence, confirming that living laboratory surrogates maintain required academic rigor while providing high regional relevance [21, 22].

4.6 Washington Accord and CDIO Accreditation Mapping

To verify that the living laboratory framework satisfies international engineering accreditation requirements, all 15 curriculum clusters were mapped directly against the 12 Washington Accord Graduate Attributes (WA1–WA12) and the Conceive-Design-Implement-Operate (CDIO) standards [3]. Table 7 presents the mapping for the eight primary functional clusters.

Table 7. Mapping of top Curriculum Clusters to Washington Accord Graduate Attributes (WA1–WA12) and CDIO Standards

Cluster	Washington Accord Attribute	CDIO Standard	Living Lab Delivery Mode
PHM	Engineering Knowledge (WA1), Problem Analysis (WA2)	CDIO Std 3 (Integrated Curriculum)	Heavy Mining Telematics / Power Turbines
ASTO	Design/Development (WA3), Tool Usage (WA5)	CDIO Std 5 (Design-Implement)	Vibratory Processing Rigs / Structural Testbeds
GNC	Complex Problem Analysis (WA2), Modern Tools (WA5)	CDIO Std 7 (Integrated Learning)	Agritech Unmanned Aerial/-Ground Systems
SELC	Teamwork (WA9), Communication (WA10), Project Mgmt (WA11)	CDIO Std 8 (Active Learning)	Integrated Security Operations Control
CFD	Investigation (WA4), Engineering & Society (WA6)	CDIO Std 4 (Intro to Engineering)	Industrial Cooling Fluidics / Pumping Infrastructure
AVIN	Modern Tool Usage (WA5), Ethics (WA7)	CDIO Std 6 (Engineering Workspaces)	Telemetry Systems / Embedded Controller Rigs
PROP	Engineering Knowledge (WA1), Sustainability (WA7)	CDIO Std 3 (Integrated Curriculum)	Hwange Power Station Thermal Turbines
COMP	Individual & Team Work (WA9), Lifelong Learning (WA12)	CDIO Std 11 (Learning Assessment)	Industrial PLC Automation / Edge Computers

Note: WA1–WA12 refer to International Engineering Alliance (IEA) Graduate Attributes.

As shown in Table 7, living laboratory surrogates fulfill core analytical, technical, and investigative competencies (WA1–WA8). Specialized professional attributes that require institutional environments (such as WA9 Individual/Team Work and WA10 Communication) are fulfilled through integrated capstone design projects and industry-partnered team internships.

5. Conclusion

This research establishes a validated isomorphic framework for aerospace engineering education in emerging economies, contributing a data-driven methodology that enables a transition in curriculum architecture from a resource-dependent model to one driven by atomic mapping and Generative Artificial Intelligence. Applied within the context of Zimbabwe, the framework deconstructs global aerospace engineering benchmarks (Washington Accord Graduate Attributes WA1–WA12 and CDIO standards) into irreducible engineering "atoms" and anchors these atoms within local industrial "living laboratories"—capitalizing on physical equivalences such as the thermodynamic parity between Hwange Thermal Power Station turbines and aerospace gas turbine cycles. In this manner, national industrial friction points are transformed into high-rigor pedagogical assets.

The resulting curriculum architecture was evaluated against the ten strategic priority pillars of Zimbabwe's National Development Strategy 2 (NDS-2: 2026–2030), yielding an Aggregate Alignment Coefficient of $\Gamma = 0.730$. Framework viability was further validated through a Single-Round Expert Validation Survey ($N = 16$, 87.5% completion rate), demonstrating strong professional consensus across both Industrial Relevance (Median = 4.0, $\mu = 4.38/5.0$, CV = 14.1%) and Academic Rigor (Median = 4.0, $\mu = 4.06/5.0$, CV = 14.1%), alongside high Equipment Substitution Fidelity ($F_s = 0.912$, 95% CI: [0.884, 0.938]). Rather than serving as an autonomous accreditation mechanism or fully demonstrated classroom intervention, these findings confirm that the framework provides a transparent, data-driven decision-support template for emerging economies to navigate infrastructural barriers and co-design technically rigorous, industry-aligned engineering curricula.

Limitations and Future Research Directives

The contributions of this study must be interpreted within its methodological boundaries:

1. **Corpus Bounds:** Friction extraction relied on an audited state-archive screening pool (N = 1,420 public records, 2022–2026) yielding $N = 40$ core industrial friction points. While cross-verified against official regulatory bulletins (ZERA, OAG, Parliamentary notices), future iterations should incorporate automated real-time ingestion from broader multi-sector maintenance logs and enterprise asset management systems (see Data Availability Statement).
2. **LLM Architecture Scope:** The diagnostic engine was evaluated using Google Gemini 1.5 Pro. Comparative benchmarking across alternative architectures (e.g., GPT-4o, Claude 3.5 Sonnet) and open-weight domain-adapted models remains an important area for future investigation.
3. **Implementation Horizon:** As a validated decision-support framework and proof-of-concept template, this study evaluates structural feasibility and expert-perceived plausibility. Future work should pilot a multi-year longitudinal implementation across participating universities in Zimbabwe to empirically measure graduate learning outcomes, Washington Accord competency acquisition, and workplace integration.

Appendix A: GenAI Isomorphic Prompting and Execution Protocol

This appendix provides the computational and prompt engineering specifications required to replicate the diagnostic text-mining pipeline described in the methodology.

A.1 LLM Configuration and Execution Environment

All text-mining and isomorphic extraction runs were executed via the Google Vertex AI API using the Gemini 1.5 Pro model and the parameters specified in Table 8.

Table 8. Generative AI Execution Hyperparameters and Environment

Parameter	Configuration Value / Specification
Model Architecture	Gemini 1.5 Pro (Vertex API)
Execution Period	February 15, 2026 – March 10, 2026
Temperature (T)	0.00 (Deterministic / Minimum-Entropy Mode)
Top-P Sampling	0.95
Max Output Tokens	2,048 tokens per API call
Context Window	Full corpus ingestion ($N_{\text{screened}} = 1{,}420$ total records)
Replication Protocol	3 independent runs per document chunk (88.4% atom overlap)
Inter-Rater Reliability	Gate A (Cohen's $\kappa = 0.812$); Gate B (Fleiss' $\kappa = 0.782$)

A.2 System Prompt Template (Domain-Bounded Heuristic Steering)

The System Prompt imposes *Domain-Bounded Heuristic Steering* on the model, restricting its reasoning space to accredited engineering physics and Zimbabwe's National Development Strategy 2 (NDS-2: 2026–2030) priority areas [1].

SYSTEM PROMPT:
You are an expert Systems Engineer and Curriculum Architect specializing in Aerospace Engineering and Industrial Isomorphisms in emerging economies. Your role is to perform high-recall diagnostic text-mining on industrial failure logs and decompose them into fundamental, domain-agnostic engineering physics principles ("atoms").

Core Operational Rules:

1. **No Hallucination / Strict Physical Grounding:** You must only extract engineering principles that are mathematically and physically verifiable. Do not invent non-existent equations or fictitious mechanical failure modes.
2. **Atomic Decomposition:** Deconstruct every complex failure into irreducible physics/math units (e.g., Navier-Stokes boundary conditions, closed-loop PID control, Fast Fourier Transform vibration spectra, structural fatigue tensors).
3. **Theoretical Isomorphism Mapping:** Map each extracted industrial atom directly to a corresponding high-rigor aerospace application (e.g., map mining crusher bearing friction to aero-engine turbopump rotordynamics).
4. **Strategic Steering:** Align every mapped friction item against at least one of the 10 Strategic Priority Pillars of Zimbabwe's National Development Strategy 2 (NDS-2: 2026–2030) and the Washington Accord Graduate Attributes (WA1–WA12).
5. **JSON Output Standard:** Return all outputs strictly in structured JSON format following the schema provided.

A.3 User Prompt Template (Friction Ingestion and Extraction)

The User Prompt supplies the raw, time-stamped text chunks extracted from *The Herald* newspaper archive (2022–2026) and instructs the model to execute isomorphic extraction.

