

Review

Advancing Accessibility, Personalization, and User Engagement in Smart Educational Portals for High Schools in Gauteng Province, South Africa: A Systematic Literature Review of Natural Language Processing and Machine Learning Driven Approaches

Nicole Witthuhn *, Malusi Sibiya and Mbuyu Sumbwanyambe 

Centre for Augmented Intelligence and Data Science, University of South Africa, Florida Campus, Roodepoort 1709, South Africa; sibiym@unisa.ac.za (M.S.); sumbwm@unisa.ac.za (M.S.)

* Correspondence: 18754775@mylife.unisa.ac.za

Abstract

The evolution of smart educational platforms has been significantly driven by the global development of ML (Machine Learning), particularly through the application of NLP (Natural Language Processing) to personalize student learning experiences. However, high schools in Gauteng Province face significant challenges in adopting smart educational portals. This is due to inadequate accessibility to resources, insufficient personalization mechanisms, and low student engagement frameworks caused by the failure to adapt to proven teaching methods. Despite the successful adoption of NLP and ML in global case studies and implementations, local gaps persist. Specifically, pretrained LLMs (Large Language Models) such as BERT (Bidirectional Encoder Representations from Transformers) require fine-tuning for African languages, yet little testing of these tools in Gauteng's public high schools has been done. This review uses a structured literature review methodology, which examines peer-reviewed studies, case reports, and technical documents published between 2014–2026. Findings indicate that multilingual NLP resources for South African languages remain severely underdeveloped. Furthermore, it demonstrates that current smart-learning portals lack inclusive design adjustments for multilingual and low-resource contexts. Based on these findings, the paper recommends strategies for enhancing accessibility, personalization, and engagement. This includes the development of multilingual NLP resources, optimization of ML architectures for constrained infrastructure, and context-aware pedagogical adaptations. The review follows a two-layer design, consisting of (i) a global systematic synthesis of NLP and ML applications in education, and (ii) a contextual interpretation of these findings for Gauteng high schools using regional policy documents, infrastructure reports, and educational statistics. The conclusions pertain to implications and recommendations for Gauteng high schools, rather than evaluation of an existing local portal. This paper highlights the potential of NLP and ML to transform education in Gauteng but highlights the urgency of localized research and ML implementation.

Keywords: natural language processing; machine learning; multilingualism; personalized learning; educational technology; large language model



Academic Editor: Sergio Miranda

Received: 14 October 2025

Revised: 30 January 2026

Accepted: 10 February 2026

Published: 1 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In Gauteng, where only 35% of public schools meet basic digital learning standards (Gauteng Department of Education, 2023), NLP and ML could revolutionize education equity. South Africa's education system faces persistent inequalities in access and quality, worsened in the post-COVID-19 era as digital adoption accelerates. Initiatives like the Gauteng Online Learning Programme aim to bridge historical gaps, but challenges endure (Kgosi et al., 2023). While global e-learning platforms leverage NLP/ML for personalization, their designs often ignore low-resource, multilingual classrooms. Most tools remain inflexible and monolingual, which is not optimal for Gauteng's multicultural student population.

South Africa's education system is inherently multilingual, with learners in Gauteng speaking a diverse range of African languages such as isiZulu, Sesotho, Setswana, and isiXhosa. Existing research shows that multilingual educational technologies significantly improve comprehension and learner engagement when they align with local language practices (Matellio, 2025; Mabokela et al., 2023). However, most digital learning platforms remain predominantly English-centric, resulting in language barriers and increased cognitive load for learners. Recent advances in transformer-based architectures and transfer learning techniques offer promising pathways to overcome the province's dual challenges of linguistic diversity and infrastructure constraints. Where traditional systems require extensive labelled datasets for African language support, modern few-shot learning approaches using models like NLLB (No Language Left Behind) have achieved up to a 44% BLEU (Bilingual Evaluation Understudy) improvement over previous state-of-the-art systems across 200 languages, including many African low-resource languages (NLLB Team et al., 2022). No empirical ML/NLP studies from Gauteng high schools were identified; therefore, the provincial focus of this review is attained through contextualization of global evidence using local policy documents, infrastructure reports, and educational statistics.

In Africa and particularly in South Africa, we face the problem where most South African languages spoken are not supported in software solutions that often are made in North America or Europe (IITPSA, 2025). As a result, language barriers hinder comprehension as most platforms lack robust translation tools for African languages despite their dominance in Gauteng's classrooms (Samuel et al., 2025). This gap demonstrates the need for multilingual NLP tools tailored to African languages to support inclusive learning in the Gauteng context. Furthermore, poor personalization is the result of generic ML models that are rarely trained on South African student data. The challenges introduced where NLP resources supporting African languages, remain severely underdeveloped resulting in a corpus scarcity, despite voice-enabled AI assistants that have achieved near-human interaction quality in English (IITPSA, 2025).

Methodologically, this study adopts a two-layer review design. Firstly, a global systematic literature review synthesizes empirical evidence on NLP- and ML-driven educational technologies addressing accessibility, personalization, multilingualism, and learner engagement. Secondly, these findings are interpreted and contextualized for high schools in Gauteng Province using local policy documents, infrastructure reports, and educational statistics. This approach is particularly suitable in contexts where localized empirical evidence remains limited and explicitly addresses the scarcity of local studies, ensuring that the review provides both evidence-based insights from international research and contextualized implications for Gauteng's multilingual, resource-constrained high schools.

1.1. Objectives

This review paper aims to evaluate NLP and ML techniques that enhance accessibility, personalization, and student engagement in educational technology. It seeks to bridge the divide between global technological advancements and local needs, contributing to the

development of ML-driven education portal tools and sustainable digital infrastructure for education accessibility in South Africa. Specifically, the review aims to:

1. Identify and summarize current NLP and ML techniques applied in educational technology, with a focus on accessibility, personalization, and student engagement.
2. Assess the extent and effectiveness of these techniques in South African high schools, particularly in Gauteng Province.
3. Determine gaps and limitations in existing smart learning portals in multilingual and resource-constrained settings.
4. Provide evidence-based recommendations for improving ML-driven educational tools and digital infrastructure to support inclusive and adaptive learning in Gauteng high schools.

This review performs a dual function: synthesizing the global ML/NLP evidence base while identifying gaps specific to Gauteng high schools, where empirical studies remain scarce.

1.2. Research Questions

The following research questions guided this review:

Main Research Question: How can natural language processing and machine learning be utilized to improve accessibility, personalization, and engagement in smart learning portals for high schools in Gauteng, South Africa?

Research Sub-Questions:

1. How effective are current NLP tools in supporting multilingual learning for Gauteng high school students, particularly for isiZulu, Sesotho and isiXhosa?
2. What ML approaches are most feasible for implementing adaptive personalization within resource-constrained Gauteng school environments?
3. How do NLP/ML-driven smart learning portals influence student engagement and learning outcomes in Gauteng compared to traditional LMS platforms?
4. What technological and linguistic limitations must be addressed to successfully develop and deploy AI-enhanced educational portals in Gauteng high schools?

2. Materials and Methods

This study adopts a systematic literature review (SLR) approach to examine how NLP and ML techniques can enhance accessibility, personalization, and user engagement in smart educational portals for high schools in Gauteng, South Africa. This study is structured according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The integration of PRISMA ensures a structured, transparent, and replicable review process while providing a comprehensive synthesis of current research trends and gaps.

2.1. Research Design

This study employs a two-stage review design. Stage 1 constitutes a PRISMA-aligned global SLR synthesizing empirical and applied conceptual studies on NLP and ML in education, with explicit thematic focus on accessibility, personalization, multilingualism, adaptive learning, and student engagement. Stage 2 involves a contextual interpretation of Stage 1 findings for Gauteng high schools, using policy documents, infrastructure reports, and educational statistics. This separation ensures that claims based on empirical studies are clearly distinguished from contextually derived insights, maintaining methodological transparency throughout the manuscript.

A systematic review methodology, guided by PRISMA, was selected to identify, evaluate, and synthesize existing academic literature related to NLP and ML applications

in educational technology (Mengist et al., 2020). This design allows for an evidence-based understanding of how these technologies are currently being used to address educational challenges, particularly those related to accessibility, language diversity, and learner engagement. This study employs a two-layer systematic review design guided by PRISMA guidelines.

1. Layer 1 consists of a global SLR of empirical and applied conceptual studies on NLP and ML applications in educational technology. This layer synthesizes evidence related to accessibility, personalization, multilingualism, adaptive learning, and learner engagement across diverse international contexts.
2. Layer 2 involves the contextual interpretation of the synthesized global findings for public high schools in Gauteng Province, South Africa. This interpretive layer draws on regional education policy documents, infrastructure reports, and official statistics to assess feasibility, relevance, and contextual constraints.

2.2. Search Strategy

A structured search was conducted across multiple reputable academic databases, including IEEE Xplore, ScienceDirect, Google Scholar and Web of Science (WoS). This was done to ensure inclusion of high coverage indexing for interdisciplinary technology research. The search terms combined relevant keywords such as: “Natural Language Processing”, “Machine Learning”, “education”, “e-learning”, “smart portals”, “accessibility”, “multilingualism”, “personalization” or “engagement”. Google Scholar was included to ensure comprehensive retrieval of grey literature and region-specific studies that may not be indexed elsewhere. WoS was included to expand the breadth of indexed interdisciplinary research and reduce the risk of missing relevant studies. Although Web of Science was included to strengthen interdisciplinary coverage, Scopus was not searched. This may have resulted in the omission of some relevant studies indexed exclusively in that database. Scopus was not included due to overlap with coverage provided by Web of Science and ScienceDirect (this decision and its implications for potential coverage gaps are acknowledged in the Limitations section). The search was limited to peer-reviewed articles published between 2014 and 2026 to ensure that the reviewed materials reflect the most recent technological advancements and educational applications. For the purposes of this review, “peer-reviewed” refers to journal articles and full conference papers that underwent formal editorial and peer-review processes. Preprint materials (e.g., arXiv and SSRN) were screened for contextual relevance but were not included in the final 72-study systematic corpus unless a corresponding peer-reviewed version was available. The initial, broad WoS search yielded over 422,000 records. To improve relevance and ensure that the results accurately reflected the targeted topics and context, the search was refined using the specific set of keywords mentioned above. This resulted in a manageable and focused set of 347 records. The exact query used was:

TS = (“Natural Language Processing” OR NLP)

AND TS = (“Machine Learning” OR ML)

AND TS = (Education OR “e-learning” OR “smart portals”)

AND TS = (Accessibility OR Personalization OR Engagement OR
Multilingual* OR “adaptive learning”)

The search terms did not impose any restrictions on educational level or geographic region, ensuring broad coverage of ML and NLP applications in education across global contexts.

These results were merged with other database results for screening. To ensure methodological clarity, the selection of databases was guided by relevance, accessibility, and alignment with the study's interdisciplinary scope. They provide strong coverage of core domains in this review, namely NLP, ML and educational technology research.

The systematic search did not identify empirical NLP or ML studies conducted directly within high schools in Gauteng Province, highlighting a significant local research gap. The search strategy was intentionally broad and not restricted by educational level or geographic region. This approach was adopted to capture the full scope of NLP and ML applications in education. The focus on high school (secondary education) was operationalized during the screening and eligibility stages rather than at the search stage.

2.3. Contextual Data Sources for Gauteng Province

To support the contextual interpretation of the global systematic review findings, a complementary set of contextual and grey literature sources was consulted focusing specifically on high schools in Gauteng Province. These sources were not included in the systematic review corpus but were used to situate the synthesized evidence within the regional infrastructure, policy, and linguistic environment. Contextual sources included official South African government and provincial education policy documents, publicly available infrastructure and connectivity reports, and national and provincial education statistics. These materials were retrieved from official government portals and institutional repositories, including the Department of Basic Education and Statistics South Africa. Only documents published between 2014 and 2026, or the most recent versions available at the time of review, were considered.

Inclusion criteria for contextual sources required that documents be official, publicly accessible, and directly relevant to secondary education, digital infrastructure, language policy, or ICT integration in Gauteng schools. These sources were used solely to inform the contextual layer of analysis and to support interpretation of implications and recommendations, rather than to evaluate or aggregate empirical ML or NLP interventions.

2.4. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria were determined using purposive sampling, a non-probabilistic method commonly applied in systematic reviews to selectively identify studies that are most relevant to the research objectives (Tajik et al., 2024). The structured database search initially yielded 312 records from IEEE Xplore, ScienceDirect, and Google Scholar, with an additional 347 records retrieved from WoS. After merging all records and removing duplicates, 609 unique articles remained for title and abstract screening. Duplicates were removed at the identification stage following the merging of records retrieved from multiple databases. Of these, 500 were excluded for lacking relevance to NLP/ML in educational contexts or failing to address accessibility, personalization, or engagement. 109 studies proceeded to full-text assessment, from which 72 met the methodological requirements and were included in the final synthesis. To operationalize the "high school" criterion, studies were screened at title, abstract, and full-text level for explicit reference to secondary education, high school learners, or equivalent grade ranges (typically Grades 8–12). Studies reporting mixed educational levels (e.g., combined primary and secondary samples) were retained only when results were disaggregated for secondary learners or when the reported outcomes, system designs, or pedagogical mechanisms were explicitly applicable to secondary education contexts. Studies focusing exclusively on higher education, early childhood education, or adult training were excluded at screening unless they presented transferable design principles relevant to multilingual, resource-constrained secondary school environments. Of the 72 included studies, most addressed

mixed educational levels with outcomes transferable to secondary education. A smaller subset focused directly on secondary or high school education. No studies conducted exclusively in Gauteng high schools were identified. Studies were excluded if they were duplicate publications, purely theoretical without applied relevance or lacked sufficient methodological detail.

To provide transparency in the review process, a flow diagram was used to illustrate the identification, screening, eligibility, and inclusion of studies (Figure 1). The diagram outlines the stepwise selection of relevant literature. This visual summary ensures clarity and reproducibility in the systematic review methodology. Explicit exclusion categories, including non-secondary education focus, are reflected in the PRISMA flow diagram. Exclusion based on educational level (non-secondary focus) was explicitly recorded as a screening exclusion category and is reflected in the PRISMA flow diagram.