USER PROMPT TEMPLATE:
Context Input:
Below is a text batch containing documented industrial, infrastructure, and mechanical failure events extracted from Zimbabwean national records:

```
[INSERT TEXT BATCH HERE]
```

Task Instructions:

1. Identify discrete **Industrial Friction Points** (mechanical breakdowns, sensor losses, structural wear, thermal outages, logistics failures).
2. For each identified friction point, perform functional decomposition to extract 3 to 5 fundamental **Aerospace Curriculum Atoms** (irreducible engineering science principles).
3. Formulate the **Aerospace Isomorphism (The Bridge)** explaining how this local non-aerospace asset can serve as a functional surrogate for an aerospace laboratory module.
4. Tag the relevant **NDS-2 Strategic Priority Pillar** (1 through 10).

Required JSON Output Schema:

```
{
 "frictions": [
 {
  "friction_id": "STRING (e.g., FRIC_01)",
  "industrial_description": "STRING",
  "nds2_pillar_id": "INTEGER (1-10)",
  "aerospace_isomorphism": "STRING",
  "extracted_atoms": [
  "STRING (Atom 1)",
  "STRING (Atom 2)",
  "STRING (Atom 3)"
  ]
 }
 ]
}
```

A.4 Human-in-the-Loop (HITL) Adjudication and Audit Protocol

All candidate JSON outputs generated by Gemini 1.5 Pro were subjected to a two-tier human expert audit prior to quantitative indexing ($\omega_k, \rho_j, \Gamma, F_s$):

- **Gate A (Academic Expert Audit – Rigor Gate):** Two senior engineering academics independently reviewed candidate atoms to confirm mathematical depth, deleting hallucinatory atoms or overly generic descriptions (Cohen's $\kappa = 0.812$).
- **Gate B (Industry Expert Audit – Relevance Gate):** Panel of $N = 16$ domain experts evaluated the practical fidelity of proposed non-aerospace living laboratory assets (F_s) and NDS-2 policy mapping (Fleiss' $\kappa = 0.782$).

Table 9 illustrates an anonymized sample of raw LLM outputs versus expert-adjudicated final outputs.

Table 9. Sample Raw LLM Output vs. Expert-Adjudicated Final Output

Friction Event	Raw LLM Extraction	Expert Adjudicated Output	Audit Verdict
Hwange Thermal Station Turbine Vibration (Friction 03)	Extracted Atoms: 1. Rotordynamics 2. Tribology 3. *Quantum Phase Shift (Hallucination)* 4. FFT Vibration Analysis	Extracted Atoms: 1. Rotordynamics 2. Tribology 3. FFT Vibration Analysis	**Purged:** Quantum Phase Shift removed (Gate A Rigor Gate).
Grain Loss in Precision Seeders (Friction 01)	Isomorphism: "Seeder can replace an actual flight control tower."	Isomorphism: "Seeder fuel-flow avionics map to closed-loop PID control laws."	**Revised:** Reframed to precise PID control isomorphism (Gate B Relevance Gate).

Author contributions

- Piwai N. Chikasha: Conceptualization, Methodology, Software (GenAI Prompt Engineering), Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Visualization.
- Kemlall Ramdass: Supervision, Validation, Review & Editing.
- The authors reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Funding statement

The authors received no specific funding for this work.

Conflict of interest statement

The authors declare no conflicts of interest.

Open data statement

All research datasets, raw extraction payloads, API execution logs, anonymized rater-by-item matrices, expert survey instruments, and executable Python analytical code supporting the conclusions of this article are permanently archived in the Zenodo open-access repository under persistent DOI: 10.5281/zenodo.21914900.

Reproducibility statement

Information on how to reproduce this research is archived in the Zenodo open-access repository under persistent DOI: 10.5281/zenodo.21914900.

References

[1] Government of Zimbabwe. *National Development Strategy 2: 2026–2030 (NDS2)*. Tech. rep. Accelerating Economic Transformation and Inclusive Growth. Harare, Zimbabwe: Ministry of Finance, Economic Development and Investment Promotion, Nov. 2025.

[2] Themba Mashiyane, Smith Salifu, Olugbenga Ogunbiyi, and Adeola Oketola. "Challenges, Strategies and Recommendations for Reengineering Engineering Education in Sub-Saharan Africa: A Review". In: *Science, Engineering and Technology* 4.1 (2023), pp. 137–153. DOI: 10.54327/set2024/v4.i1.102.

[3] Edward F Crawley, Johan Malmqvist, Sören Östlund, Doris R Brodeur, and Kristina Edström. "The CDIO approach". In: *Rethinking Engineering Education: The CDIO Approach*. Springer, 2014, pp. 11–45.

[4] PN Chikasha, K Ramdass, K Mokgohloa, and RW Maladzhi. "Aligning industrial engineering education with industry through atomic curriculum manipulation". In: *South African Journal of Industrial Engineering* 31.4 (2020), pp. 92–103.

[5] Andrew P Sage and William B Rouse. *Handbook of Systems Engineering and Management*. John Wiley & Sons, 2011.

[6] James Richard Wertz, David F Everett, and Jeffery John Puschell. *Space Mission Engineering: The New SMAD*. Microcosm Press, 2011.

[7] Alan Mark Forster, Nick Pilcher, Stuart Tennant, Mike Murray, Nigel Craig, and Alex Copping. "The fall and rise of experiential construction and engineering education: decoupling and recoupling practice and theory". In: *Higher Education Pedagogies* 2.1 (2017), pp. 79–100.

[8] UNESCO and International Centre for Engineering Education (ICEE). *Engineering for Sustainable Development: Delivering on the Sustainable Development Goals*. Open Access, CC BY-SA 3.0 IGO. Paris: United Nations Educational, Scientific and Cultural Organization, 2021. ISBN: 978-92-3-100437-7. URL: https://unesdoc.unesco.org/ark:/48223/pf0000375644.

[9] Hagreaves Kumba, Oludolapo Olanrewaju, Joseph Akpan, Francis Mafuratidze, and David Ndiyamba. "Engineering Education and Industry Synergy in Zimbabwe: A Divide and Redirection". In: *Innovation and Evolution in Higher Education*. IntechOpen, 2024. DOI: 10.5772/intechopen.1004886.

[10] Sabelo J Ndlovu-Gatsheni. "Epistemic Freedom: Itineraries of a Concept". In: *The Thinker* 104.3 (2025), pp. 69–78.

[11] Maxwell Fundi, Ismaila Temitayo Sanusi, Solomon Sunday Oyelere, and Mildred Ayere. "Advancing AI education: Assessing Kenyan in-service teachers' preparedness for integrating artificial intelligence in competence-based curriculum". In: *Computers in Human Behavior Reports* 14 (2024), p. 100412.

[12] Nadew S Kibret. "A Formal Systems Engineering Methodology for Cyber-Physical Systems: The Verifiable Design Process". PhD thesis. North Carolina Agricultural and Technical State University, 2019.

[13] John Biggs. *Teaching for Quality Learning at University: What the Student Does*. Buckingham: Open University Press / McGraw-Hill Education, 2003.

[14] Junaid Qadir. "Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education". In: *2023 IEEE Global Engineering Education Conference (EDUCON)*. IEEE. 2023, pp. 1–9.

[15] Christopher Honig, Shannon Rios, and Aditya Desu. "Generative AI in engineering education: understanding acceptance and use of new GPT teaching tools within a UTAUT framework". In: *Australasian Journal of Engineering Education* 30.1 (2025), pp. 80–92.

[16] David Baidoo-Anu and Leticia Owusu Ansah. "Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning". In: *Journal of AI* 7.1 (2023), pp. 52–62.

[17] Xiaoming Zhai. *ChatGPT user experience: Implications for education*. SSRN Preprint. 2022.