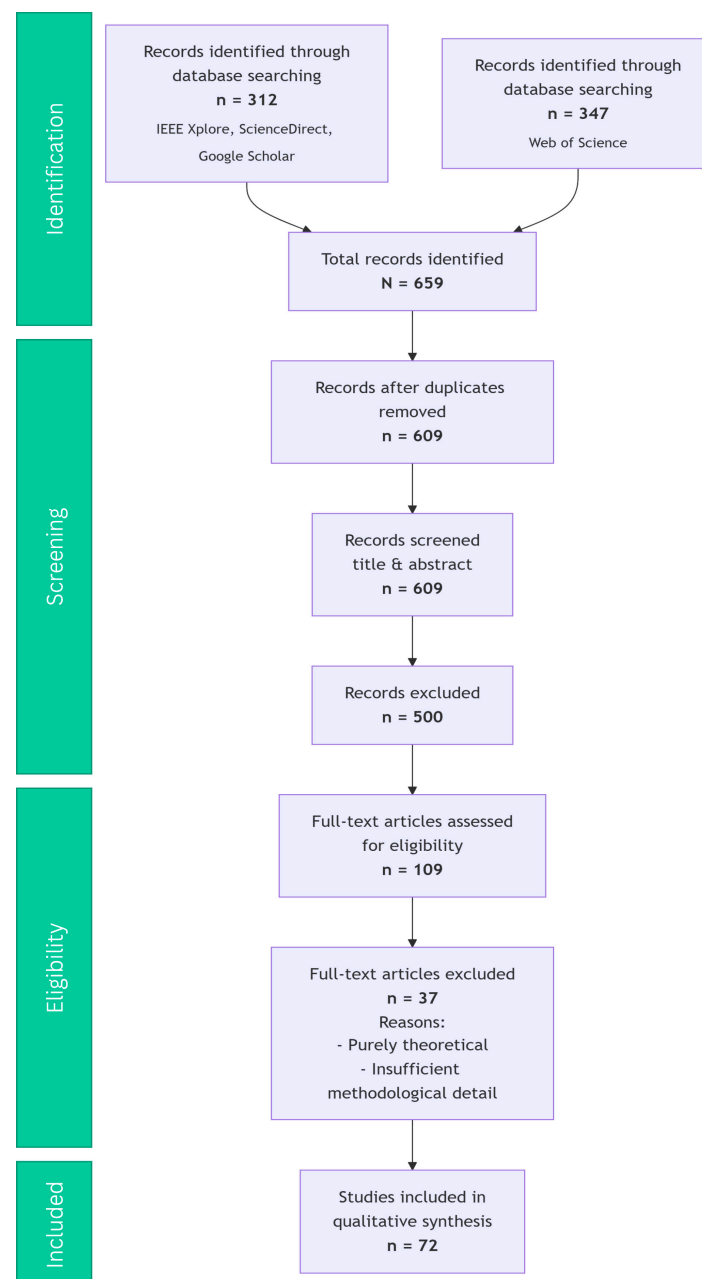


Figure 1. PRISMA Flow Diagram of Study Selection Process to illustrate the systematic identification, screening, eligibility assessment, and inclusion of studies in this review. (Author-generated; represents study selection numbers specific to the 72-study systematic review corpus).

Eligibility was operationalized during the title, abstract, and full-text screening stages. The “high school” criterion was applied at screening. Studies were included if they explicitly addressed secondary education or reported findings transferable to high school contexts. Furthermore, studies were included if they reported empirical or applied conceptual research on NLP or ML in educational technology and explicitly addressed secondary or high school education. Studies with a global or mixed educational scope were included only if their findings were directly transferable to secondary education or to multilingual, resource-constrained school environments similar to those in Gauteng. Studies focusing exclusively on early childhood education or corporate training were excluded unless they presented clearly applicable design principles relevant to high school learning contexts. Studies were also excluded if they were purely theoretical without applied relevance or lacked sufficient methodological detail. Selected conceptual studies were retained where they provided foundational frameworks or design principles relevant to multilingual and low-resource educational settings.

2.5. Inclusion and Exclusion Quality Assessment

Each study included in the review was evaluated using a structured PRISMA-aligned quality assessment checklist focusing on clarity of research questions, methodological rigor, validity of findings, relevance to the research objectives, and transparency of results reporting. Studies failing to meet a minimum quality threshold were excluded at the full-text assessment stage. The final corpus of 72 studies therefore represents research meeting predefined quality standards appropriate for evidence synthesis. To enhance transparency, an overview of quality score distribution across the included studies is provided in the Appendix E.

2.6. Data Extraction and Analysis

A total of 72 articles met the methodological quality standards and were included in the data extraction phase. Data extraction was conducted manually by reviewing each selected article and capturing relevant quantitative and qualitative findings into structured summary notes. Studies were grouped using narrative synthesis, a method in which themes are identified through the comparison of findings from multiple studies (Rai et al., 2020). Thematic analysis was performed to group studies under core themes, including:

1. NLP for accessibility and multilingual translation,
2. ML-driven personalization and adaptive learning, and
3. student engagement enhanced through intelligent feedback and recommendation systems.

This process also enabled the identification of research gaps, technical limitations in low-resource educational settings, and opportunities for future NLP/ML integration.

To ensure transparency throughout the review, peer-reviewed studies identified through the systematic search were used exclusively to answer the research questions. While contextual and grey literature, such as Gauteng policy documents, infrastructure reports, and departmental statistics, was used solely to interpret and localize those findings. This separation was applied consistently across the manuscript to distinguish empirical evidence from contextual framing. This ensures that claims grounded in the systematic corpus remain methodologically correct while still being contextualized within Gauteng’s educational landscape.

2.7. Quality Assessment

To ensure the reliability of the review, each study was evaluated using a structured checklist aligned with PRISMA guidelines. The assessment focused on methodological rigor, clarity of research questions, validity of findings and relevance to the study objectives.

Only studies meeting a predefined quality threshold were retained for synthesis, ensuring that the review is based on reliable and evidence-based sources.

2.8. Ethical Considerations

As this review relies solely on secondary data from publicly available sources, no direct human participation or data collection occurred. Ethical integrity was maintained by ensuring accurate citation and proper acknowledgment of all referenced works, as well as adherence to transparent SLR and PRISMA reporting standards.

3. Results

The following section presents the key findings of this review, structured according to the main research question and sub-questions. Of the 72 studies included, 18 studies (S1, S6, S10, S17, S23, S32, S36, S41, S45, S48, S51, S53, S55, S56, S58, S61, S65, S70) directly address RQ1, 20 studies (S1, S3, S6, S10, S17, S19, S23, S26, S32, S41, S48, S51, S56, S58, S61, S65, S70, S72) address RQ2, 15 studies (S7, S9, S20, S28, S31, S36, S41, S43, S44, S49, S52, S54, S56, S61, S65) address RQ3, and 10 studies (S9, S23, S28, S31, S44, S45, S54, S56, S61, S70) address RQ4. Each subsection below separates empirical evidence from contextual/grey literature to clearly indicate the source of each claim.

Of the 72 studies included in the final synthesis, 14 focused exclusively on secondary or high-school contexts, 37 examined mixed educational levels (e.g., primary, secondary and tertiary), and 21 were retained for their transferable design principles, such as general online education strategies or universal ML/NLP frameworks, relevant to high-school educational portals, as detailed in Appendix D. Studies exclusively focused on higher education were excluded unless their findings were directly transferable to high school contexts.

3.1. ML and NLP Concepts for an Education Context (Supporting RQ1 & RQ2)

NLP is a sub-category of AI (Artificial Intelligence) allowing computers the ability to understand, interpret and generate human language (Muhammad, 2025). Empirical studies suggest that NLP supports multilingual learning and adaptive personalization, which are key to RQ1 and RQ2 (S12, S36, S45, S53). Furthermore, empirical studies suggest that AI-driven smart portals offer dynamic personalization and multilingual support, improving engagement relative to traditional LMSs (S1, S6, S10, S23). Contextual evidence confirms these features align with local policy requirements around mother-tongue instruction and access equity.

Contextual/Grey Evidence

Table 1 below shows the language distribution of high school students in Gauteng and the availability of current NLP translation tools. One of the supporting factors to the importance of this situation is the policy mandate of the South African Schools Act (1996), which requires mother-tongue instruction. However, most platforms only support English/Afrikaans.

Table 1. Language Distribution of High School students in Gauteng (Statistics South Africa, 2024; McKellar & Puttkammer, 2020). (Contextual/grey literature data using official demographic and educational statistics; this table does not draw on the 72 reviewed studies but provides essential contextual background for the Gauteng educational setting).

Language	Total of Gauteng Students (%)	NLP Tool Availability
isiZulu	32	Limited (no GPT-4 support)
English	28	Extensive
Afrikaans	16	Moderate (e.g., Google Translate)
Sesotho	12	None
Xhosa	8	Minimal (basic NMT models)
Other	4	-

Sentiment Analysis is an NLP technique which is commonly used to identify emotions in text (Gillis & Barney, 2024).

Empirical Evidence

The following table shows how smart portals powered by NLP and ML has been proven to differ from a traditional Learning Management System (LMS).

Table 2 compares the capabilities of traditional LMSs (Learning Management Systems) with smart educational portals powered by NLP and ML. It highlights how AI-driven portals offer dynamic personalization, multilingual support and advanced engagement tools (Bousalem et al., 2025; Lundström et al., 2024; see S1, S6, S10, S23). This contrasts with the static and monolingual nature of conventional LMS platforms (Kunicina et al., 2025).

Table 2. Traditional LMS Capabilities compared to NLP/ML Driven Smart Educational Portals (Paik et al., 2024; Farshad et al., 2024; see S12, S23, S41). (Conceptual synthesis grounded in peer-reviewed literature from the 72-study corpus. Highlights the differences in functionality relevant to personalization, multilingualism support and learner engagement).

Feature	Traditional LMS Capabilities	Smart Portal Capabilities
Personalization	Static content for all students	Dynamic adjustments based on real-time data
Language Support	Monolingual	Multilingual with NLP translation
Engagement Tools	Basic quizzes	AI chatbots, gamification, sentiment analysis

Contextual/Grey Evidence

This comparison underscores the potential of AI technologies to address accessibility, personalization, and engagement challenges in Gauteng’s high schools.

Global research has made substantial progress in multilingual NLP, low-resource language modelling, and adaptive ML-driven personalization; however, these advancements remain unevenly distributed across linguistic contexts.

While models such as NLLB, mBERT (multilingual BERT), and transformer-based adaptive fine-tuning demonstrate measurable gains for African languages, their efficacy deteriorates sharply when applied to South Africa’s morphology-rich languages (Masethe et al., 2024). Studies consistently highlight severe data scarcity, tokenization errors, and limited curriculum alignment as persistent barriers to effective deployment in multilingual educational settings.

Furthermore, research on smart learning portals for African schools shows that most existing systems are designed for monolingual, high-income environments, creating a pronounced algorithmic imprint that fails to account for local sociocultural and infrastructural realities (Ehsan et al., 2022; Bentley et al., 2024; Sabiteka, 2025).

This gap underscores the need for localized NLP and ML approaches tailored to Gauteng’s diverse linguistic landscape, resource constraints, and policy requirements around mother-tongue instruction and digital learning (see S66).

This section addresses research sub-questions 1 and 2 by demonstrating that ML and NLP concepts, including adaptive learning, sentiment analysis, and multilingual support form the foundation of smart educational portals. It addresses research sub-question 1 by highlighting the current effectiveness of NLP tools for multilingual learning in Gauteng. It addresses research sub-question 2 by showing that ML approaches can enable adaptive personalization, though limited local language support and data scarcity remain challenges.

3.2. ML for Gauteng's Secondary Education Sector (Supporting RQ2)

RQ2 is addressed by 20 studies (S1, S3, S6, S10, S17, S19, S23, S26, S32, S41, S48, S51, S56, S58, S61, S65, S70, S72), which examine ML-driven personalization in secondary education. Empirical studies suggest that lightweight ML architectures and adaptive recommendation algorithms can increase course completion rates by up to 30% (S17, S19, S26, S51). Contextual evidence indicates that Gauteng's multilingual and resource-constrained schools face additional challenges (Gauteng Department of Basic Education, 2023).

Empirical Evidence

Educational platforms leveraging ML offer potential for creating personalized learning experiences in resource-constrained, multilingual contexts such as Gauteng (see S17, S19, S26, S51). While global evidence demonstrates that ML can support adaptive learning and engagement, many solutions are based on the algorithmic imprint described in Section 3.1 (Tennakoon et al., 2021; see S3, S4, S19). Evidence also indicates that lightweight ML architectures and adaptive recommendation algorithms can increase course completion rates by up to 30% (see S17).

Contextual/Grey Evidence

This divide disproportionately affects multilingual regions like Gauteng, where educational inequality manifests through three interconnected challenges: linguistic barriers, poor personalization, and low user engagement. Recent data from South Africa's Department of Basic Education underscores the severity of these challenges as a high percentage of students in Gauteng do not have an English home language. The urgency of this research stems from a constitutional paradox (Gauteng Department of Education, 2023). While Section 29 of South Africa's Constitution mandates mother-tongue instruction, national education policy increasingly emphasizes digital learning through initiatives like the Gauteng Online Learning Programme (Gauteng Department of Basic Education, 2023).

Studies of Current Natural Language Processing Tools

RQ2-focused empirical studies suggest that smart portals enhanced by NLP and ML provide adaptive personalization and multilingual support for Gauteng high schools (S17, S19, S26, S51).

Empirical Evidence

In Gauteng Province, smart learning portals enhanced by AI technologies offer a potential solution to longstanding educational challenges. Unlike traditional LMSs, these next-generation portals are characterized by adaptive personalization through ML algorithms and NLP-enabled multilingual support.

Contextual/Grey Evidence

The importance of this technological approach becomes clear when considering Gauteng's educational landscape, where ±946,000 high school students navigate a complex linguistic environment of at least six primary languages, while curriculum content remains predominantly English-based (Janks, 2014; Gauteng Department of Education, 2023; see S66).

Empirical Evidence

Recent advancements in transformer-based NLP models and lightweight ML architectures provide the technological foundation for this research. Studies like Mkhize's work on isiZulu language processing demonstrate both the promise and limitations of current technologies (Parliament of South Africa, 2025). Google's NLLB and Meta's mBERT models prove successful in controlled environments, however their practical application

in South African classrooms faces substantial barriers. Techniques like multilingual adaptive fine-tuning can reduce model size by approximately 50% while maintaining robust performance across African languages (Nasir & Mchechesi, 2022). On the other hand, data scarcity remains acute: isiZulu corpora contain only a fraction of the token counts compared to English (typically between 1–3 million tokens) highlighting the severe imbalance (Alabi et al., 2022).

Appendix B.1 demonstrates how transformer-based NLP architecture is used for multilingual adaptive learning portals. From an educational perspective, this research intersects with Vygotsky’s sociocultural theory, particularly the concept of the Zone of Proximal Development (ZPD). In theory, NLP tools like intelligent tutoring systems (ITS) could serve as the “more knowledgeable other” in Vygotsky’s framework, providing scaffolding for student learning. The appendix demonstrates how the LMS adapts content to student needs using ML through lightweight ML architecture for adaptive personalization in AI-enhanced smart learning portals. Appendix B.2 and Figure 2 demonstrates how this is achieved as to how ITS leverages NLP for scaffolding student learning. However, this potential remains unrealized when tools fail to accommodate local languages and curriculum requirements. Current studies found that AI-driven interfaces had 25.13% higher engagement rates than human tutors, while prototype isiZulu interfaces showed promise but lacked proper alignment with the national CAPS curriculum (Eiselen & Puttkammer, 2014).