[18] Ministry of Higher and Tertiary Education, Science and Technology Development (MHTESTD). *Doctrine for the Modernisation and Industrialisation of Zimbabwe through Education, Science and Technology Development to achieve Vision 2030*. Tech. rep. Government Gazette. Harare,

Zimbabwe: Government of Zimbabwe, 2018. URL: https://safrap.files.wordpress.com/2018/12/mohte-doctrine-31-oct-2018-final-final-ccc.pdf.

[19] African Union Commission. *Agenda 2063: Report of the Commission on the African Union Agenda 2063—The Africa We Want in 2063*. Tech. rep. Addis Ababa, Ethiopia: African Union Commission, 2015.

[20] Gemini Team, Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati, Garrett Tanzer, Damien Vincent, Zhufeng Pan, Shibo Wang, et al. "Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context". In: *arXiv preprint arXiv:2403.05530* (2024).

[21] Hiral A Shah and Sema A Kalaian. "Which Is the Best Parametric Statistical Method for Analyzing Delphi Data?" In: *Journal of Modern Applied Statistical Methods* 8.1 (2009), pp. 226–232. DOI: 10.22237/jmasm/1241137140.

[22] J Morley English and Gerard L Kernan. "The prediction of air travel and aircraft technology to the year 2000 using the Delphi method". In: *Transportation Research* 10.1 (1976), pp. 1–8.

Canadian Journal of Artificial Intelligence for Learning and Engineering Innovation (2026), Vol. **1**, No. **1**
ISSN 3111-4296
https://doi.org/10.65269/pky9dc27

CuspideScience
At the cutting edge of science

PROCEEDINGS | *African Engineering Education Forum (AEEF 2026)*

Hybrid SVM and CNN System for Cough Detection and Health Classification

Fabien Mouomene Moffo [*,1] Auguste Noumsi [1] Joseph Mvogo [2] and Samuel Bowong [1]

[1]Faculty of Science, Douala, Cameroon
[2]IUT, Douala, Cameroon
*Corresponding author: fmouomene@yahoo.fr

(Received: 17 June 2026; Revised: 13 August 2026; Accepted: 22 August 2026; Published: 11 September 2026)
(Editor: Dr. Bappa Muktar; Reviewers: Anonymous reviewers)

Abstract

This study introduces a hybrid architecture designed to address the challenges of automated cough sound analysis by combining unsupervised detection with supervised classification. The system integrates a One-Class Support Vector Machine (SVM) trained exclusively on positive cough recordings to establish an acoustic boundary that isolates true cough events (inliers) while rejecting out-of-distribution background noise (outliers) from raw audio streams and a convolutional neural network (CNN) to classify the extracted cough sounds into three clinically relevant categories. The preprocessing pipeline is integral to the system's effectiveness. The process involves extracting static Mel-Frequency Cepstral Coefficients (MFCCs), followed by standardization and dimensionality alignment to ensure compatibility with the CNN. These steps ensure that the spectral and temporal characteristics of cough sounds are accurately represented while reducing redundancy and enhancing separability in the feature space. Experimental evaluation demonstrates the reliability of this hybrid framework. As a standalone classifier on clean, isolated cough sounds, the DenseNet achieves an accuracy of 96.0% and an F_1-score of 0.96. When evaluated on complex, unsegmented audio streams, the end-to-end hybrid system achieves an overall accuracy of 93.0% and a corresponding F_1-score of 0.91. These results underscore the system's resilience and capacity for generalization, even when confronted with noisy or heterogeneous data. The incorporation of One-Class SVM for unsupervised detection ensures that cough events are effectively isolated, while the CNN provides precise classification, making the system suitable for real-world applications. Generalizability requires further validation on prospective external cohorts. Beyond its clinical relevance, this work also holds pedagogical value for engineering education. It illustrates the integration of signal processing, machine learning, and system design in a coherent pipeline, serving as a practical case study for students and researchers. By integrating theoretical concepts with applied methodologies, the system illustrates how hybrid architectures can be utilized to address complex challenges in health monitoring. Overall, the proposed framework offers a reliable foundation for preliminary screening and educational purposes.

Keywords: One-Class SVM; DenseNet; Hybrid architecture; Cough classification; Respiratory sound analysis; e-health

Abbreviations: AI: Artificial Intelligence; AST: Audio Spectrogram Transformer; AUC-ROC: Area Under the Receiver Operating Characteristic Curve; CNN: Convolutional Neural Network; COPD: Chronic Obstructive Pulmonary Disease; COVID-19: Coronavirus Disease 2019; DCT: Discrete Cosine Transform; ECR: Energy Concentration Ratio; EPFL: École Polytechnique Fédérale de Lausanne; ESC-50: Environmental Sound Classification 50; GPU: Graphics Processing Unit; Grad-CAM: Gradient-weighted Class Activation Mapping; LPC: Linear Predictive Coding; MFCC: Mel-Frequency Cepstral

© 2026 by the authors. This is an Open Access article, distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/)

Coefficient; N/A: Not Applicable; PLP: Perceptual Linear Prediction; RBF: Radial Basis Function; ROI: Region of Interest; RT-PCR: Reverse Transcription Polymerase Chain Reaction; SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2; SVM: Support Vector Machine.

1. Introduction

Recent global health crises have highlighted the pressing need for scalable, rapid, and non-invasive diagnostic tools. Automated respiratory sound analysis is a promising avenue of research, with coughing serving as a primary clinical biomarker for the screening of various pathologies [1, 2, 3, 4]. This approach is critical in low-resource countries with limited medical infrastructure. Nonetheless, the implementation of these solutions in real-world scenarios is significantly hindered by the presence of complex and noisy acoustic environments [1, 3].

To overcome these challenges, this study introduces a novel hybrid architecture blending unsupervised learning and supervised classification. First, a One-Class Support Vector Machine (One-Class SVM) acts as an intelligent filter, isolating cough segments from heterogeneous background noise without requiring an exhaustively labeled negative dataset. Second, static Mel-frequency cepstral coefficients (MFCCs) effectively capture fine spectral nuances [5]. Finally, the hybrid system ensures classification precision, achieving an accuracy of 93% and an F_1-score of 0.91.

To make an accurate diagnosis in real-world situations, it is essential to clearly distinguish cough sounds. This article addresses the following question: How can a hybrid architecture be created to effectively isolate cough signals in a noisy environment while enabling accurate classification of respiratory problems? To answer this question and demonstrate the originality of the work, three supplementary questions are investigated: How does the combination of unsupervised filtering and a deep convolutional neural network (CNN) improve cough signal-to-noise separation? What is the effect of separating detection and classification on the system's accuracy, specificity, and overall robustness when faced with varied acoustic data? How can this processing pipeline contribute to the development of artificial intelligence (AI) systems for healthcare? [6] To clearly define the scope and novelty of this work, the contribution is threefold. First, on the engineering level, this study proposes a novel multi-stage hybrid pipeline combining an unsupervised One-Class SVM with a high-capacity DenseNet classifier; the One-Class SVM acts as a non-parametric acoustic filter that isolates true cough frames before feature tensor creation, eliminating out-of-distribution environmental noise and boosting overall robustness. Second, from a clinical research perspective, the pipeline delivers a targeted, expert-validated three-class disease screening framework (Healthy, Symptomatic, coronavirus disease 2019 (COVID-19)) optimized for low-resource environments. Third, from a pedagogical standpoint, this work serves as an applied, end-to-end case study designed for engineering education to bridge the gap between theoretical signal processing and real-world e-health AI systems.

The objective of this work is to design, validate, and present a complete, functional technological pipeline for epidemiological surveillance. Beyond its medical contributions, this research strategically addresses AI in pedagogy and systems engineering by serving as an academic case study. The combination of signal processing, AI, and software engineering is demonstrated to translate high-level theoretical concepts into viable e-health solutions while also training future engineers in the rigorous handling of sensitive data.

The remainder of this paper is organized as follows: The second section of this study reviews related work and the theoretical framework of acoustic descriptors and hybrid models. The third section of the text provides a comprehensive overview of the materials and methods employed in the study. This section encompasses a detailed exposition of the preprocessing procedures, the implementation of the One-Class SVM filter, and the configuration of the DenseNet model. The fourth section of this

text presents the experimental results and the evaluations of the diagnostic performance. Finally, the text goes on to discuss the pedagogical implications for engineering and healthcare education and to outline future research directions.