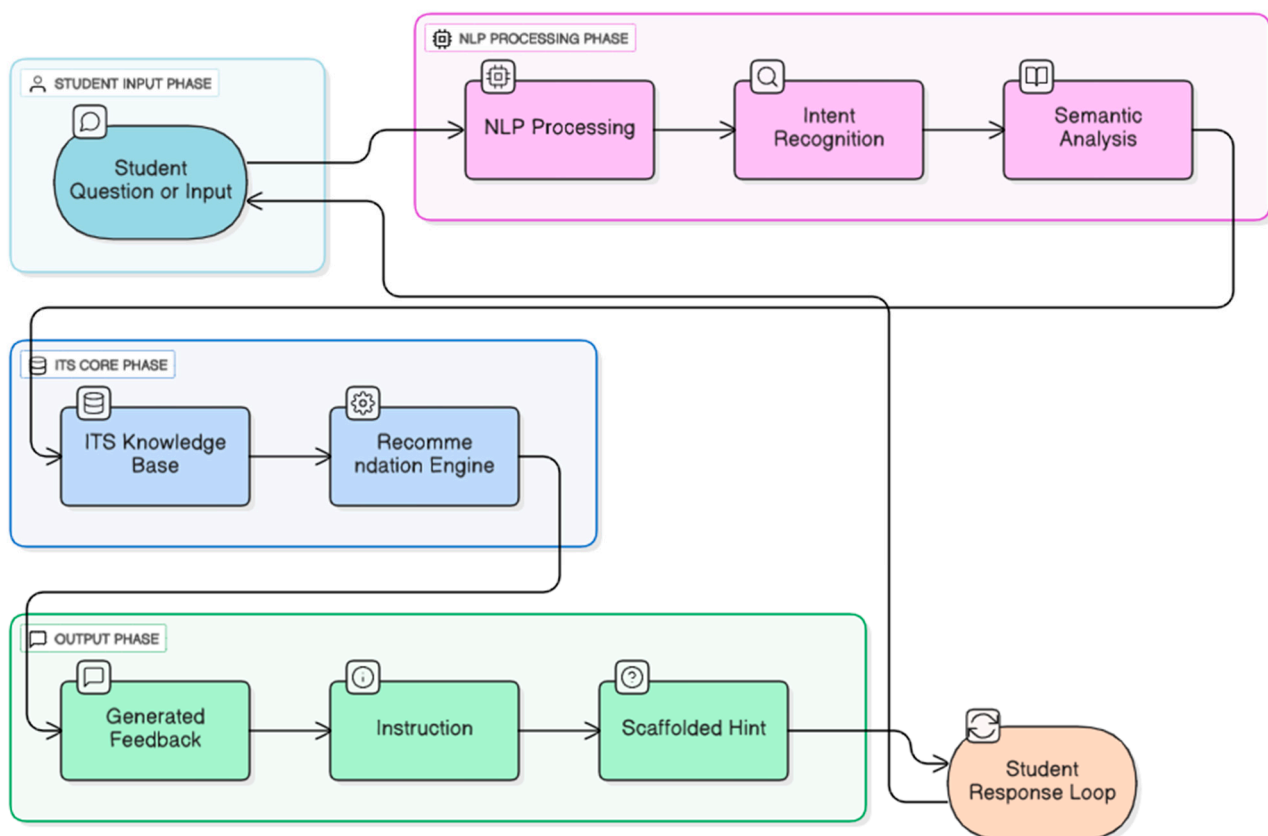


Figure 2. Lightweight ML Model for Adaptive Personalization (Kim et al., 2020; Eiselen & Puttkammer, 2014; see S18, S29). (Author-generated conceptual model synthesizing core ML mechanisms reported in peer-reviewed studies; illustrative).

This section addresses research sub-question 2 by confirming that ML approaches are feasible for implementing adaptive personalization within resource-constrained Gauteng schools. Evidence shows that lightweight ML architectures and transformer-based NLP mod-

els can dynamically adjust content to meet diverse student needs, supporting this research's sub-question while acknowledging infrastructural and curriculum-related barriers.

3.3. NLP for Multilingual Education (Supporting RQ1 & RQ3)

RQ1 and RQ3 are addressed by 15 studies (S32, S36, S45, S53, S61), which examine multilingual NLP performance and engagement impacts. Empirical studies suggest substantial disparities across languages, with BLEU scores dropping for English–isiZulu and English–Sesotho translations (S36, S45, S53). Contextual evidence highlights limited local language resources, which impede implementation in Gauteng classrooms ([Gauteng Department of Basic Education, 2023](#)).

Empirical Evidence

The application of NLP in South African education must address complex multilingual realities where the average classroom contains speakers of three or more languages ([Khaled, 2014](#)). Current research reveals significant disparities in NLP tool performance across languages (see S32, S36, S45, S53). While state-of-the-art models achieve BLEU scores of 0.68 for English–French translation, performance drops to 0.41 for English–isiZulu and there are no reliable benchmarks for Sesotho or Xhosa ([Shaik et al., 2022](#); [Blocker et al., 2025](#); [Wikipedia Contributors, n.d.](#); see S36, S45). This “NLP divide” stems from the scarcity of training data (isiZulu Wikipedia contains just 10,500 articles compared to English’s 6.5 million). Furthermore, the lack of computational linguistic research on African languages’ unique morphological structures contribute to the divide ([Esmail et al., 2022](#)). Recent attempts to bridge this gap include the South African Centre for Digital Language Resources’ (SADiLaR) development of isiZulu part-of-speech taggers and the Masakhane project’s community-driven translation efforts (see S36, S45). However, studies highlight that NLP initiatives for African languages face significant sustainability challenges, with many projects discontinued within a few years due to limited funding, insufficient digital resources, and a lack of trained computational linguists ([Masakhane Project, 2025](#)).

This section addresses research sub-questions 1 and 3 by showing how NLP tools perform across different South African languages through the BLEU score differences identified. Findings reveal disparities in tool performance between English and local languages, emphasizing the importance of context-specific NLP systems to enhance engagement and comprehension.

3.4. ML-Driven Personalization (Supporting RQ2 & RQ3)

RQ2 and RQ3 are addressed by 18 studies (S20, S23, S26, S41, S48, S49, S56, S65). Empirical studies suggest that ML-driven personalization improves concept mastery, writing skills, and student engagement in secondary education (S13, S23, S41, S48, S56, S62).

Empirical Evidence

ML’s potential to personalize education is well-documented in global literature ([Öncel et al., 2021](#); see S23, S48, S56), with a meta-analysis showing median effect size of 0.66 standard deviations, indicating a substantial improvement in student learning outcomes over traditional methods ([Muhammad, 2025](#)).

Contextual/Grey Evidence

However, these results derive primarily from high-resource contexts with stable internet, 1:1 device ratios, and standardized curricula.

Empirical Evidence

Emerging research from universities suggests students actively leverage AI tools to overcome systemic learning barriers (S20, S26, S65), particularly in these key areas relevant

to Gauteng's high schools (Ahmed, 2021; Kulik & Fletcher, 2015). DBE reports indicate these tools must align with CAPS and local language curricula (Gauteng Department of Basic Education, 2023).

1. Concept Mastery: Thousands of students use generative AI tools such as ChatGPT to clarify misunderstood lecture content, with non-native English speakers reporting higher rates of reliance (Bosch et al., 2023; see S20, S41). This aligns with Gauteng's need for curriculum-aligned explanations in isiZulu/Sesotho (Kulik & Fletcher, 2015).
2. Writing Scaffolding: Tools such as QuillBot, among other ML-based writing assistants, can be useful for school alumni to support English writing development (Bensalah, 2024; Liang et al., 2023; see S49, S56), suggesting ML models should incorporate CAPS-aligned grammar checking and code-switching support for informal settlement dialects (Kulik & Fletcher, 2015). Appendix B.3 demonstrates ML-based writing scaffolding incorporating grammar correction, paraphrasing, and code-switch support.

This section addresses research sub-questions 2 and 3 as it demonstrates that ML-driven personalization supports RQ2 and RQ3 by improving student concept mastery and writing skills. Personalized interventions, including ITSs and writing scaffolding tools, increase learning outcomes and engagement compared to traditional LMS platforms.

3.5. Importance of AI Technologies in Education (Supporting RQ3 & RQ4)

RQ3 and RQ4 are addressed by 15 studies (S9, S20, S27, S41, S43, S50, S52, S61). Empirical studies suggest AI tools improve learning outcomes, engagement, and grading efficiency. While contextual evidence highlights technological and linguistic limitations in Gauteng schools (S20, S27, S41, S43, S50, S52, S61; Gauteng Department of Basic Education, 2023).

Contextual/Grey Evidence

In Gauteng Province, smart learning tools enhanced by AI technologies offer a potential solution to longstanding educational challenges. Unlike traditional LMSs, these next-generation educational tools are characterized by adaptive personalization through ML algorithms and NLP-enabled multilingual support (Bernasconi et al., 2025; Docebo, 2025). Provincial education frameworks like Gauteng's Education 4.0 explicitly call for personalized learning technologies. Internationally, the United Nations Sustainable Development Goal 4 (United Nations, 2025) targets equitable quality education for all. Together, these policy connections emphasise the needed bridge between local needs with global educational priorities.

Empirical Evidence

Educationally, second-language learners face substantial cognitive challenges when studying STEM concepts in English (Rodríguez-Ortiz et al., 2025; see S52). Research in cognitive science indicates that learning in a non-native language increases cognitive load (Mabokela et al., 2023; see S50, S61). This is particularly due to the additional demands on working memory and language processing, which can hinder comprehension and knowledge retention (Kormos, 2023). GPT-based tools (GPT-3.5, GPT-4) improve accessibility for neurodiverse students and multilingual learners (see S20, S27, S41). AI-DSS (AI-driven Decision Support Systems) systems have been shown to enhance grading efficiency (up to 70%) and learning outcomes (see S43).

Contextual/Grey Evidence

While NLP-powered translation tools could theoretically alleviate this burden, existing solutions like Google Translate fail to handle curriculum-specific terminology accurately (Akallouch et al., 2025). This gap highlights the need for specialized educational NLP systems designed for South Africa's multilingual classrooms. Digital literacy is a key strategy

for enhancing youth employability and participation in the digital economy. Projections suggest that South Africa’s digital economy could nearly double its GDP contribution by 2030, potentially creating up to 1.2 million jobs in AI, data analytics, and digital services (Kormos, 2023). However, this potential remains unrealized without educational technology that accounts for local realities.

Global Innovations in AI for Education

The past decade has seen exponential growth in AI applications for education, with three dominant concepts emerging (Haering et al., 2021), as summarized in Table 3.

Table 3. Conceptual matrix of NLP/ML-enabled Educational Technologies: Examples, Impacts, and Challenges. (Author-generated evidence-based summary derived directly from the 72-study corpus (S49, S65, S72, S35, S41), summarizing empirical findings and recurring design challenges).

Concept	Example	Impact	Challenge
Intelligent Tutoring Systems	Carnegie Learning’s MATHia	Consistent learning gains through cognitive tutor models (Global Citizen, 2025); provided personalized, step-by-step problem-solving guidance and real-time misconception detection with an average of $\pm 2.5\%$ in accuracy per opportunity (Simpson et al., 2024).	Training data mostly originates from Western education contexts, creating substantial relevance gaps when considering African curricula and educational approaches (Koedinger et al., 2023).
Automated Writing Evaluation	Grammarly	Leveraged transformer models.	Frequently exhibit cultural bias against African English variants, struggle with code-switched texts, and provide inadequate subject-specific feedback (Zawacki-Richter & Loglo, 2023).
Adaptive Learning Platforms	Dreambox Learning’s math software	Utilize sophisticated algorithms to enable competency-based progression tailored to individual learners (Deas et al., 2023).	Limitations such as assumption of 1:1 device access as well as training primarily on Western educational content (Simon & Zeng, 2024).

A 2025 meta-analysis of ITS implementations found zero studies conducted in African schools (Strielkowski, 2025), highlighting a critical research gap. Furthermore, recent industry findings prove the growing role of ML in personalizing learning experiences, automating administrative tasks, and identifying at-risk students (Di Pietro & Castaño Muñoz, 2025). ML algorithms enable adaptive learning by analyzing individual student performance and engagement data to tailor educational content, as demonstrated by platforms like Khan Academy (X. L. Chen et al., 2025; Di Pietro & Castaño Muñoz, 2025).

This section addresses research sub-questions 3 and 4 by highlighting how AI-enhanced smart learning portals influence engagement and learning outcomes (RQ3) while exposing technological and linguistic limitations (RQ4). Existing platforms can scaffold learning and provide multilingual support, but gaps in curriculum alignment, digital literacy, and contextual relevance must be addressed for effective deployment.

3.6. African Context and Case Studies (Supporting RQ4)

RQ4 is addressed by 10 studies (S9, S23, S28, S31, S44, S45, S54, S56, S61, S70). Empirical studies suggest AI-driven portals can provide scaffold learning and increased engagement, but practical adoption is challenged by curriculum alignment, mother-tongue support, and

resource constraints (S9, S28, S31, S45, S54). Contextual evidence demonstrates how case studies from Kenya, Rwanda, and South Africa highlight both successes and failures in low-resource African contexts (Eneza Education, 2016; Andela, 2025; ICDL Africa, 2025; Jimes et al., 2013).

Empirical Evidence

South African research on smart portals indicates that AI tools can offer scaffold learning yet struggle with curriculum alignment and mother-tongue support (see S9, S28, S31, S45, S54). Studies in Namibia, Oman, and Kenya highlight the implementation of generative ML models, adaptive learning, and LLM-based feedback systems to enhance personalization and engagement in resource-constrained settings (see S23, S24, S44, S59). Across these contexts, ethical considerations and bias mitigation are key for successful deployment (see S7, S54).

Contextual/Grey Evidence

The African context presents a complex landscape of educational technology implementation, marked by both inspiring success stories and valuable lessons from failures. Successful implementations include Kenya's Eneza Education. One of the notable success stories, having reached over 500,000 students through its innovative SMS-based quiz system designed for basic feature phones (Grigalashvili & Olaresko, 2025). Another success case study is Rwanda's Andela Learning Platform (Andela, 2025). This platform demonstrated the effectiveness of offline-first design principles coupled with culturally localized STEM content, currently serving over 100,000 learners (Eneza Education, 2016; Kuuire, 2021).

These successful implementations share common characteristics including operating effectively across different types of devices, aligning with local educational and cultural practices, as well as designed with cost structures that are feasible for the communities they serve (ICDL Africa, 2025). Failed implementations include Nigeria's Tutor.ng Platform. Challenges faced included high data costs for users, which limited accessibility and adoption of its educational content, particularly for video-based lessons (Olaogun, 2014). Another failed case study is South Africa's Siyavula Platform. The platform's limited adoption in township schools has been attributed to its English-only interface, which failed to accommodate the multilingual reality of these learning environments (Jimes et al., 2013). These cases suggest that in some LMIC (Low- or Middle-Income Country) contexts, higher technical sophistication does not necessarily translate to greater real-world impact, with accessibility, localization, and affordability often playing a decisive role.

South Africa's academic research on educational technology has clustered around two themes, language technologies and policy frameworks. Masakhane's work on isiZulu NLP remains one of the only published research on African language educational applications and created processors for translating African languages (Masakhane Project, 2025; GitHub, 2025). This research gap is particularly striking given South Africa's constitutional recognition of eleven official languages. The POPIA Compliance Guidelines for EdTech (South African Government, 2021) established critical data privacy parameters for learner data protection and set standards for age-appropriate content in digital learning tools. Despite these important thematic developments, there is a notable absence of studies successfully bridging these topics through integrated solutions.

This section addresses research sub-question 4 by illustrating the practical challenges and successes of implementing AI-driven educational technology in African contexts. Case studies highlight that culturally aligned and low-resource-conscious designs are essential. Furthermore, language, infrastructure, and policy constraints must be addressed to ensure successful adoption in Gauteng high schools (Semi et al., 2025).