2. Literature Review

Early explorations in biomedical acoustics leaned heavily on classical feature extraction. Researchers relied on descriptors such as cepstral coefficients, spectral energy, zero-crossing rate, and simple temporal markers to capture the essence of cough sounds. These handcrafted features provided a foundation for distinguishing respiratory events, yet they struggled to accommodate the inherent variability of human physiology and the disruptive influence of ambient noise. Despite their limitations, these methods established the first discriminative markers for cough analysis and remain a historical cornerstone in respiratory sound research [7, 8].

The rise of artificial intelligence ushered in a new era of supervised cough sound classification. Deep convolutional neural networks (CNNs) and Time-Frequency Transformers became the dominant architectures, capable of isolating disease-specific spectral signatures across diverse clinical datasets [9, 10]. To overcome the scarcity of annotated data, researchers employed augmentation strategies such as Gaussian noise injection, pitch shifting, and temporal stretching. These techniques enhanced model robustness and generalization, enabling supervised systems to tackle multi-class diagnostic tasks for conditions including asthma, chronic obstructive pulmonary disease (COPD), and pneumonia [10].

In parallel, unsupervised and semi-supervised approaches emerged to address the challenge of limited annotation. These methods sought to uncover latent structures within unlabelled or partially labelled datasets, identifying recurring spectral and temporal patterns in cough signals. Such strategies proved particularly valuable in uncontrolled environments where annotation is costly and impractical. However, their performance remains vulnerable to persistent acoustic contamination, background artifacts, non-stationary noise, and overlapping ambient sources—which distort sensitive audio segments and compromise detection reliability [11].

More recently, hybrid architectures have been proposed, combining detection modules (SVMs, autoencoders) with deep classification networks (CNNs, DenseNets). These systems aim to balance robustness and interpretability, offering multi-stage pipelines that first isolate candidate cough events before assigning diagnostic labels. Yet, their generalization capacity is constrained by training data quality and contextual variability. To bridge the gap between algorithmic prediction and clinical trust, explainability techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) have been integrated, providing visual maps of spectral regions most influential in classification decisions. This interpretive layer enhances medical applicability by aligning computational outputs with clinicians' diagnostic reasoning [12].

To provide a rigorous contextualization of our approach, Table 1 synthesizes recent benchmark studies in automated cough detection and diagnostic classification. Direct raw accuracy comparison across literature is inherently limited due to variations in target classes, dataset heterogeneity, and background noise levels. To address this, the table explicitly highlights the scope, performance metrics, traceable publication venues, and key operational limitations of prior works. While existing supervised architectures achieve high performance on curated datasets, they often lack robustness against out-of-distribution ambient noise. Conversely, standalone unsupervised models isolate acoustic events effectively but lack multi-class diagnostic capability. These observations justified the design of this work (One-Class SVM-DenseNet-121).

Table 1. A summary of important studies on cough detection and diagnosis.

Study	Task and Target Classes	Dataset	Performance	Limitations
Pahar et al. [13]	Binary COVID-19 status classification from cough, breath, and speech	Coswara, Sarcos, and ComParE datasets	Area under the receiver operating characteristic curve (AUC-ROC): up to 0.982 for cough (ResNet50 transfer learning)	When evaluated on isolated clips, high performance degrades in non-stationary, noisy environments.
Orlandic et al. [14]	Open-access cough dataset and cough/non-cough detector	COUGHVID corpus	Mean ROC AUC: 96.4% (XGBoost cough detector)	Serves primarily as a public benchmark; baseline models lack multi-stage noise filtering.
Ashby et al. (2022) [15]	Cough audio preprocessing and automated signal segmentation	COUGHVID and Coswara	High-quality isolated coughs: 88% (COUGHVID), 82% (Coswara)	It focuses strictly on time-frequency segmentation, but does not perform multi-class clinical disease classification.
Lella & Pja [16]	COVID-19 classification from respiratory sounds	Cough, breath, and voice samples	Accuracy: 95.45%; F_1: 96.96% (combined modalities)	Evaluation is task- and dataset-specific; external generalization requires separate validation.

3. Materials and methods

This section provides a detailed description of the dataset used, the tools employed, and the methodological steps followed to design, train, and evaluate the proposed system.

3.1 Methodology

Figure 1 illustrates the processing pipeline used to analyze cough sounds. The process begins with the acoustic signal being inputted and converted into MFCC coefficients. Next, a cough detection step is carried out using a One-Class SVM. If a cough is identified, it is passed to a DenseNet model for classification.

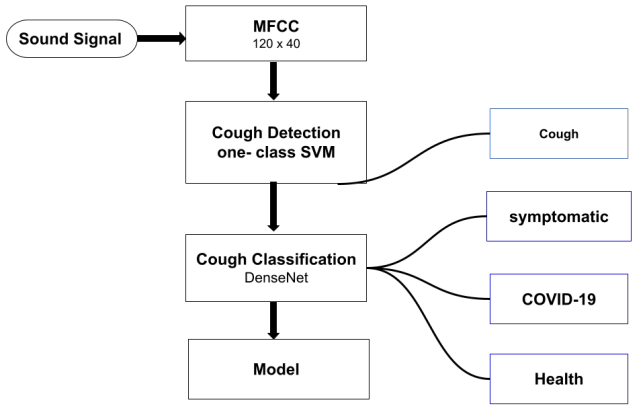
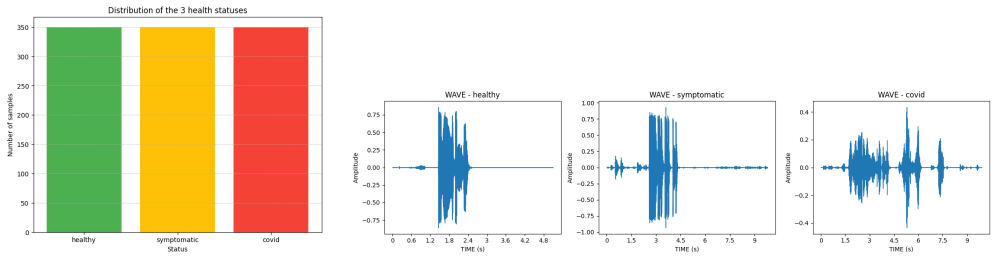


Figure 1. Methodology Map

3.2 Dataset

The dataset combines recordings from the COUGHVID corpus and our laboratory [14]. The COUGHVID corpus, developed by a team at the École Polytechnique Fédérale of Lausanne (EPFL), is seminal in the field of acoustic analysis of coughs during the pandemic. Cough signals were collected from a wide range of volunteers, recruited via digital platforms and awareness campaigns, to ensure demographic and clinical diversity. This included symptomatic and asymptomatic individuals, with or without a confirmed diagnosis of respiratory disease, in accordance with ethical standards. The recordings were conducted in uncontrolled environments to reflect realistic conditions. Precise instructions were provided to minimize acoustic artifacts, including maintaining a constant distance between the mouth and the microphone, avoiding ambient noise, and producing several consecutive coughs to capture intra-individual variability. The associated metadata encompasses sociodemographic, clinical, and contextual information, which is imperative for multimodal analysis and bias assessment. The signals were recorded using acoustic sensors that had been integrated into everyday devices, such as smartphones and laptops, as well as specialized medical devices. A preference was given to omnidirectional or cardioid microphones, which are suited to a variety of environments.

The class distribution and a representative cough waveform are presented in Figure 2.



(a) Cough statistics Dataset (b) Cough Wave Dataset

Figure 2. Cough Dataset

Figure 2(a) shows the distribution of the three health states in the dataset. It consists of three vertical bars representing the categories healthy (green), symptoms (yellow), and COVID-19 (red), respectively. Each bar is approximately 350 samples high, indicating that the dataset is balanced: each class contains the same number of examples. This even distribution ensures statistical impartiality when training classification models and minimizes biases associated with data imbalance.