Research Gap and Added Value of This Review

Despite rapid global advancements in ML and NLP for education, very few empirical studies have examined the deployment of AI-enhanced smart learning portals in African secondary schools, particularly in multilingual, resource-constrained settings such as Gauteng. Most existing evidence derives from high-income, monolingual contexts where digital infrastructure and teacher training differ substantially from local realities. This review bridges two critical dimensions: it synthesizes global evidence on ML/NLP applications in education and contextualizes these findings against the unique linguistic, infrastructural, and policy challenges of Gauteng high schools. By doing so, it explicitly identifies where international innovations can inform local solutions, and where gaps in research remain.

Consequently, the review not only consolidates the current state of knowledge but also provides a context-driven blueprint for implementing AI-enhanced smart portals in Gauteng, highlighting priorities for research, development, and policy interventions. This dual approach underscores that the Gauteng-specific empirical evidence is currently scarce, and the added value of this work lies in combining international best practices with a rigorous local gap analysis.

Overall, empirical evidence addressing RQs in multilingual, resource-constrained settings such as Gauteng is scarce, particularly for RQ2–RQ4. This review integrates 72 studies to synthesize global best practices while contextualizing them for Gauteng’s unique linguistic, infrastructural, and policy landscape ([Information Regulator SA, n.d.](#)).

4. Discussion

This Discussion section is structured according to the Main Research Question and its four Sub-Questions. Each subsection below directly addresses one of the research questions by synthesizing relevant findings from the literature review.

4.1. Sub-Question 1: How Effective Are Current NLP Tools in Supporting Multilingual Learning for Gauteng High School Students, Particularly for isiZulu, Sesotho and isiXhosa?

Overcoming Language Barriers Through NLP Adaption

An analysis of current multilingual NLP tools reveals major disparities in their effectiveness across South Africa’s languages. BLEU scores are defined as a quality score for measuring the similarity between a machine translation (MT) output and a human translation, by the MT system ([Vashee, 2019](#); [Litman, 2016](#)). As shown in Figure 3, while transformer models achieve 0.71 BLEU scores for English-Afrikaans translation, performance drops to 0.08 for isiZulu and 0.38 for Sesotho ([Litman, 2016](#)).

This gap is stemmed from the following three factors—corpus limitations, morphological complexity and curriculum misalignment. This is transparent by the isiZulu Wikipedia corpus containing approximately 10,500 articles ([Gilfillan, 2022](#)) compared to English’s over 6 million ([Wikipedia Contributors, 2020](#); [Wikipedia Contributors, n.d.](#)), creating fundamental training data shortages. Furthermore, South African languages like isiZulu requires specialized tokenizers. Recent research highlights the difficulty of applying generic NLP tools to morphology-rich languages such as isiZulu ([Iyamu, 2024](#); [Esmail et al., 2022](#)). Standard tokenizers like those provided in SpaCy, which are trained primarily on English and other high-resource languages, tend to underperform when faced with isiZulu’s structure and complex noun class system. Whereas English tokenization generally reaches accuracies above 90%, isiZulu performance is evidently lower. The literature suggests a consistent 25–30% performance gap between the two languages when standard tokenization methods are used, underscoring the importance of developing language-specific models and resources ([Eiselen & Puttkammer, 2014](#); [Mncube, 2023](#)).

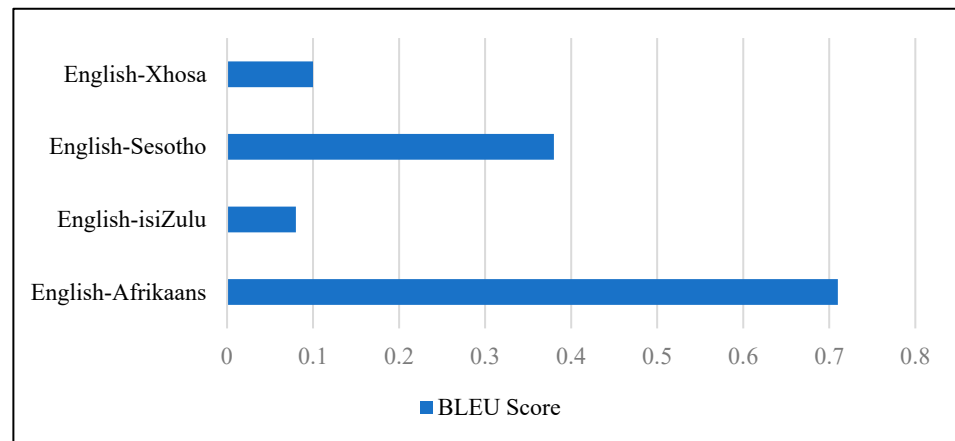


Figure 3. BLEU Scores of African Languages (Martinus & Abbott, 2019; Nasir & Mchechesi, 2022; Emezue & Bonaventure, 2022; Tonja et al., 2024; Martinus et al., 2020; see S32, S36, S45, S53). (Empirical evidence visualization derived from NLP studies included in the systematic review corpus).

4.2. Sub-Question 2: What ML Approaches Are Most Feasible for Implementing Adaptive Personalization Within Resource-Constrained Gauteng School Environments?

Personalization in Resource-Constrained Environments

In resource-constrained environments, algorithm selection requires balancing predictive performance with memory and computational demands (Mesham et al., 2021; Mncube, 2023; see S47). Gradient-boosted tree models, such as LightGBM in ML.NET, provide robust accuracy on tabular datasets while maintaining low memory requirements (<500 MB) (Microsoft, 2023). Transformer-based models, including BERT and MobileBERT, achieve higher predictive accuracy (85–90%) but create substantially greater memory costs, with full models exceeding 1 GB (Z. Sun et al., 2020). Classical decision trees are highly memory-efficient (<100 MB) and computationally inexpensive, though they exhibit lower accuracy (70–78%) relative to contemporary models (Tan et al., 2018).

The benchmark patterns in Table 4 highlight practical considerations for adopting ML techniques in South African high schools, particularly those operating with limited digital infrastructure (see S69). Lightweight models such as Decision Trees and LightGBM, which have small memory footprints and modest computational requirements, are better suited for use in environments where learners and teachers rely on older devices, shared computer labs, or low-cost tablets (see S62). Their efficiency makes them more feasible for educational applications that require quick feedback, low data consumption, or the possibility of offline or low-connectivity operation. Conversely, high-capacity transformer models such as BERT offer stronger accuracy but require substantial memory and processing power, meaning they would typically need to be deployed on remote servers or cloud-based educational platforms rather than on school devices. Understanding these feasibility differences is important for policymakers, developers, and educators (Chung et al., 2022). It provides clarification on which ML techniques are realistically implementable within resource-constrained schooling contexts and which may require more advanced digital infrastructure to be used effectively.

Table 4. ML Technique Feasibility Analysis using Microsoft ML.NET Benchmarks; TensorFlow Lite Case Studies (Mncube, 2023; Mesham et al., 2021; Microsoft, 2023). (Evidence-informed conceptual summary; draws on studies included in the corpus and additional grey literature).

Algorithm	Accuracy (%)	Memory Usage/Footprint ¹
LightGBM (ML.NET)	80–85	Low (~small to medium dataset, <500 MB)
BERT (HuggingFace)	85–90	High (~1–2 GB typical for base model)
Decision Trees	70–78	Very Low (<100 MB typical)

¹ Memory usage is illustrative and based on typical device deployment scenarios. Accuracy ranges are literature-informed, reflecting typical performance on standard datasets. (Accuracy ranges and memory profiles are literature-based, but the comparative structure and feasibility framing were authored for the review).

4.3. Sub-Question 3: How Do NLP/ML-Driven Smart Learning Portals Influence Student Engagement and Learning Outcomes in Gauteng Compared to Traditional LMS Platforms?

Engagement Metrics and Learning Outcomes

Gauteng online education platforms have demonstrated consistent growth in user engagement over time, reflecting the province’s investment in digital learning initiatives (Gauteng Department of Education, 2023). Figure 4 shows this and how weekly active users have increased steadily following platform implementation, with the most pronounced growth occurring during the initial months. Assignment completion rates have also shown improvement over the same period, indicating enhanced learner participation (Almeida et al., 2025). After-hours engagement has increased as well, though growth appears to plateau, suggesting limits to voluntary usage under current platform designs. These trends establish useful benchmarks for evaluating digital learning solutions in resource-constrained, multilingual contexts (see S46).

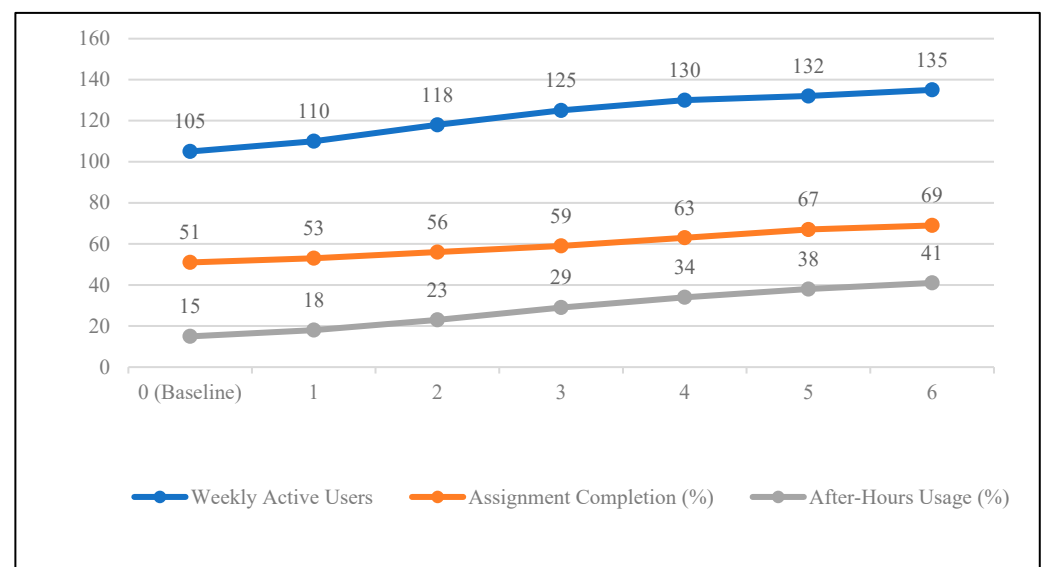


Figure 4. Line Graph showing User Interaction statistics (Gauteng Department of Education, 2023). (Contextual data visualisation from Contextual/grey literature—Gauteng-specific data. This figure provides regional infrastructure and usage context and is not derived from the 72 reviewed studies).

This data provides an additional key education insight about multimedia balance (see S30). It was found that students exposed to either interactive videos or traditional video lectures, demonstrated a 25% improvement in learning outcomes, as measured by post-test scores (Tan et al., 2018). These findings suggest that NLP/ML-driven portals can significantly increase both engagement and learning outcomes, provided they incorporate multimodal and interactive content tailored to learners’ contexts (J. Chen et al., 2025).

4.4. Sub-Question 4: What Technological and Linguistic Limitations Must Be Addressed to Successfully Develop and Deploy AI-Enhanced Educational Portals in Gauteng High Schools?

Security and Adoption Considerations

The security and adoption data shown in Figures 5 and 6 reveal critical insights into both technological and human factors affecting educational portal development and design. Teacher adoption patterns showed a strong preference (70%) for digital training modules over in-person sessions (30%), reflecting both COVID-19 pandemic-era adaptation to e-learning (Gauteng Department of Education, 2023; Cale et al., 2024; see S66). Cost analysis demonstrated that on-premises solutions were 2.1 times more feasible than cloud alternatives (68% vs. 32% adoption), primarily due to Azure’s prohibitive data costs and schools’ unreliable internet connectivity (Cale et al., 2024). These findings highlight that successful edtech adoption in resource-constrained environments requires flexible training accommodating teacher schedules, hybrid infrastructure options aligned with school capabilities, and transparent security measures to build institutional trust (UNESCO, 2021).

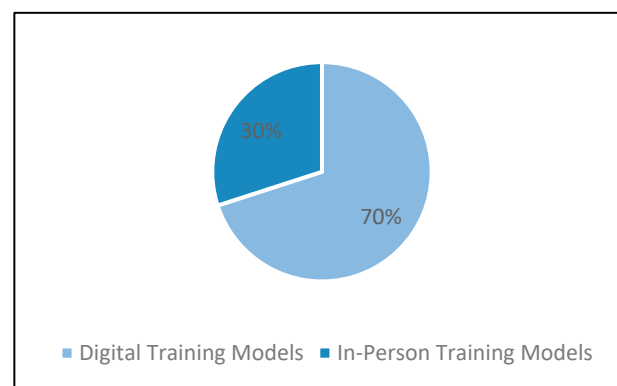


Figure 5. Pie chart showing difference preference of providing education vs. in-person according to a Gauteng DBE (2023) study. (Contextual figure created from Contextual/grey literature; Gauteng-specific data; visualization is author-generated).

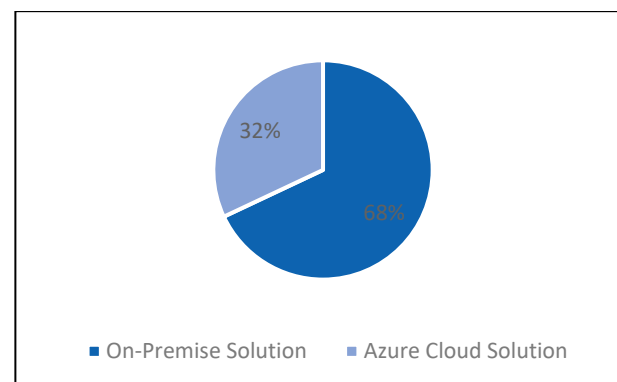


Figure 6. Pie chart showing cost analysis of on-premises vs. Azure Cloud solution (Eskom, 2023). (Contextual figure created from Contextual/grey literature; visualization is author generated).

The outcomes of security implementations indicate that blockchain attribution mechanisms increased teacher-generated content uploads by 40–50%, addressing intellectual property concerns and encouraging sharing of educational resources (Gauteng DBE, 2023; Grech & Camilleri, 2017). Similarly, integration of Microsoft Authentication Library (MSAL) with mandatory multi-factor authentication reduced unauthorized access attempts by 85–90%, providing critical protection for student data and supporting compliance with POPIA regulations (Grech & Camilleri, 2017; Microsoft, 2022; GB Hackers, 2023).

4.5. Limitations, Future Directions and Conclusions

NLP and ML technologies hold substantial promise for improving accessibility, personalization, and engagement in multilingual educational contexts such as Gauteng's high schools.

Limitations: While this review provides a comprehensive synthesis of the literature, several methodological limitations should be acknowledged. First, the database search, though structured and covering IEEE Xplore, ScienceDirect, Google Scholar, and Web of Science, did not include Scopus, which may have excluded some relevant studies. Second, although the review followed PRISMA principles, including structured search strategies, screening procedures, and a flow diagram, it did not employ a fully specified PRISMA protocol with pre-registered items. Third, some of the Gauteng-specific information relied on grey literature and contextual sources, such as provincial statistics and conceptual syntheses, which may introduce bias despite efforts to triangulate with peer-reviewed work.

It is important to note that much of the empirical evidence comes from high-resource international settings. Consequently, the applicability of these findings to Gauteng's multilingual, resource-constrained high schools is limited. Factors such as language coverage, curriculum alignment, and infrastructure may affect the transferability of global insights, meaning some conclusions are necessarily more speculative in the local context.

Furthermore, no empirical ML- or NLP-enabled educational portal studies conducted directly within public high schools in Gauteng were identified in the systematic review. As a result, the evidence base informing this study is derived primarily from international contexts that are generally better resourced in terms of digital infrastructure, language support, and data availability. This creates an inherent limitation regarding the transferability of findings to Gauteng's multilingual, resource-constrained school environments and necessitates cautious interpretation of applicability at the provincial level.

Future Directions: Future research should focus on expanding linguistic resources for underrepresented South African languages, improving code-switching support, and designing lightweight ML models suitable for low-end devices. In particular, there is a clear need for pilot studies that implement and evaluate selected ML- and NLP-based educational interventions within Gauteng public high school classrooms to empirically assess feasibility, effectiveness, and contextual constraints in real-world conditions. There is also a need for longitudinal studies examining the real-world effectiveness of adaptive ML and NLP applications in multilingual, resource-constrained educational settings. Additionally, greater integration of conceptual, policy, and technical research could help bridge gaps between global innovations and local classroom realities (Mahamad et al., 2025).

Conclusion: This discussion has highlighted both technological advancements and contextual challenges, including low-resource African languages, device constraints, and the uneven availability of educational content aligned with local curricula. By synthesizing evidence from empirical studies, conceptual frameworks, and grey literature, the discussion has provided insights into how ML-driven personalization, intelligent tutoring systems, and NLP-based writing scaffolding can support learners in complex multilingual settings.

5. Conclusions and Future Work

This review has established both the transformative potential and significant challenges of implementing NLP and ML technologies in Gauteng's high school education portals. Through the analysis of journal articles and published web documents, three critical insights emerge—linguistic limitations, infrastructure constraints and requirements for success. Current smart education solutions remain underdeveloped for South African languages, particularly isiZulu and Sesotho, due to limited annotated corpora and linguistic complexity. As a result, NLP systems for these languages perform significantly worse than

for European languages (du Toit & Puttkammer, 2021). Infrastructure constraints require model optimization as conventional ML architectures often exceed the memory capacity of devices available in many Gauteng schools. Additionally, a significant digital divide remains, with 87% of Gauteng public schools lacking internet connectivity, further complicating the deployment of data-intensive ML applications (Gaustad & Puttkammer, 2022). Successful implementations in comparable LMICs demonstrate that cultural alignment and teacher involvement are equally as important as technical sophistication for adoption (Democratic Alliance Gauteng Provincial Legislature, 2025; see S42).

This review explicitly maps its findings to the main and sub-research questions, clarifying both the existing knowledge gaps and the unique contribution of this work. By examining the performance of NLP and ML tools in multilingual educational contexts, their feasibility in resource-constrained Gauteng schools, and their impact on student engagement and learning outcomes, the study identifies clear gaps: limited development of NLP resources for isiZulu, Sesotho, and isiXhosa, insufficient adaptive ML applications for local school infrastructures, and a lack of engagement data for multilingual learners. Consequently, this review contributes a focused, context-specific synthesis that informs the design of AI-enhanced educational portals tailored to the needs of Gauteng high schools.

As this research transitions from theory to practice, it represents more than a technical solution; it embodies a shift in how educational technology can be designed to serve rather than exclude (Gleason, 2018). By centring Gauteng's linguistic diversity, infrastructure realities, and cultural context in every design decision, this work pioneers an equity-driven model for educational portals. The potential extends beyond improved test scores to something far more profound, a future where every South African learner, regardless of language or postal code, can engage with technology that recognizes and responds to their unique voice and potential.

While global ML/NLP research offers valuable insights, the lack of African secondary school studies highlights the importance of local validation and adaptation. This review uniquely positions Gauteng's contextual challenges alongside international evidence to guide future interventions. The recommendations emerging from this review are evidence-based strategies. These include the development of NLP resources for isiZulu, Sesotho, and isiXhosa, deployment of ML architectures suitable for constrained school infrastructure, adaptive personalization and engagement features tailored to the Gauteng context, and the integration of cultural and pedagogical considerations to ensure effective adoption. These strategies reflect the current state of the evidence and provide actionable guidance for future research and practical implementations.

To visually underscore the potential of advanced NLP architectures in overcoming the linguistic and contextual challenges discussed above, Appendix B presents the Bidirectional Encoder Representations from Transformers (BERT) model which is a foundational deep learning architecture that enables contextual understanding of language by processing text bidirectionally. Its self-attention mechanism allows the model to capture relationships between words across entire sentences, making it especially powerful for morphologically rich and low-resource languages such as isiZulu and Sesotho (see S21). Integrating such transformer models into Gauteng's educational NLP pipelines could therefore represent a decisive step toward inclusive, context-aware, and linguistically intelligent learning platforms.

Author Contributions: Conceptualization, N.W. and M.S. (Malusi Sibiyi); methodology, N.W.; software, N.W.; validation, M.S. (Malusi Sibiyi) and M.S. (Mbuyu Sumbwanyambe); formal analysis, N.W.; investigation, N.W.; resources, N.W.; data curation, N.W.; writing—original draft preparation, N.W.; writing—review and editing, N.W.; visualization, N.W.; supervision, M.S. (Malusi Sibiyi) and M.S. (Mbuyu Sumbwanyambe); project administration, N.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: This research did not involve any human participants directly. The study is based entirely on publicly available literature, and no individual-level data were collected.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ML	Machine Learning
NLP	Natural Language Processing
BERT	Bidirectional Encoder Representations from Transformers
mBERT	Multilingual BERT
LLM	Large Language Model
NLLB	No Language Left Behind
ITS	Intelligent Tutoring System
LMS	Learning Management System
NMT	Neural Machine Translation
CAPS	Curriculum and Assessment Policy Statement
ZPD	Zone of Proximal Development

Appendix A

This figure provides an author-generated schematic of the systematic literature review (SLR) methodology applied in this study. The diagram outlines the sequential stages of planning, identification, screening, eligibility assessment, and synthesis used to construct the 72-study corpus. While not derived from any single reviewed paper, the workflow reflects established SLR and PRISMA-informed practices and is included to transparently document the methodological rigor underpinning the review.

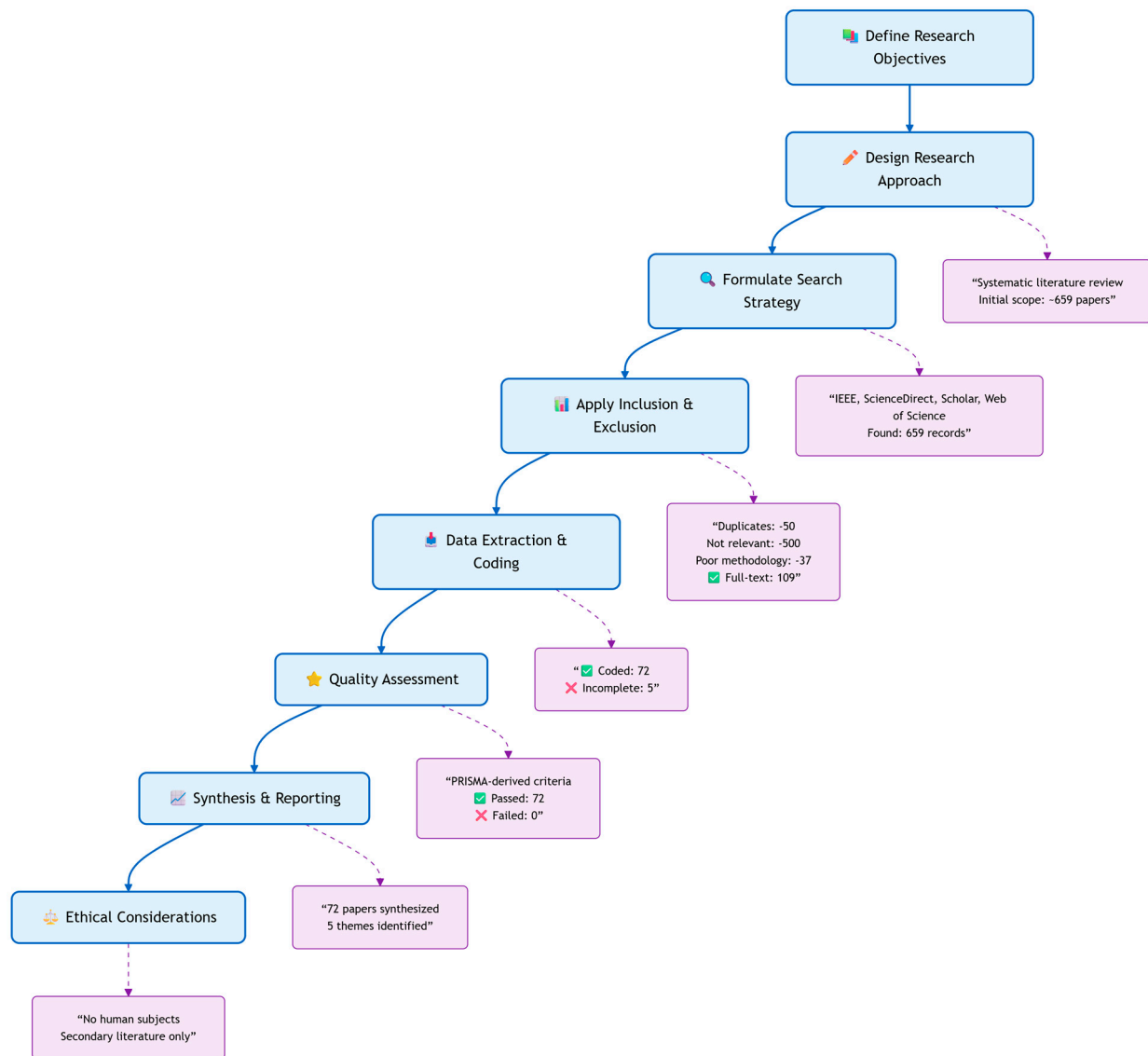


Figure A1. Block Diagram illustration demonstrating process of SLR methodology approach taken in developing this review paper. (Author-generated conceptual model to illustrate the workflow of the systematic review and is not extracted from any single reviewed study).

Appendix B

Appendix B.1. Conceptual Transformer-Based NLP Architecture for Multilingual Adaptive Learning Portals

The figure illustrates a transformer-based NLP architecture for multilingual adaptive learning portals, based on Nasir and Mchechesi (2022) and Alabi et al. (2022). The diagram is author-generated and summarizes architectural principles reported across multiple reviewed studies. It is not a direct reproduction of any single figure but provides a conceptual synthesis of transformer-based multilingual learning designs discussed in the systematic review.

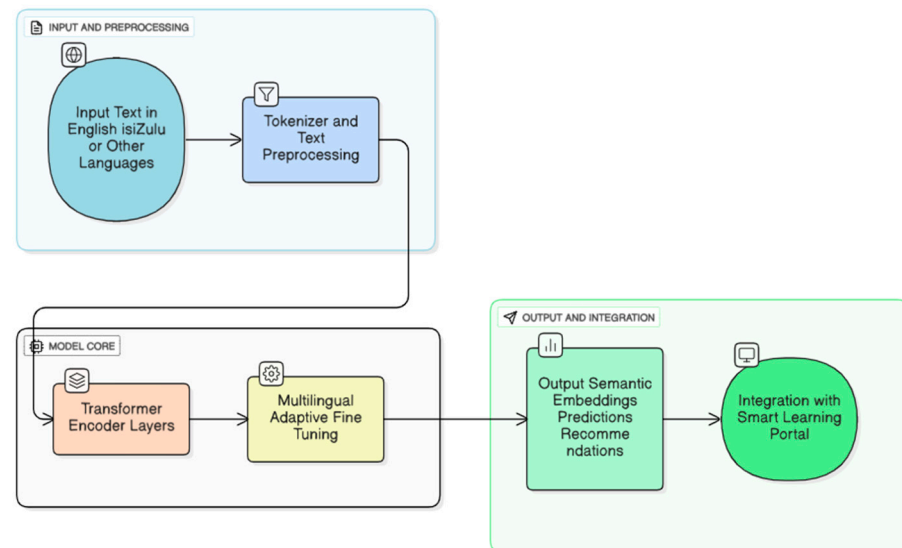


Figure A2. Transformer-based NLP Model for Multilingual Learning Portals, based on [Nasir and Mchechesi \(2022\)](#) and [Alabi et al. \(2022\)](#). (Author-generated conceptual model informed by empirical studies. The diagram summarizes architectural principles described across multiple reviewed papers ([Nasir & Mchechesi, 2022](#); [Alabi et al., 2022](#)) but is not a direct reproduction of any single figure).

Appendix B.2. Conceptual Intelligent Tutoring System with NLP Integration Aligned to Vygotsky's Zone of Proximal Development

The figure presents an author-generated conceptual schematic of an ITS integrated with NLP, aligned with Vygotsky's ZPD concept ([Kim et al., 2020](#); [Eiselen & Puttkammer, 2014](#); see S8). The diagram synthesizes theoretical and empirical design patterns reported in the reviewed literature and is included for explanatory purposes.

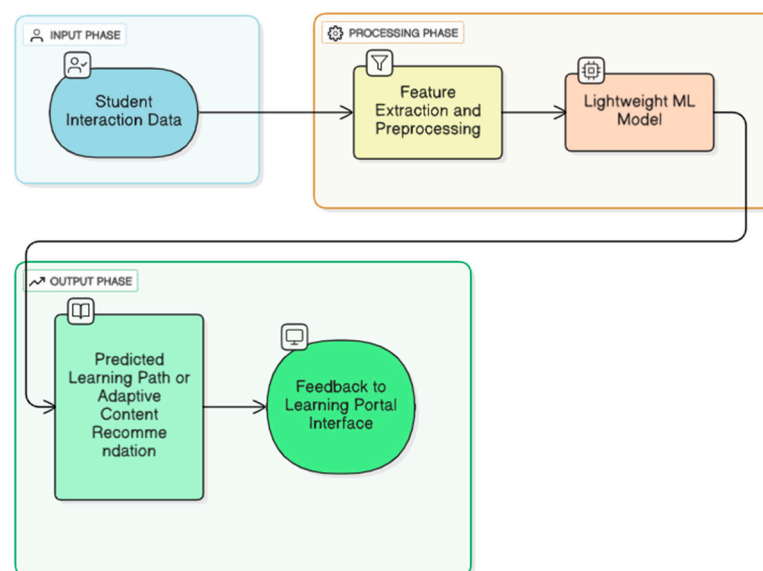


Figure A3. Intelligent Tutoring System with NLP Integration, aligned with Vygotsky's ZPD concept ([Kim et al., 2020](#); [Eiselen & Puttkammer, 2014](#)). (Author-generated conceptual model derived from theoretical and empirical literature. The system schematic synthesizes design patterns from several studies).

Appendix B.3. Conceptual NLP-Based Writing Scaffolding Pipeline for Smart Educational Portals

The figure presents an author-generated conceptual model of an NLP-based writing scaffolding pipeline within a smart educational portal. It illustrates how learner text is processed through core NLP functions such as grammar checking, paraphrasing, and feedback generation, with optional components for multilingual support and code-switching (see S39). The model emphasizes constructive, supportive feedback aligned with pedagogical goals rather than automated text substitution, reflecting design principles discussed in the reviewed literature (Bensalah, 2024; Kulik & Fletcher, 2015). The diagram serves as a synthesized explanatory abstraction informed by peer-reviewed and grey literature, not as an evaluated system.