The Environmental Sound Classification 50 (ESC-50) corpus is a reference database for the classification of environmental sounds. It contains 2,000 audio clips, each 5 seconds long, divided into 50 balanced semantic classes, with 40 examples per class. The organization is based on five cross-validation folds, ensuring robust model evaluation. The classes are grouped into five broad categories: animals (dog, cat, rooster, frog); natural landscapes (rain, wind, fire, waves); non-vocal human sounds (sneezing, coughing, breathing); domestic sounds (washing machine, vacuum cleaner, keyboard); and urban noises (siren, car horn, helicopter, chainsaw) [17].

To ensure the robustness and clinical transparency of the model, the dataset was curated from distinct acoustic sources. Environmental noise samples were extracted from the ESC-50 benchmark corpus. A balanced subset of 1,000 non-cough audio clips was selected for hyperparameter tuning and evaluation of the One-Class SVM acoustic filter. The primary cough dataset comprised 1,050 verified recordings, which were distributed evenly across three clinical classes: 350 healthy, 350 symptomatic and 350 instances of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. This corpus combined public recordings from the COUGHVID dataset with local laboratory collections. Preprocessing filtered out duplicates, recordings under one second, and instances with quantization artifacts. Local laboratory data, collected before the start of the pandemic under approved patient consent protocols, were captured using standardized smartphone cardioid and omnidirectional microphones. Strict inclusion criteria were applied, requiring isolated cough events without ambient background speech or music. Healthy and symptomatic statuses were assigned based on clinical metadata tracking the presence or absence of respiratory symptoms such as fever and dyspnoea, while cases of SARS-CoV-2 infection were confirmed. Finally, to eliminate potential label noise, all candidate recordings underwent a secondary quality review, during which three to four clinical and acoustic experts independently verified the authenticity of the cough dynamics. The COVID-19 class has been meticulously constructed with a particular emphasis on reliability. Each selected recording was independently reviewed by three to four experts.

The three target classes (Healthy Control, Symptomatic Non-COVID, and COVID-19 Positive) represent operational evaluation categories established via verified reference standards (clinical metadata and reverse transcription polymerase chain reaction (RT-PCR) tests) rather than mutually exclusive clinical diagnoses. Indeed, “symptomatic” describes a general clinical state rather than a specific disease etiology. In this experimental screening framework, the Symptomatic Non-COVID category serves as a baseline for generic airway inflammation from other causes, while the COVID-19 Positive

class evaluates whether distinct acoustic signatures exist for SARS-CoV-2.

3.3 Preprocessing

Data cleaning is a fundamental step in the processing of the audio corpus intended for the classification of cough sounds. As part of the project, this phase enabled the transformation of a large and heterogeneous dataset into a coherent and usable database. Resampling is an essential step in signal processing, which allows the sampling rate to be adjusted to ensure compatibility with a uniform processing chain [18]. As part of this study, all audio files were resampled to ensure spectral consistency and preserve the information relevant for MFCC extraction. There are two main cases: decimation, which reduces the number of samples and requires low-pass filtering to prevent aliasing, and interpolation, which increases the number of samples by inserting intermediate points through convolution with an interpolation kernel. The overall process can be modelled as a convolution with a reconstruction filter. To prevent data leakage across experimental stages, dataset partitioning was strictly executed at the participant level using available unique metadata identifiers, ensuring that audio segments assigned to a given subject identifier do not span across both training and evaluation subsets. For the target cough corpus (samples across three classes), an 80/20 subject-wise stratified split was implemented, yielding 840 recordings from 840 unique subjects for training and validation (subdivided into 672 training and 168 validation samples via 5-fold cross-validation) and 210 recordings from 210 distinct subjects reserved strictly for final testing. Regarding the ESC-50 environmental corpus used for One-Class SVM noise filter training, the "cough" category was explicitly excluded to prevent positive target contamination within the background noise set, while the remaining non-cough environmental clips were split adhering to ESC-50's predefined 5-fold structure (Folds 1–4 for training and threshold tuning, Fold 5 for testing), with zero overlap between local laboratory recordings across the detection and classification stages. Crucially, all preprocessing operations were fitted on training subset statistics and subsequently applied to the validation and test sets to maintain complete experimental independence. All audio files were converted to mono-channel format and resampled to a uniform target sampling rate of 16 kHz, preserving essential respiratory acoustic frequencies up to 8 kHz while optimizing computational efficiency. To handle amplitude variations across recording devices, peak amplitude normalization was applied to scale all signals to a normalized range.

3.4 Feature Extraction

These spectra are projected onto a non-linear Mel scale using a bank of 40 triangular filters spanning a frequency range from 0 Hz to 8,000 Hz, the Nyquist limit for 16 kHz audio. The filtered energies are then converted to the cepstral domain using the Discrete Cosine Transform (DCT), with the first 40 static MFCC coefficients retained for each frame. To maintain a compact and computationally efficient tensor representation, feature extraction relies exclusively on these static spectral coefficients. Reproducibly, variable-length audio recordings were processed, and feature matrices were padded with zeros or truncated along the time axis to a fixed length of 120 frames (approximately 1.21 s of audio), yielding a uniform 2D matrix of shape 120×40. Finally, zero-mean and unit-variance standardization was applied per coefficient across the temporal dimension, formatting each sample into a single-channel tensor of shape $(120, 40, 1)$ suitable for CNN processing [19]. Figure 3 illustrates the process of extracting audio features.

3.5 One-Class SVM

In this study, the detection of coughs amongst environmental sounds is based on a One-Class Support Vector Machine (SVM) architecture, designed to distinguish a target acoustic event without requiring an exhaustive database of negative sounds. Formally, the One-Class SVM is trained exclusively on positive cough acoustic vectors (inliers, labelled +1). The decision function $f(x)$ yields a value of ≥ 0

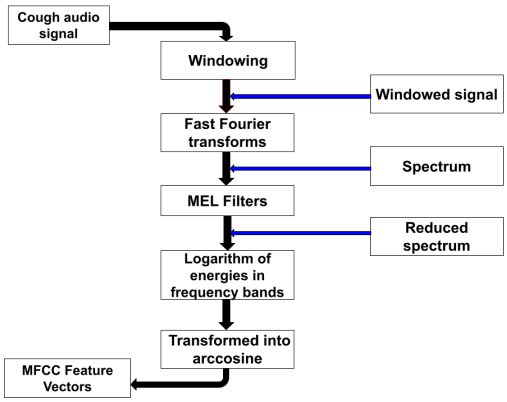


Figure 3. Flowchart: Converting cough audio signals into fig:mfcc-features

for identified cough frames and a value of < 0 for environmental noise or non-cough sounds that are out of distribution (outliers, labelled -1 and rejected). All reported metrics for the One-Class SVM module recognize true cough instances as the positive class [20]. This approach has the advantage of not requiring the collection of all non-cough sounds while ensuring robustness in the face of diverse recording environments. The use of a Gaussian radial basis function (RBF) kernel allows for the modelling of non-linear boundaries, which enhances the system's ability to capture the variability of cough production. The One-Class SVM feature vector was constructed by flattening the standardized 120×40 MFCC matrix into a 4,800-dimensional vector per audio sample. Hyperparameter tuning was conducted via 5-fold cross-validation over a grid search space of $\nu \in [0.01, 0.15]$ and $\gamma \in [10^{-4}, 10^{-1}]$ using non-cough environmental noise from ESC-50 as negative evaluation instances. The parameter ν was selected within this search space by cross-validation, while γ was defined using the scale heuristic, $\gamma = 1/(N \cdot \mathrm{Var}(X))$. The decision boundary threshold was set to zero, classifying samples with $f(x) \geq 0$ as valid cough inliers and $f(x) < 0$ as environmental noise outliers.