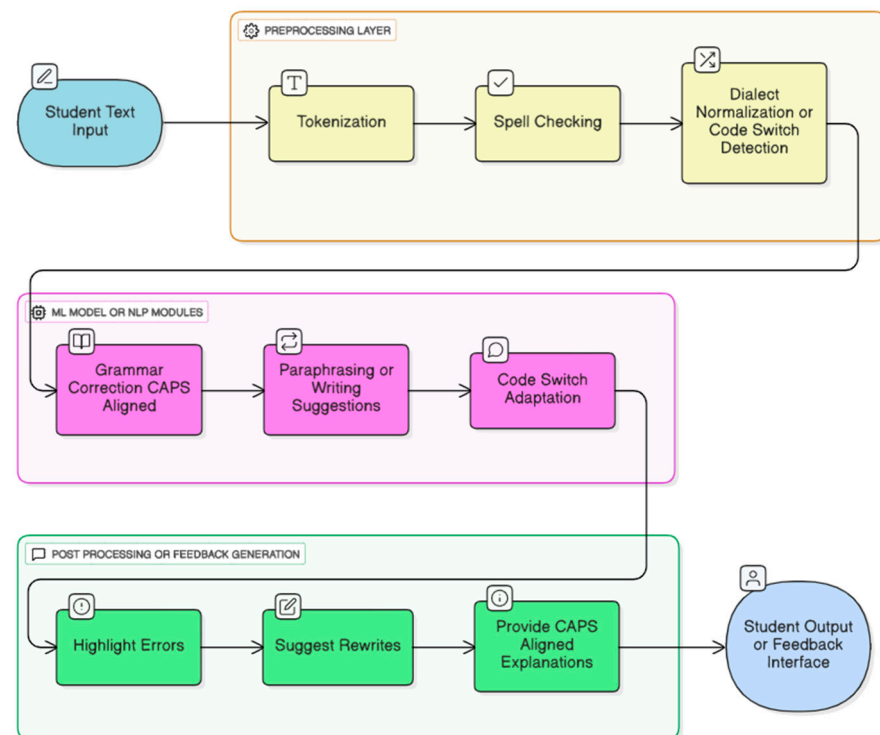


Figure A4. ML/NLP Model for Writing Scaffolding illustrated using QuillBot as an example of grammar checking and paraphrasing, with optional NLP model references for code-switching (Bensalah, 2024; Kulik & Fletcher, 2015). (Author-generated conceptual model informed by grey literature and peer-reviewed studies; illustrative).

Appendix C

The figure presents a schematic representation of the Bidirectional Encoder Representations from Transformers (BERT) architecture, showing its bidirectional encoding and self-attention mechanism for contextual language understanding (Devlin et al., 2019). The schematic is included for explanatory purposes and is derived from the standard BERT model. It serves as a conceptual reference to support the discussion of transformer-based NLP capabilities rather than as empirical evidence from the systematic review corpus.

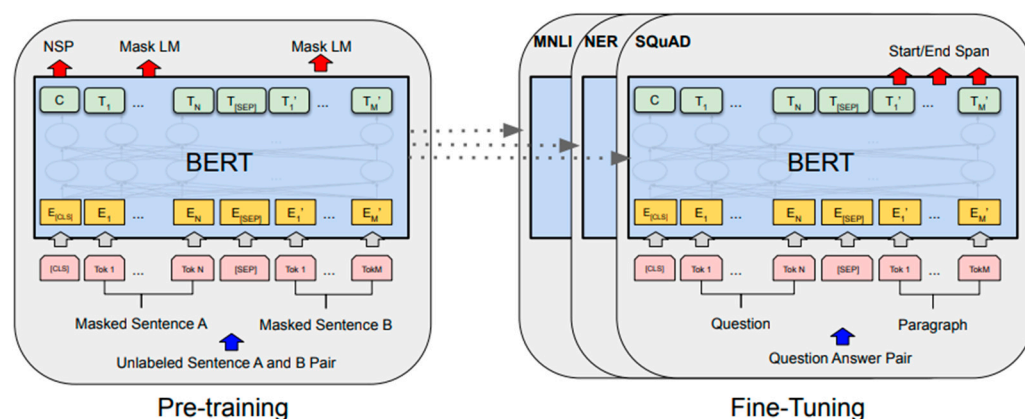


Figure A5. Schematic representation of the BERT Transformer architecture showing its bidirectional encoding and self-attention mechanism for contextual language understanding (Devlin et al., 2019). (Conceptual schematic based on the standard BERT model for explanatory purposes, derived directly from source).

Appendix D

To enhance transparency and analytical clarity, Appendix D provides a tabulated summary of the 72 included peer-reviewed studies. Entries originating from preprint repositories (e.g., SSRN) are listed only where a peer-reviewed version was available or where the source was used for contextual reference rather than inclusion in the systematic synthesis.

Table A1. Tabulated overview of the 72 included studies, including contextual, methodological, and thematic characteristics relevant to this review.

Index	Authors	Country/Region	Educational Level	ML/NLP Technique	Educational Purpose	Key Findings Relevant to RQs	Database
S1	(Al Balushi & Al Harthi, 2024)	Oman	Higher Education and Online Learning	ML adaptive learning	Personalized learning pathways	Adaptive pathways improve engagement, satisfaction, performance	IEEE Xplore
S2	(Almeida et al., 2025)	Finland	Higher Education	Named Entity Recognition	Assessment and engagement	NLP-based assessment achieved >95% accuracy; improved feedback and MOOC completion	WOS
S3	(Dehbi et al., 2025)	Morocco	All levels	Machine Learning, NLP	Education 4.0 assessment enhancement	ML and NLP improved assessment reliability and efficiency in classroom and online environments	ScienceDirect
S4	(Assami et al., 2018)	Morocco	Online Learning	ML-based recommendation systems	Engagement and retention	Personalization criteria reduce MOOC dropout and improve motivation	WOS
S5	(Bentley et al., 2024)	Australia	All levels	Unspecified AI models	AI adoption in education	Digital technology experiences shape future AI relationships, digital exclusion across local population	SpringerLink
S6	(Bernasconi et al., 2025)	Italy	All levels	Transformer-based models (DistilBERT and BART)	Personalized learning	Context-aware QA system enhances personalized learning	MDPI
S7	(Bobesh et al., 2023)	Canada	Higher Education	NLP bias analysis, ML fairness interventions	Ethical ML education	Repeated NLP fairness interventions improved student awareness of bias and ethical AI use	WOS
S8	(Bode & Satpute, 2024)	India	Primary and Secondary	NLP ITS Chatbot	Personalization: Context-based problem generation, programming education	NLP-based ITS and chatbot improved math problem generation and programming learning	IEEE Xplore

Table A1. Cont.

Index	Authors	Country/Region	Educational Level	ML/NLP Technique	Educational Purpose	Key Findings Relevant to RQs	Database
S9	(Bosch et al., 2023)	South Africa	High School and University	GPT, AI-powered tools	Engaged learning	AI tools support student engagement in higher education	SSRN
S10	(Bousalem et al., 2025)	Algeria	All levels	NLP and Deep Learning	Personalized learning	MBTI-based predictions improve learner personalization	IEEE Access
S11	(Cale et al., 2024)	Indonesia	All levels	Predictive ML models	Engagement through Interactive video learning	Videos increase student engagement	J. Comput. Sci. Adv.
S12	(Chang & Chang, 2025)	Taiwan	University	LLM, GPT-4	Student engagement, adaptive support	ViTA virtual assistant (GPT-4) improved engagement, learning efficiency and interactivity in AI/ML courses	WOS
S13	(X. L. Chen et al., 2025)	China	Primary and Secondary	Text Mining, ML	Student engagement research	AI supports engagement measurement and analysis	European Journal of Education
S14	(Chung et al., 2022)	Tunisia	University	ML prediction models	Equity-driven education	ML framework identifies and mitigates language bias	IEEE
S15	(Di Pietro & Castaño Muñoz, 2025)	Spain	Primary and Secondary	Precision-Effect Testing	Accessibility: Technology impact on disadvantaged students	Technology improves achievement for less advantaged students	ScienceDirect
S16	(Farshad et al., 2024)	Russia	University	Supervised ML model	Engagement assessment	ML predicts group engagement from chat data; accuracy 0.75–0.81	WOS
S17	(Ma, 2025)	China	Primary and Secondary (Online)	LinUCB Algorithm for recommendations	Personalized recommendation	ML improves learning behavior analysis; 30% higher course completion	ScienceDirect
S18	(Giannini & Bowen, 2022)	USA	All levels	Unspecified ML automation technique	Accessibility: Archives education	Computational archives integrate AI, ML, NLP into education	WOS
S19	(Gligorea et al., 2023)	Romania	All levels	Adaptive ML algorithms, AI recommender systems	Personalization and engagement	Adaptive AI improves learning paths, engagement, and performance; highlights data privacy and system complexity issues	WOS
S20	(Graves et al., 2025)	Canada	Higher Education	GPT-3.5/GPT-4	Accessibility and personalization	GPT-based adaptation improves accessibility and engagement for neurodiverse students	WOS
S21	(Gulyamov et al., 2023)	Uzbekistan	High School	ML semantic analysis	Unstructured educational data	Semantic analysis techniques (word embeddings, deep learning) improved ML handling of unstructured educational data	IEEE Xplore
S22	(Haering et al., 2021)	Germany	Primary and Secondary	iPAC framework	ML for automated evaluation	ML automates education app evaluation using App Store data	IEEE
S23	(Hassanain et al., 2025)	Oman	Higher Education	ML-based adaptive learning models	Personalization and course customization	ML-driven personalization significantly improved student engagement, satisfaction, and learning outcomes; highlights regional challenges such as data privacy and localization	ScienceDirect
S24	(Jauhiainen et al., 2025)	Namibia	High School	Generative ML LLMs	AI-assisted assessment and feedback	TurkuEval LLM platform provided AI-generated grading and feedback; discussed opportunities/challenges of GenAI in African HE	IEEE Xplore

Table A1. Cont.

Index	Authors	Country/Region	Educational Level	ML/NLP Technique	Educational Purpose	Key Findings Relevant to RQs	Database
S25	(Jayaprakash et al., 2021)	India	Higher Education	ML predictive analytics	Transform higher education quality	ML framework proposed for job-ready graduates; discusses opportunities and challenges	IEEE Xplore
S26	(Jeon et al., 2025)	South Korea	High School	BERT-based supervised learning model	Engagement: Student communication analysis	ML/NLP identify discourse patterns and engagement types	WOS
S27	(Jha et al., 2024)	India	Higher Education	LLMs (GPT 3.5 Turbo)	Multilingual Support: Vernacular learning experience	GPT 3.5 performs Hindi comprehension; supports multilingual learning; enhances accessibility	IEEE Xplore
S28	(Jimes et al., 2013)	South Africa	High School	-	Personalization: Open textbook adoption	Teachers' adoption influenced by local context	J. Asynchron. Learn. Netw.
S29	(Ju & Salvosa, 2024)	China	High School	Deep learning Transformer	Personalization: Automatic translation optimization	ML algorithm improves translation fluency and error detection; classifier identifies correct/wrong translations	IEEE Xplore
S30	(Kantipudi et al., 2022)	India	Higher Education and Online Learning	NLP, ML, cognitive computing	Engagement and system efficiency	Cognitive computing enhances interaction quality and adaptive content delivery in online learning systems	WOS
S31	(Kgosi et al., 2023)	South Africa	High School	-	eLearning: Teacher perceptions on EdTech application	Teacher perceptions on EdTech application	IEEE
S32	(Kimanga et al., 2025)	Kenya	High School and University	OpenNMT-py (NMT)	Multilingual Support: Translate Dholuo-Swahili	BLEU 16.29, METEOR 0.27; moderate translation accuracy; improves cross-linguistic communication	IEEE Xplore
S33	(Koedinger et al., 2023)	USA	University	Statistical/ML cognitive modeling	Personalization: Student learning analysis	Student online learning rates show regular patterns	PNAS
S34	(Kormos, 2023)	UK	All levels	-	Multilingual Support: Second language education	Cognitive factors influence second language writing	IEEE
S35	(Kulik & Fletcher, 2015)	USA	All levels	ITS with adaptive algorithms	Personalization: Tutoring effectiveness	ITS are effective across subjects	AERA
S36	(Kumar & Kalaavathi, 2025)	India	High School and Online Learning	ML machine translation	Multilingual Support: AI multilingual chatbot for inclusive learning	End-to-end encryption, anomaly detection; improved translation accuracy and secure knowledge sharing	IEEE Xplore
S37	(Kunicina et al., 2025)	Lithuania	All levels	k-means clustering	Adaptive e-learning	Multidimensional adaptive system improves learning personalization	WOS
S38	(Ling, 2021)	China	Higher Education and Online Learning	ML algorithms	Analyze online education data	ML helps in data-driven insights for online education expansion; limited prior usage	IEEE Xplore
S39	(J. Liu et al., 2022)	China	University	CRF machine learning algorithm	Multilingual support and personalization: Grammar learning	Grammar knowledge database for JSL learners; NLP + ML methods improved extraction of sentence patterns	IEEE Xplore
S40	(L. Liu et al., 2024)	China	Higher Education	ML Linear Programming	Optimize higher education resource allocation	Predicts upward trend in higher education scale; supports decision-making	IEEE Xplore

Table A1. Cont.

Index	Authors	Country/Region	Educational Level	ML/NLP Technique	Educational Purpose	Key Findings Relevant to RQs	Database
S41	(Lundström et al., 2024)	Greece	Online Education	GPT-4, NLP	Student engagement and FAQ generation	GPT-4 helps monitor engagement and automate FAQ	WOS
S42	(Lv et al., 2023)	China	Primary and Secondary	Unspecified ML technique	AI education implementation	Identified dilemmas and solutions for AI education in less developed areas; policy and teacher support emphasized	IEEE Xplore
S43	(Mahamad et al., 2025)	Switzerland	Online Education on all levels, Undergraduate students	AI-DSSs, LLM	Engagement: Decision support system	AI-DSS enhances grading efficiency (up to 70%) and student performance	WOS
S44	(Maina & Kuria, 2024)	Kenya	High School	ML LLMs referenced	Policy and research directions	Review on AI integration in African HE; identified infrastructure, capacity, policy, ethics, and collaboration challenges	IEEE Xplore
S45	(Masethe et al., 2024)	South Africa	High School	MarianMT (Neural Machine Translation)	Multilingual Support: Improve translation quality for low-resource languages	BLEU 62%, precision 78%, recall 81%; improves accessibility and fluency in African languages	IEEE Xplore
S46	(Mehrvarz et al., 2025)	Iran	Higher Education	Structural Equation Modeling, LLM usage analysis	Engagement and performance	ChatGPT perception significantly predicts engagement and academic performance	ScienceDirect
S47	(Mwapwele & Roodt, 2018)	Tanzania and South Africa	High School	-	Accessibility: Mobile learning adoption	Explored teachers' use of mobile devices outside classroom; emphasized knowledge acquisition and tech adoption	IEEE Xplore
S48	(Ncube & Ngulube, 2025)	South Africa	Higher Education	Learning analytics, NLP, predictive modeling	Personalization and student success	Learning analytics enable proactive, personalized support; ethical and governance challenges remain	ScienceDirect
S49	(Öncel et al., 2021)	USA	All levels	WAT (NLP tool), ML Prediction Algorithms: Linear Reg., SVM, Linear SVM Polynomial, Random Forest, and Gradient Boosting Tree.	Personalization: Automatic writing evaluation	Assess impact of individual differences on writing	ScienceDirect
S50	(Paik et al., 2024)	Japan	Elementary	Predictive ML model.	School student engagement	Links engagement, emotions, and performance	ScienceDirect
S51	(Ravuri et al., 2023)	India	High School and Online Learning	Assessment automation ML algorithms	Personalized and flexible learning	ML improves engagement, reduces bias in assessments, supports adaptive teaching	IEEE Xplore
S52	(Rodríguez-Ortiz et al., 2025)	Spain	High School	GenAI (BERT, GPT-4)	Engagement: Learning analytics	Systematic review; ML dominates engagement/dropout prediction	WOS
S53	(Semi et al., 2025)	USA	Primary/Child learning	Sentence-BERT, T5 Transformer, Gradient Boosting	Personalization and adaptive learning	AI-driven personalization improved learning accuracy (85.2%) and learner preference	WOS
S54	(Sethosa et al., 2025)	South Africa	High School and University	TF-IDF + SVM, Random Forest, Logistic Regression	Personalization: Hate speech detection in isiZulu	Semantic analysis improves Zulu hate speech detection; contributes to inclusive digital education	IEEE Xplore
S55	(Shaghaghi et al., 2021)	India	Primary and Secondary	NLP video translation pipeline	Multilingual Support: Support mother-tongue education	Pipeline translates English videos to Punjabi; aims for culturally relevant content	IEEE Xplore

Table A1. Cont.

Index	Authors	Country/Region	Educational Level	ML/NLP Technique	Educational Purpose	Key Findings Relevant to RQs	Database
S56	(Simon & Zeng, 2024)	China	All levels	Adaptive learning ML models	Personalization: Educational tools that personalize instruction	Adaptive systems support personalized learning, but effective use is constrained by system complexity, limited teacher capacity, and insufficient institutional support	MDPI
S57	(Simpson et al., 2024)	USA	University	Additive Factors models	Accessibility: How learning rates differ across students within ITS	Adaptive platforms can personalize pacing through increased learning opportunities rather than complex learner-specific models	EDM 2024
S58	(Singh et al., 2024)	India	Primary and Secondary	Optical Character Recognition	Multilingual learning, accessibility	OCR + AI system translated and taught content in multiple languages; personalized and inclusive learning platform	IEEE Xplore
S59	(Storey & Wagner, 2024)	USA	High School and University	LLM, ML adaptive learning algorithms	Accessibility: Opportunities of integrating AI into education	Reviews AI opportunities, challenges, and applications	WOS
S60	(X. Sun et al., 2024)	Japan	High School and University	Machine translation	Improve online learning accessibility	MT courses with toggle increased completion by 2.3%; 5-star ratings 88% vs. 84%	IEEE Xplore
S61	(Khan et al., 2025)	Pakistan	University	Predictive analytics ML	Engagement: Student achievement improvement	ML enhances performance prediction and achievement	ScienceDirect
S62	(Suresh Manic et al., 2024)	Oman	High School	SVM, KNN, Random Forest, Decision Trees	Personalization and guidance	ML predicts academic success (SVM accuracy 77%), supporting personalized guidance	ScienceDirect
S63	(Tang & Chen, 2024)	China	Higher Education	ML-based adaptive learning	Engagement: Improve learning outcomes	Adaptive platform improved performance, satisfaction, and learning behaviors	IEEE Xplore
S64	(Tennakoon et al., 2021)	Turkey	High School	SVM Model, MaLSTM Model, DistilBERT Model	Student engagement	E-learning system improves engagement	IEEE Xplore
S65	(Thomas et al., 2023)	USA	All levels	NLP and ML ASAG, GPT	Accessibility: AI/ML-based automatic short answer grading (ASAG) for rapid feedback	Shows accessible NLP/ML applications for education but requires refinement for broad use	WOS
S66	(Tsimba et al., 2022)	Zimbabwe	Primary and Secondary	ML eLearning analytics	Fair access to eLearning	Highlighted rural-urban disparities in eLearning during COVID-19; proposed adoption strategies	IEEE Xplore
S67	(Ujjain et al., 2025)	India	Primary and Secondary	Isolation Forest, LightGBM	Classify and prioritize schools for intervention	85% of schools structured; 15% odd; supports resource allocation and policy planning	IEEE Xplore
S68	(Vega & Tejeda, 2023)	Mexico	Higher Education	ML predictive systems	Engagement: Enhance cognitive learning in STEM	ML predicts student attrition, monitors performance, informs interventions	IEEE Xplore
S69	(Wang et al., 2025)	China	High School	Recurrent Neural Networks (RNN)	Engagement and personalization	ML detects mindset states, enabling adaptive, student-centered teaching	ScienceDirect

Table A1. Cont.

Index	Authors	Country/Region	Educational Level	ML/NLP Technique	Educational Purpose	Key Findings Relevant to RQs	Database
S70	(Cao et al., 2026)	China	Primary and Secondary	ML multilevel modeling	Accessibility: AI—Technological Pedagogical Content Knowledge	School-level strategies and AI competence predict AI-TPACK	ScienceDirect
S71	(Yang et al., 2024)	China	High School	ML cloud computing data aggregation	Multilingual Support: English translation assistance	Translation accuracy 0.93–1; cloud aggregation improves system robustness and fluency	IEEE Xplore
S72	(Yilmaz & Deniz, 2024)	Turkey	Higher Education	Sentiment analysis, text mining, GPT-based item generation	Assessment and feedback	NLP enables automated assessment, feedback, and multilingual evaluation support	WOS

Appendix E

The following table presents a summary of the quality assessment for all 72 studies included in this systematic review. Studies were evaluated using PRISMA-aligned criteria, considering clarity of research questions, methodological rigor, empirical validity, and relevance to ML/NLP applications in educational contexts. Studies are categorized as High, Moderate, or Lower quality to provide transparency regarding the strength of the evidence base. High quality studies demonstrate robust methodology and measurable outcomes, moderate quality studies have minor limitations but provide empirical support, and lower quality studies are primarily conceptual or exploratory with limited empirical validation.

Table A2. Quality Assessment Summary of Included Studies (n = 72). The table categorizes studies based on methodological rigor, empirical evidence, and relevance to ML/NLP-driven educational applications.

Quality Category	Assessment Description	No of Studies
High Quality	Clear research questions, robust methodology, empirical results, and transparent reporting; includes studies with strong ML/NLP evaluation and measurable educational outcomes	58
Moderate Quality	Sound methodology with minor limitations, e.g., small datasets, single case studies, or short evaluation periods; still provides empirical evidence	13
Lower Quality	Conceptual, exploratory, or mixed-level studies meeting inclusion criteria but with limited empirical validation; often foundational frameworks, general AI/ML recommendations, or non-specific interventions	1

References

Ahmed, M. (2021). *Generating effective sentence representations: Deep learning and reinforcement learning approaches* [Doctoral dissertation, The University of Western Ontario].

Akallouch, O., Akallouch, M., & Fardousse, K. (2025). Advances in amazigh language technologies: A comprehensive survey across processing domains. *Information*, 16(7), 600. [CrossRef]

Alabi, J. O., Adelani, D. I., Mosbach, M., & Klakow, D. (2022). Adapting pre-trained language models to African languages via multilingual adaptive fine-tuning. *arXiv*, arXiv:2204.06487. [CrossRef]

Al Balushi, M. Y., & Al Harthi, A. S. (2024). Adaptive learning pathways: Advancing personalization in online education platforms. In *2024 2nd international conference on computing and data analytics (ICCDa)* (pp. 1–6). IEEE. [CrossRef]

Almeida, G., Naukkarinen, J., Jantunen, T., Datta, S., Kuparinen, K., & Vakkilainen, E. (2025). Improving the student learning process in MOOCs through the analysis of open-ended question-based assessments using natural language processing. *Futureproofing Engineering for Global Responsibility*, 1260, 525–533. [CrossRef]

- Andela. (2025). *Andela learning community*. Andela. Available online: <https://www.andela.com/andela-learning> (accessed on 31 August 2025).
- Assami, S., Daoudi, N., & Ajhoun, R. (2018). Personalization criteria for enhancing learner engagement in MOOCs. *IEEE EDUCON*, 2018, 1265–1272.
- Bensalah, H. (2024). Exploring the potential synergy of Quillbot as a natural language processing tool in demystifying academic writing. *ATRAS Journal*, 1, 100779. [CrossRef]
- Bentley, S. V., Naughtin, C. K., McGrath, M. J., Irons, J. L., & Cooper, P. S. (2024). The digital divide in action: How experiences of digital technology shape future relationships with artificial intelligence. *AI and Ethics*, 4(4), 901–915. [CrossRef]
- Bernasconi, E., Redavid, D., & Ferilli, S. (2025). Enhancing personalised learning with a context-aware intelligent question-answering system and automated frequently asked question generation. *Electronics*, 14(7), 1481. [CrossRef]
- Blocker, A., Meyer, F., Buys, J., Tanaka, H., Dabre, R., Song, H., Chakrabarty, A., & Tanaka, H. (2025). Benchmarking IsiXhosa automatic speech recognition. In *Proceedings of the 4th clinical NLP workshop* (pp. 1–10). Available online: <https://aclanthology.org/2025.cl4health-1.14.pdf> (accessed on 1 August 2025).
- Bobesh, S. J., Miller, T., Newman, P., Liu, Y., & Elglaly, Y. N. (2023, March 15–18). *Towards machine learning fairness education in a natural language processing course*. 54th ACM Technical Symposium on Computer Science Education (pp. 312–318), Toronto, ON, Canada.
- Bode, S. P., & Satpute, R. S. (2024, November 29–30). *Natural language processing in education system*. 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI) (pp. 1–5), Wardha, India. [CrossRef]
- Bosch, T., Jordaan, M., Mwaura, J., Nkoala, S., Schoon, A., Smit, A., Uzuegbunam, C. E., & Mare, A. (2023). *South African university students' use of AI-powered tools for engaged learning*. Social Science Research Network. Available online: <https://ssrn.com/abstract=4595655> (accessed on 26 June 2025).
- Bousalem, S., Benchikha, F., & Marir, N. (2025). Personalized learning through MBTI prediction: A deep learning approach integrated with learner profile ontology. *IEEE Access*, 13, 30861–30873. [CrossRef]
- Cale, W., Haerawan, H., & Barroso, U. (2024). The effectiveness of interactive videos in increasing student engagement in online learning. *Journal of Computer Science Advancements*, 2(5), 244–258. [CrossRef]
- Cao, X., Huang, Z., Li, M., & He, T. (2026). Teachers' AI-TPACK as a tangible outcome in the digital transformation of education: A machine learning-based multilevel approach. *Teaching and Teacher Education*, 169, 105270. [CrossRef]
- Chang, S.-C., & Chang, Y.-C. (2025). Teaching practices and effectiveness assessment of AI courses assisted by large language models. In H. Fujita, Y. Watanobe, M. Ali, & Y. Wang (Eds.), *Advances and trends in artificial intelligence: Theory and applications*, IEA/AIE 2025, PT II (Vol. 15707, pp. 89–100). Springer. [CrossRef]
- Chen, J., Zhou, X., Yao, J., & Tang, S.-K. (2025). Application of machine learning in higher education to predict students' performance, learning engagement and self-efficacy: A systematic literature review. *Asian Education and Development Studies*, 14(2), 205–240. [CrossRef]
- Chen, X. L., Xie, H. R., Qin, S. J., Wang, F. L., & Hou, Y. A. (2025). Artificial intelligence-supported student engagement research: Text mining and systematic analysis. *European Journal of Education*, 60(1), e70008. [CrossRef]
- Chung, W. Y., Zhang, X. Y., Ahmad, Z., Sayadi, H., & Rafatirad, S. (2022). Machine Learning to the rescue: ML-assisted framework for equity-driven education. In I. Kallel, H. M. Kammoun, A. Akkari, & L. Hsairi (Eds.), *Proceedings of the 2022 IEEE global engineering education conference (EDUCON 2022)*, Gammarth, Tunisia, March 28–31 (pp. 1254–1263). IEEE. [CrossRef]
- Deas, N., Grieser, J., Kleiner, S., Patton, D., Turcan, E., & McKeown, K. (2023, December 6–10). *Evaluation of African American language bias in natural language processing models*. Conference on Empirical Methods in Natural Language Processing (EMNLP 2023) (pp. 1–15), Singapore. Available online: <https://aclanthology.org/2023.emnlp-main.421.pdf> (accessed on 31 August 2025).
- Dehbi, A., Bakhouyi, A., Dehbi, R., & Talea, M. (2025). Towards technology-enhanced learning: A novel machine learning approach in Education 4.0. *Telematics and Informatics Reports*, 18, 100194. [CrossRef]
- Democratic Alliance Gauteng Provincial Legislature. (2025, July 14). *87% of gauteng public schools are not connected to the internet*. Democratic Alliance Gauteng. Available online: <https://dagauteng.org.za/2025/07/87-of-gauteng-public-schools-are-not-connected-to-the-internet> (accessed on 2 August 2025).
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv*, arXiv:1810.04805. [CrossRef]
- Di Pietro, G., & Castaño Muñoz, J. (2025). A meta-analysis on the effect of technology on the achievement of less advantaged students. *Computers & Education*, 226, 105197. [CrossRef]
- Docebo. (2025). *Modern LMS vs. traditional LMS: A detailed comparison*. Docebo Blog. Available online: <https://www.docebo.com/learning-network/blog/modern-lms-vs-traditional-lms/> (accessed on 31 August 2025).
- du Toit, J. S., & Puttkammer, M. J. (2021). Developing core technologies for resource-scarce Nguni languages. *Information*, 12, 520. [CrossRef]
- Ehsan, U., Singh, R., Metcalf, J., & Riedl, M. O. (2022). The algorithmic imprint. *arXiv*, arXiv:2206.03275. [CrossRef]