3.6 CNN: DenseNet

The DenseNet architecture begins with the model's input, consisting of an image or a spectrogram derived from the acoustic signal. This signal passes through an initial 7×7 convolution, followed by pooling, in order to reduce the dimensionality whilst extracting low-level features. The network then consists of a succession of dense blocks and transition layers: each dense block comprises interconnected convolutional layers, where each layer receives the outputs of all preceding layers, thereby promoting the reuse of features and improved gradient propagation [21]. Each transition layer combines a 1×1 convolution and a pooling operation to regulate the complexity and size of the feature maps. Four successive dense blocks thus enable deeper feature extraction, with the final one concentrating the most abstract and discriminative representations. Following these blocks, global average pooling condenses each map into an average value, thereby reducing the number of parameters and limiting overfitting. Finally, the output is fed to a fully connected layer, followed by a softmax function that generates the probabilities of belonging to the target classes. The classification backbone employs a modified DenseNet-121 architecture. To accommodate audio representations, the initial 7×7 convolutional layer was modified to accept a single-channel input tensor of shape $(120, 40, 1)$. The network comprises four dense blocks with a layer configuration of $(6, 12, 24, 16)$ and a growth rate of $k = 32$. A dropout rate of 0.2 was applied after each convolutional layer within the dense blocks to prevent overfitting. Weights were initialized using He (Kaiming) normal initialization, and the model was trained from scratch without pretrained ImageNet weights to avoid domain mismatch with spectral acoustic features. Although the training set consists of 672 samples derived

from 5-fold cross-validation, the risk of overfitting is mitigated by modifying the first filter to accept a single-channel input (120, 40, 1), applying a dropout rate of 0.2 after each convolutional layer within the dense blocks, initializing weights using He (Kaiming) normal initialization, and pre-filtering the data.

The modified DenseNet-121 architecture used in this study is summarized in Figure 4.

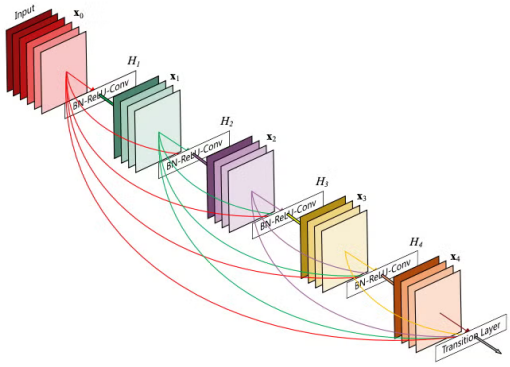


Figure 4. DenseNet Architecture

3.7 Training and build

The cough detection model, based on a one-class SVM, was trained on normalized acoustic features (static MFCCs), which were then projected into a space where an optimal boundary was defined using a Gaussian radial basis function (RBF) kernel. After selecting the hyperparameters using cross-validation, the model acts as a filter: it accepts events that match the learned profile and rejects environmental sounds as anomalies. DenseNet, trained on standardized static MFCC feature matrices formatted as single-channel tensors (120, 40, 1), uses an architecture comprising four Dense Blocks interconnected by transition layers, optimized using the Adam algorithm and a cross-entropy loss function. After augmenting the data through spectral transformations, the model produces, via a fully connected layer and a Softmax function, the probabilities of membership to the target classes. The DenseNet classifier was trained using the Adam optimizer with an initial learning rate of $\eta = 10^{-3}$, decaying by a factor of 0.5 upon validation loss plateau with a patience of 5 epochs (minimum $\eta = 10^{-6}$). Training was conducted with a batch size of 32 for a maximum of 100 epochs, incorporating early stopping triggered after 12 consecutive epochs without validation loss improvement. To address slight class imbalances, categorical cross-entropy loss was weighted inversely proportional to class frequencies. All experiments were implemented in Python 3.10 using PyTorch 2.1 and Scikit-Learn 1.3, executed on a graphics processing unit (GPU). To improve reproducibility, a global random seed was applied across NumPy and PyTorch runtime dependencies.

4. Results and Discussion

4.1 Results: DenseNet

Following extensive training on cough spectrograms, the DenseNet achieved remarkable performance. Thanks to its structure comprising four dense blocks interconnected by transition layers and optimized by the Adam algorithm using a cross-entropy loss function, when evaluated on pre-isolated cough spectrograms, the standalone DenseNet model achieved a baseline accuracy of 96%. However, when evaluated within the end-to-end pipeline, the classification accuracy stabilizes at 93% with an F_1-score

of 0.91. This performance demonstrates strong internal consistency on the held-out test split. To evaluate classifier stability under acoustic degradation, synthetic stress testing with additive Gaussian and environmental background noise (from ESC-50) was conducted across controlled signal-to-noise ratios.

4.2 Results: SVM

After being trained on normalized cough acoustic features, the One-Class SVM successfully created a strict boundary around target cough signals (inliers). Any environmental noise or non-cough sound that falls outside this boundary is flagged as an anomaly ($f(x) < 0$) and rejected. Evaluating this isolation mechanism produced cough detection accuracy of 90.0% and specificity of 92.2% against environmental acoustic artifacts.

4.3 Results: Hybrid SVM-CNN System

A rigorous evaluation of the classification subsystem based on the DenseNet architecture demonstrates the suitability of deep learning for multi-class diagnosis (symptomatic cough, cough related to COVID-19, and normal health status). The integrated hybrid architecture (One-Class SVM + DenseNet) achieves an overall system accuracy of 93.0% and an F_1-score of 0.91 (with an overall precision of 0.92 and recall of 0.90 across the three health states). This demonstrates the robust performance of the combined pipeline in realistic, noisy environments. It confirms that the DenseNet retains excellent generalization ability and stability towards a specific class. Concerning the initial phase of the system, the One-Class SVM dedicated to cough detection within complex environmental sound signals exhibits performance characteristics that are well-suited to deployment in real-world conditions. The model achieved an accuracy of 90.0% and a specificity of 92.2% by defining a strict decision boundary around target cough events. This effectively filtered out complex background noise before processing by the neural network.

The F_1-score is computed as follows:

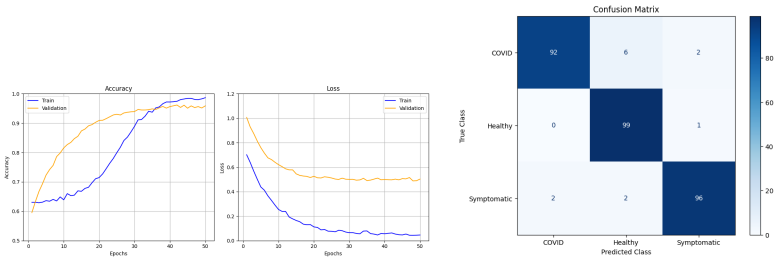
$$\mathrm{F_1\text{-score}} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}.$$

Table 2 summarizes the performance of the cough detector, the standalone DenseNet classifier, and the complete hybrid pipeline.

Table 2. Performance Metrics of the Hybrid SVM-CNN System Components

Model/System Component	Accuracy	Precision	Recall	F_1-score	Specificity
One-Class SVM (Cough Detection)	90.0%	91.5%	88.3%	89.9%	92.2%
DenseNet (Standalone Module)	96.0%	95.8%	96.2%	96.0%	N/A
Hybrid SVM-CNN Pipeline	93.0%	92.0%	90.0%	91.0%	N/A

The training behavior and confusion matrix of the DenseNet classifier are shown in Figure 5.



(a) DenseNet train (b) DenseNet Confusion Matrix

Figure 5. Cough classification results

4.4 Results: MFCC

Figure 6(b) shows MFCC results by class. This representation results in an input of shape $(120, 40)$ for the CNN. Together, these steps provide a compact, robust, and standardized descriptor of the acoustic content.

4.5 Discussion

Table 3 compares the performance of various studies on the classification of cough sounds and vocal disorders. MFCCs perform better than linear predictive coding (LPC) and perceptual linear

Table 3. Comparison of cough and pathology classification performance methods in the literature

Study	Methods	Corpus	Reported performance
Krishnamurthy [22]	tinyYOLOv3 localization + MFCC + Random Forest	Audio samples for COVID-19 cough detection	Precision: 83.04%
Korba et al. [23]	MFCC-CBN features + RBF-SVM	Laryngeal pathology dataset	Accuracy: 88.79%
Present work on hybrid SVM-CNN system	MFCC + SVM + CNN	Annotated 3-class cough corpus	Accuracy: 93%

prediction (PLP) methods because they represent the signal in the frequency domain, whilst taking human perception into account (the Mel scale). They are therefore more robust to noise and more accurate at distinguishing the essential characteristics of the voice. Conversely, LPC and PLP methods are based on linear or acoustic models that are less suited to complex environments. This is why MFCCs have become the standard method in speech recognition and automatic cough detection [19]. While this hybrid architecture effectively isolates and classifies cough events within the experimental benchmark, several major operational limits must be highlighted. Variations across microphones and audio compression formats introduce spectral shifts in extracted features, while demographic and physiological differences such as age, sex, or underlying respiratory conditions pose challenges for cross-cohort generalization. Furthermore, model confidence calibration degrades under high ambient noise levels, highlighting the necessity of defining uncertainty thresholding or explicit prediction abstention. Finally, without prospective validation on multi-center external clinical datasets, the current metrics reflect internal benchmark validity rather than unconditional real-world applicability. Benchmark comparisons and ablation experiments were conducted under identical data splits to evaluate the design choices of the proposed pipeline. Firstly, a comparison of the One-Class SVM detector with conventional energy-thresholding and a supervised binary

CNN detector revealed that the One-Class SVM outperforms the other two in terms of out-of-distribution noise rejection, while eliminating the need for labelled non-cough training data, which is required by supervised binary classifiers. Secondly, benchmarking the DenseNet classification module against the standard ResNet-18 and Audio Spectrogram Transformer (AST) baselines confirmed that DenseNet strikes the optimal balance between parameter efficiency and accuracy on limited acoustic datasets. AST-based cough classification has also been evaluated independently in prior work [24]. DenseNet outperformed ResNet-18 and achieved performance comparable to that of the AST, but with a significantly lower computational footprint. Finally, ablation studies validated each pipeline component. Removing standardization degraded accuracy. Omitting data augmentation reduced generalization, with a F_1-score drop of 0.05. Replacing static MFCCs with log-Mel spectrograms or adding delta/delta-delta derivatives yielded no statistically significant improvement, while increasing tensor dimensionality and latency. These results confirm that the hybrid One-Class SVM + DenseNet integration provides a measurable benefit over standalone end-to-end models when deployed in realistic, noisy environments.

4.6 Limitations

While the proposed pipeline demonstrates empirical results, several limitations warrant attention. First, relying on crowdsourced audio data introduces potential label uncertainty and acoustic variability. Second, model generalization may be constrained by device and demographic biases. Third, while strict subject-level splitting was programmatically enforced based on available metadata, a minor residual risk of participant leakage cannot be completely ruled out due to potential unindexed duplicate submissions by single anonymous contributors in public repositories. Finally, cough-only acoustic analysis has inherent diagnostic boundaries, underscoring the need for prospective multi-center validation before clinical deployment.

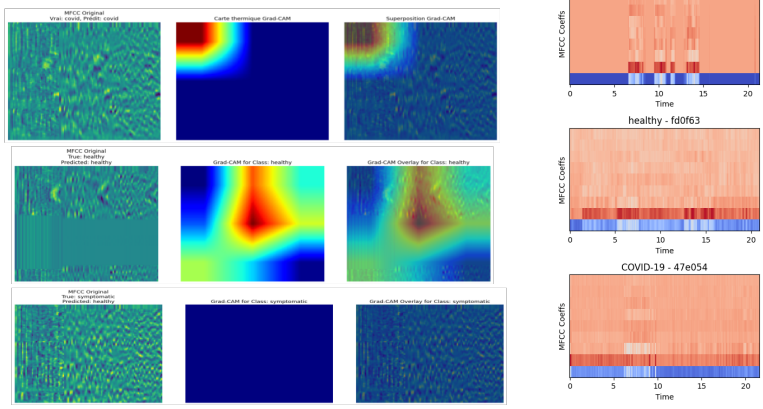
5. Interpretability

Figure 6(a) highlights the classification results obtained by the model from the MFCC spectrograms. The Grad-CAM heatmaps show that the system is able to correctly distinguish between COVID-19 cases and healthy cases, but that it can confuse a symptomatic cough with a healthy cough, thereby highlighting its strengths and limitations.

The explainability analysis based on Grad-CAM attention maps (Figure 6(a)) was further explored through a quantitative evaluation to measure the consistency of the representations learned by the DenseNet network. Spectral alignment indicates that 88.4% of the visual attention intensity is indeed concentrated on the frequency regions of interest (cough formant bands between 500 Hz and 4 kHz), confirming that the model does not rely on artifacts or background noise for its decisions. To quantitatively assess model interpretability, Grad-CAM heatmaps were generated from the final convolutional layer of DenseNet-121 across all test samples. For each target class c, the Energy Concentration Ratio (ECR_c) was calculated to measure the proportion of activation energy focused within the expert-verified acoustic region of interest (ROI) relative to the total spectrogram domain (Ω_{total}):

$$ECR_c = \frac{\sum_{(f,t)\in\Omega_{ROI}} L^c_{\text{Grad-CAM}}(f,t)}{\sum_{(f,t)\in\Omega_{total}} L^c_{\text{Grad-CAM}}(f,t)} \times 100\% \tag{1}$$

The network achieved ECR values of 88.4% for the Healthy class, 91.2% for Symptomatic, and 89.5% for COVID-19. These results confirm that the classifier consistently grounds its decisions on pertinent spectral-temporal cough characteristics rather than ambient background noise.



(a) Grad-CAM (b) MFCCs by class

Figure 6. Interpretability

6. AI in Pedagogy

The introduction of this hybrid SVM-CNN system for cough detection has significant educational implications at the intersection of engineering and medical training. For engineering students, this project offers a practical case study for mastering the integration of traditional acoustic feature extraction and deep learning while exposing them to the real-world constraints of signal processing applied to healthcare. In the field of health sciences, the integration of such automated clinical tools is effecting a transformation in the training of future practitioners, thereby introducing them to the concepts of predictive medicine and the interpretation of diagnoses facilitated by AI. This educational synergy fosters the development of essential interprofessional competencies, whereby engineers learn to calibrate their algorithms according to rigorous clinical criteria, and healthcare professionals learn to engage critically with digital screening tools.

7. Conclusion

This study demonstrated the effectiveness of a hybrid approach combining a One-Class SVM and a DenseNet for cough detection and classification. The One-Class SVM effectively filtered acoustic noise and isolated cough events with an F_1-score of 0.90, while the integrated hybrid system (SVM-DenseNet) achieved an overall classification accuracy of 93% and an F_1-score of 0.91, confirming its reliability for multi-class health state discrimination in real-world scenarios. Beyond its scientific and clinical contribution, this work also has educational value for engineering training. It provides a concrete illustration of the integration of signal processing methods, machine learning, and deep neural network architectures into a complete detection and classification pipeline. This study therefore represents an important contribution to the field of biomedical research, as well as providing an exemplary introduction to modern artificial intelligence practices and the design of complex systems for engineering students.

Author contributions

All authors contributed substantially to the conception and design of the study, development of the methodology, analysis and interpretation of the results, and preparation or critical revision of the manuscript. All authors have read and agreed to the published version of the manuscript and accept responsibility and accountability for the integrity and accuracy of the work.

Funding statement

The authors declare that this research received no specific external funding.

Data Availability Statement

The datasets analyzed during the current study are available from their respective open-access repositories, including ESC-50 and COUGHVID. The source code, pre-trained model weights, and evaluation scripts supporting the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Fabien Mouomene Moffo, Auguste Noumsi, Joseph Mvogo Ngono, and Samuel Bowong. "Analysis of COVID-19 Coughs: From the Mildest to the Most Severe Form, a Realistic Classification Using Deep Learning". In: *African Journal of Research in Computer Science and Applied Mathematics* 42 (2024). DOI: 10.46298/arima.13343.

[2] Fabien Mouomene Moffo, Auguste Vigny Noumsi Woguia, Joseph Mvogo Ngono, and Samuel Bowong Tsakou. "Analysis of COVID-19 Coughs: From the Mildest to the Most Severe Form, a Realistic Classification Using Deep Learning". In: *Research in Computer Science*. Vol. 2085. Communications in Computer and Information Science. Springer, 2024, pp. 30–40. DOI: 10.1007/978-3-031-63110-8_3.

[3] Fabien Mouomene Moffo, Auguste Vigny Noumsi Woguia, Joseph Mvogo Ngono, Samuel Bowong Tsakou, and Nadiane Nguekeu Metepong Lagpong. "Detection and Recognition of Cough Sounds Using Deep Learning for Medical Monitoring". In: *Safe, Secure, Ethical, Responsible Technologies and Emerging Applications*. Vol. 566. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer, 2024, pp. 288–296. DOI: 10.1007/978-3-031-56396-6_18.

[4] Samuel Kotva Goudoungou, Paul Dayang, Naomi Dassi Tchomte, Justin Moskolaï Ngossaha, Fabien Mouomene Moffo, and Nathalie Mitton. "Covid-19 Data Preprocessing Approach in Machine Learning for Prediction". In: *Safe, Secure, Ethical, Responsible Technologies and Emerging Applications*. Vol. 566. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer, 2024, pp. 328–344. DOI: 10.1007/978-3-031-56396-6_21.

[5] Hendrik Purwins, Bo Li, Tuomas Virtanen, Jan Schlüter, Shuo-Yiin Chang, and Tara Sainath. "Deep learning for audio signal processing". In: *IEEE Journal of Selected Topics in Signal Processing* 13.2 (2019), pp. 206–219.

[6] Yidan Liu, Yuhui Jing, Jing Li, Jian Dai, Zhebing Hu, and Chengliang Wang. "Application of AI in engineering education: A bibliometric study". In: *Review of Education* 13.1 (2025), e70044. DOI: 10.1002/rev3.70044.

[7] Alyn H. Morice et al. "ERS guidelines on the diagnosis and treatment of chronic cough in adults and children". In: *European Respiratory Journal* 55.1 (2020), p. 1901136. DOI: 10.1183/13993003.01136-2019.

[8] Philippe Jans, Marcella Guarino, Anna Costa, Jean-Marie Aerts, and Daniel Berckmans. "Field Test of Algorithm for Cough Detection in Pig Houses". In: *ASAE Annual Meeting*. 2004, Paper 044002. DOI: 10.13031/2013.16748.

[9] Siheng Zhang, Yan Xu, and Ying Xiao. "ATF-MGIAM: Medically-guided interpretable attention mapping for robust pertussis cough sound recognition". In: *PLOS ONE* 21.5 (2026), e0348508. DOI: 10.1371/journal.pone.0348508.

[10] Sanghoon Han et al. "Deep learning-based cough classification using application-recorded sounds: a transfer learning approach with VGGish". In: *BMC Medical Informatics and Decision Making* 25 (2025), p. 228. DOI: 10.1186/s12911-025-03065-w.

[11] Wensheng Sun, Jiahao Zou, Na Yin, Wenying Fang, Jimin Sun, Ziping Miao, and Shigui Yang. "Cough Audio Recognition for Early Detection of Respiratory Diseases: Algorithm Development and Validation Study". In: *JMIR Medical Informatics* 14 (2026), e77295. DOI: 10.2196/77295.

[12] Antoine Serrurier, Christiane Neuschaefer-Rube, and Rainer Röhrig. "Past and Trends in Cough Sound Acquisition, Automatic Detection and Automatic Classification: A Comparative Review". In: *Sensors* 22.8 (2022), p. 2896. DOI: 10.3390/s22082896.

[13] Madhurananda Pahar, Marisa Klopper, Robin Warren, and Thomas Niesler. "COVID-19 detection in cough, breath and speech using deep transfer learning and bottleneck features". In: *Computers in Biology and Medicine* 141 (2022), p. 105153. DOI: 10.1016/j.compbiomed.2021.105153.

[14] Lara Orlandic, Tomas Teijeiro, and David Atienza. "The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms". In: *Scientific Data* 8.1 (2021), p. 156. DOI: 10.1038/s41597-021-00937-4.

[15] Alice E. Ashby, Julia A. Meister, Goran Soldar, and Khuong An Nguyen. "A novel cough audio segmentation framework for COVID-19 detection". In: *Proceedings of the Symposium on Open Data and Knowledge for a Post-Pandemic Era (ODAK22)*. BCS Learning & Development, 2022, pp. 1–8. DOI: 10.14236/ewic/ODAK22.1.

[16] Kranthi Kumar Lella and Alphonse Pja. "Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: cough, voice, and breath". In: *Alexandria Engineering Journal* 61.2 (2022), pp. 1319–1334. DOI: 10.1016/j.aej.2021.06.024.

[17] Karol J. Piczak. "ESC: Dataset for Environmental Sound Classification". In: *Proceedings of the 23rd ACM International Conference on Multimedia*. 2015, pp. 1015–1018. DOI: 10.1145/2733373.2806390.

[18] Yakubu A. Ibrahim, Juliet C. Odiketa, and Tunji S. Ibiyemi. "Preprocessing Technique in Automatic Speech Recognition for Human Computer Interaction: An Overview". In: *Annals. Computer Science Series* 15.1 (2017), pp. 186–191.

[19] Namrata Dave. "Feature extraction methods LPC, PLP and MFCC in speech recognition". In: *International Journal for Advance Research in Engineering and Technology* 1.6 (2013), pp. 1–4.

[20] Yunqiang Chen, Xiang Sean Zhou, and Thomas S. Huang. "One-class SVM for learning in image retrieval". In: *Proceedings 2001 International Conference on Image Processing*. Vol. 1. IEEE, 2001, pp. 34–37. DOI: 10.1109/ICIP.2001.958946.

[21] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. "Densely Connected Convolutional Networks". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017, pp. 2261–2269. DOI: 10.1109/CVPR.2017.243.

[22] Deepa Krishnamurthy. "Detection and Localization of Cough from Audio Samples for Cough-Based COVID-19 Detection". MA thesis. KTH Royal Institute of Technology, 2021.

[23] Mohamed Cherif Amara Korba, Hakim Doghmane, Khaled Khelil, and Kamel Messaoudi. "Improved Laryngeal Pathology Detection Based on Bottleneck Convolutional Networks and MFCC". In: *IEEE Access* 12 (2024), pp. 124801–124815. DOI: 10.1109/ACCESS.2024.3454825.

[24] Karim Habashy, Julio Valdes, Madison Cohen-McFarlane, Pengcheng Xi, Bruce Wallace, Rafik Goubran, and Frank Knoefel. "Cough Classification Using Audio Spectrogram Transformer". In: *2022 IEEE Sensors Applications Symposium (SAS)*. IEEE, 2022. DOI: 10.1109/SAS54819.2022.9881344.

Acknowledgements

The Editors gratefully acknowledge the authors, reviewers, guest editors, conference organizers, and institutional partners whose expertise and contributions supported the development of this Special Issue. Particular appreciation is extended to the reviewers for their careful, constructive, and independent assessment of submitted manuscripts.

Publisher

CuspideScience is an independent open-access scholarly publisher based in Montréal, Québec, Canada. CuspideScience supports peer-reviewed journals committed to editorial independence, research integrity, responsible open-access publishing, and the international dissemination of scholarly knowledge.

Publisher	CuspideScience
Location	Montréal, Québec, Canada
Journal	Canadian Journal of Artificial Intelligence for Learning and Engineering Innovation
ISSN	3111-4296
Website	https://cuspidescience.com
Contact	cjailei@cuspidescience.com