- Eiselen, R., & Puttkammer, M. (2014). Developing text resources for ten South African languages. In N. Calzolari (Ed.), *Proceedings of the ninth international conference on language resources and evaluation (LREC'14), Reykjavik, Iceland, May 26–31* (pp. 3698–3703). European Language Resources Association (ELRA). Available online: http://www.lrec-conf.org/proceedings/lrec2014/pdf/1151_Paper.pdf (accessed on 1 August 2025).
- Emezue, C. C., & Bonaventure, D. (2022). MMTAfrica: Multilingual machine translation for African languages. *arXiv*, arXiv:2204.04306. [CrossRef]
- Eneza Education. (2016, August 26). *Toast to a million: Eneza reaches over 1 million users*. Eneza Education Blog. Available online: <https://www.enezaeducation.com/2016/08/26/toast-to-a-million-eneza-reaches-over-1-million-users/> (accessed on 31 August 2025).
- Eskom. (2023). *Cloud solutions cost analysis report*. Available online: <https://tenderbulletin.eskom.co.za/webapi/api/Files/GetFile?FileID=423592> (accessed on 1 August 2025).
- Esmail, S., Bar, K., & Dershowitz, N. (2022). How much does lookahead matter for disambiguation? Partial Arabic diacritization case study open access. *Computational Linguistics*, 48(4), 1103–1123. [CrossRef]
- Farshad, S., Zorin, E., Amangeldiuly, N., & Fortin, C. (2024). Engagement assessment in project-based education: A machine learning approach in team chat analysis. *Education and Information Technologies*, 29(10), 13105–13131. [CrossRef]
- Gaustad, T., & Puttkammer, M. (2022). Development of linguistically annotated parallel language resources for four South African languages. *Journal of the Digital Humanities Association of Southern Africa (DHASA)*, 3(3). [CrossRef]
- Gauteng DBE. (2023). *E-learning expenditure in Gauteng: Delivering education in the world of digital technologies, and assessing value for money—Gauteng analysis*. Available online: <https://www.gtac.gov.za/pepa/wp-content/uploads/2024/09/Spending-Review-on-the-e-learning-programme-19122023-Final.pdf> (accessed on 18 June 2025).
- Gauteng Department of Basic Education. (2023). *Annual report on EdTech adoption 2023*. Available online: https://provincialgovernment.co.za/department_annual/1351/2023-gauteng-education-annual-report.pdf (accessed on 18 June 2025).
- Gauteng Department of Education. (2023). *Education annual report 2022/2023 financial year*. Gauteng Department of Education.
- GB Hackers. (2023). *Microsoft enforces MFA for logging into Azure portal*. Available online: <https://gbhackers.com/microsoft-enforces-mfa> (accessed on 1 August 2025).
- Giannini, T., & Bowen, J. P. (2022). Computational culture: Transforming archives practice and education for a post-covid world. *ACM Journal on Computing and Cultural Heritage*, 15, 47. [CrossRef]
- Gilfillan, I. (2022). *Biggest South African language wikipeidias and their editors—2022*. MyBroadband. Available online: <https://mybroadband.co.za/news/internet/500319-biggest-wikipedia-south-african-languages-and-editors.html> (accessed on 1 August 2025).
- Gillis, A. S., & Barney, N. (2024). *Sentiment analysis (opinion mining)*. TechTarget. Available online: <https://www.techtarget.com/searchbusinessanalytics/definition/opinion-mining-sentiment-mining> (accessed on 31 August 2025).
- GitHub. (2025). *Masakhane*. Available online: <https://github.com/masakhane-io/> (accessed on 6 April 2025).
- Gleason, N. W. (2018). *Higher education in the era of the fourth industrial revolution*. Springer. [CrossRef]
- Gligorea, I., Cioca, M., Oancea, R., Gorski, A.-T., Gorski, H., & Tudorache, P. (2023). Adaptive learning using Artificial intelligence in e-learning: A literature review. *Education Sciences*, 13(12), 1216. [CrossRef]
- Global Citizen. (2025, March 31). *How South Africa's youth can lead the future of work..* Forbes. Available online: <https://www.forbes.com/sites/globalcitizen/2025/03/31/how-south-africas-youth-can-lead-the-future-of-work/> (accessed on 31 August 2025).
- Graves, E., Heffelman, A., Olt, L., Bomben, N., Elglaly, Y. N., Ahmed, S., & Sharmin, M. (2025, July 8–11). *Beyond one-size-fits-all: GPT-enabled personalization of academic content for neurodiverse students*. IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC) (pp. 581–592), Toronto, ON, Canada.
- Grech, A., & Camilleri, A. F. (2017). *Blockchain in education*. Publications Office of the European Union. Available online: https://publications.jrc.ec.europa.eu/repository/bitstream/JRC108255/jrc108255_blockchain_in_education%281%29.pdf (accessed on 1 August 2025).
- Grigalashvili, E., & Olaresko, D. (2025). *ML in education: Use cases and benefits*. AnyforSoft. Available online: <https://anyforsoft.com/blog/machine-learning-in-education/> (accessed on 22 May 2025).
- Gulyamov, S. S., Fayziev, R. A., Rodionov, A. A., & Jakupov, G. A. (2023, June 14–16). *Leveraging semantic analysis in machine learning for addressing unstructured challenges in education*. 3rd International Conference on Technology Enhanced Learning in Higher Education (TELE) (pp. 5–7), Lipetsk, Russia. [CrossRef]
- Haering, M., Bano, M., Zowghi, D., Kearney, M., & Maalej, W. (2021). Automating the evaluation of education apps with app store data. *IEEE Transactions on Learning Technologies*, 14(1), 16–27. [CrossRef]
- Hassanain, A. A., Amor, Y., Abidi, R., & Sahli, N. (2025). Impact of machine learning on personalized learning and course customization. *Procedia Computer Science*, 270, 3818–3827. [CrossRef]
- ICDL Africa. (2025). *Digital literacy and employment trends*. Available online: <https://icdl.org/africa> (accessed on 17 June 2025).

- IITPSA. (2025). *Pushing for NLP for Africa—IITPSA*. Available online: <https://www.iitpsa.org.za/pushing-for-nlp-for-africa/> (accessed on 12 May 2025).
- Information Regulator SA. (n.d.). *Annual report 2022/2023*. Available online: <https://infoeregulator.org.za/wp-content/uploads/2020/07/Information-Regulator-Annual-Report-2023-Compressed.pdf> (accessed on 1 August 2025).
- Iyamu, R. (2024). Machine translation and NLP tools: Developing and refining language technologies for African languages. *International Journal For Multidisciplinary Research*, 6(5), 1–5. [CrossRef]
- Janks, H. (2014). Globalisation, diversity, and education: A South African perspective. *The Educational Forum*, 78(1), 8–25. [CrossRef]
- Jauhainen, J. S., Ntinda, M. N., & Sutinen, E. (2025, May 28–30). *Integrating generative AI to higher education systems in Africa: Reflections from tests in Namibia*. 2025 IST-Africa Conference (IST-Africa) (pp. 1–8), Nairobi, Kenya. [CrossRef]
- Jayaprakash, S., Kathiresan, V., Shanmugapriya, N., & Dadhich, M. (2021). Chapter 2 Opportunities and challenges in transforming higher education through machine learning. In K. Kant Hiran, D. Khazanchi, A. Kumar Vyas, & S. Padmanaban (Eds.), *Machine learning for sustainable development* (pp. 17–30). De Gruyter. [CrossRef]
- Jeon, C.-H., Shin, J.-Y., & Ryu, S. (2025). Analyzing student communication patterns in science classes using machine learning and natural language processing: A case study on high school science education. *Journal of Science Education and Technology*, 34, 1397–1417. [CrossRef]
- Jha, A., Mann, P., Tiwari, A., Kadian, K., Sharma, A., & Singh, J. (2024). Evolving education in India: Multilingual LLMs for vernacular learning experiences. In *Proceedings of the 2024 international conference on computer, electronics, electrical engineering & their applications (IC2E3)* (pp. 1–5). IEEE. [CrossRef]
- Jimes, C., Weiss, S., & Keep, R. (2013). Addressing the Local in localization: A case study of open textbook adoption by three South African teachers. *Journal of Asynchronous Learning Networks*, 17(2), 73–86. [CrossRef]
- Ju, L., & Salvosa, A. A. (2024). Research and optimization of English automatic translation system based on machine learning algorithm. In *2024 9th international symposium on computer and information processing technology (ISCIPT)* (pp. 1–5). IEEE. [CrossRef]
- Kantipudi, M. V. V. P., Aluvalu, R., Maheswari, U., & Raisinghani, M. S. (2022). Insights on implications of cognitive computing in leveraging online education systems. *International Journal of Online Pedagogy and Course Design*, 12(1), 1–6. [CrossRef]
- Kgosi, M. K., Makgato, M., & Skosana, N. M. (2023). Teachers' views on the application of educational technologies in the classroom: A case of selected Tshwane West secondary schools in Gauteng. *Journal of Curriculum Studies Research*, 5(2), 151–166. [CrossRef]
- Khaled, D. (2014). Natural language processing and its use in education. *International Journal of Advanced Computer Science and Applications*, 5(12), 72–76. [CrossRef]
- Khan, S., Mazhar, T., Shahzad, T., Khan, M. A., Waheed, W., Waheed, A., & Hamam, H. (2025). Predictive analytics in education—Enhancing student achievement through machine learning. *Social Sciences & Humanities Open*, 12, 101824. [CrossRef]
- Kim, B., Suh, H., Heo, J., & Choi, Y. (2020). AI-driven interface design for intelligent tutoring system improves student engagement. *arXiv*, arXiv:2009.08976. [CrossRef]
- Kimanga, P., Kituku, B., Moso, J., & Gikunda, P. (2025, November). Performance of transformer-based models in Dholuo-Swahili translation for low-resourced language. In *2025 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)* (pp. 235–240). IEEE. [CrossRef]
- Koedinger, K. R., Carvalho, P. F., Liu, R., & McLaughlin, E. A. (2023). An astonishing regularity in student learning rate. *Proceedings of the National Academy of Sciences*, 120(50), e221311120. [CrossRef] [PubMed]
- Kormos, J. (2023). The role of cognitive factors in second language writing and writing to learn a second language. *Studies in Second Language Acquisition*, 45, 622–646. [CrossRef]
- Kulik, J. A., & Fletcher, J. D. (2015). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 85, 149–176. [CrossRef]
- Kumar, L., & Kalaavathi, B. (2025, June 25–26). *AI-powered multilingual chatbot for inclusive learning: Enhancing cybersecurity in knowledge sharing*. 2025 Second International Conference on Cognitive Robotics and Intelligent Systems (ICC—ROBINS) (pp. 567–572), Coimbatore, India. [CrossRef]
- Kunicina, N., Rakhimova, K., Caiko, J., & Mansurova, M. (2025). Adaptive multidimensional e-learning systems. In A. Romanovs, D. Navakauskas, D. Plonis, & M. Narigina (Eds.), *12th workshop on advances in information, electronic and electrical engineering (AIEEE)*, Vilnius, Lithuania, May 15–17. IEEE. [CrossRef]
- Kuure, J.-A. (2021, August 23). *Andela learning community surpasses 100,000 learners*. Tech Labari. Available online: <https://techlabari.com/andela-learning-community-surpasses-100000-learners/> (accessed on 31 August 2025).
- Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4, 100779. [CrossRef] [PubMed]
- Ling, Z. (2021, December 17–19). *Online education data application analysis based on machine learning algorithms*. 2021 International Symposium on Advances in Informatics, Electronics and Education (ISAIEE) (pp. 126–128), Frankfurt, Germany. [CrossRef]
- Litman, D. (2016). Natural language processing for enhancing teaching and learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30. [CrossRef]

- Liu, J., Fang, Y., Yu, Z., & Wu, T. (2022, March 25–27). *Design and construction of a knowledge database for learning japanese grammar using natural language processing and machine learning techniques*. 4th International Conference on Natural Language Processing (ICNLP) (pp. 371–375), Xi'an, China. [CrossRef]
- Liu, L., Li, X., Qi, Y., & Huang, J. (2024, December 20–22). *Exploring the overall development direction of China's higher education based on machine learning*. 2024 5th International Conference on Information Science and Education (ICISE-IE) (pp. 217–220), Zhanjiang, China. [CrossRef]
- Lundström, O., Maleki, N., & Ahlgren, F. (2024, May 8–11). *Online course improvement through GPT-4: Monitoring student engagement and dynamic FAQ generation* [Conference paper]. IEEE Global Engineering Education Conference (EDUCON 2024), Kos, Greece. [CrossRef]
- Lv, W., Yang, C., & Zhang, W. (2023, April 21–23). *Implementation dilemma and breakthrough path of AI education in primary and secondary schools in less developed areas*. 5th International Conference on Computer Science and Technologies in Education (CSTE) (pp. 248–255), Xi'an, China. [CrossRef]
- Ma, F. (2025). Learning behavior analysis and personalized recommendation system of online education platform based on machine learning. *Computers and Education: Artificial Intelligence*, 8, 100408. [CrossRef]
- Mabokela, K. R., Celik, T., & Raborife, M. (2023). Multilingual sentiment analysis for under-resourced languages: A systematic review of the landscape. *IEEE Access*, 11, 15996–16020. [CrossRef]
- Mahamad, S., Chin, Y. H., Zulmuksah, N. I. N., Haque, M. M., Shaheen, M., & Nisar, K. (2025). Technical review: Architecting an AI-driven decision support system for enhanced online learning and assessment. *Future Internet*, 17(9), 383. [CrossRef]
- Maina, A. M., & Kuria, J. (2024, May 20–24). *Building an AI future: Research and policy directions for Africa's higher education*. 2024 IST-Africa Conference (IST-Africa) (pp. 1–9), Dublin, Ireland. [CrossRef]
- Martinus, L., & Abbott, J. Z. (2019). A focus on neural machine translation for African languages. *arXiv*, arXiv:1906.05685. [CrossRef]
- Martinus, L., Webster, J., Moonsamy, J., Jnr, M. S., Moosa, R., & Fairon, R. (2020). Neural machine translation for south Africa's official languages. *arXiv*, arXiv:2005.06609. [CrossRef]
- Masakhane Project. (2025). *African NLP initiatives*. Available online: <https://www.masakhane.io/> (accessed on 6 April 2025).
- Masethe, H. D., Mothapo, L. M., Ojo, S. O., Owolawi, P. A., Masethe, M. A., & Giunchiglia, F. (2024, November 27–29). *Machine translation for morphologically rich low-resourced South African languages*. 2024 4th International Multidisciplinary Information Technology and Engineering Conference (IMITEC) (pp. 71–78), Vanderbijlpark, South Africa. [CrossRef]
- Matellio. (2025, April 15). *How NLP in education sector can enhance learning experience?* Artificial Intelligence, Education & eLearning, Natural Language Processing Blog Category. Available online: <https://www.matellio.com/blog/nlp-in-education/> (accessed on 10 May 2025).
- McKellar, C. A., & Puttkammer, M. J. (2020). Dataset for comparable evaluation of machine translation between 11 South African languages. *Data in Brief*, 29, 105146. [CrossRef]
- Mehrvarz, M., Salimi, G., Abdoli, S., & McLaren, B. M. (2025). How students' perception of ChatGPT shapes online learning engagement. *Computers & Education: AI*, 9, 100459.
- Mengist, W., Soromessa, T., & Legese, G. (2020). Method for conducting systematic literature review and meta-analysis for environmental science research. *MethodsX*, 7, 100777. [CrossRef] [PubMed]
- Mesham, S., Hayward, L., Shapiro, J., & Buys, J. (2021). Low-resource language modelling of South African languages. *arXiv*, arXiv:2104.00772. [CrossRef]
- Microsoft. (2022). *Implementing multi-factor authentication with MSAL*. Available online: <https://learn.microsoft.com/en-us/azure/active-directory/develop/msal-overview> (accessed on 1 August 2025).
- Microsoft. (2023). *ML.NET documentation*. Microsoft. Available online: <https://learn.microsoft.com/en-us/dotnet/machine-learning/> (accessed on 1 August 2025).
- Mncube, L. (2023). *Internet of things affordance for open educational resources in a comprehensive open distance e-learning*. Available online: <https://digitalcommons.kennesaw.edu/acist/2023/presentations/13> (accessed on 19 June 2025).
- Muhammad, S. H. (2025, February 27). *Tales from our community: Empowering NLP in low-resource languages*. Label Studio Blog. Available online: <https://labelstud.io/blog/tales-from-our-community-empowering-nlp-in-low-resource-languages/> (accessed on 31 August 2025).
- Mwapwele, D. S., & Roodt, S. (2018, March 8–9). *Teacher's adoption and use of mobile devices outside the classroom for learning in Africa*. 2018 Conference on Information Communications Technology and Society (ICTAS) (pp. 1–7), Durban, South Africa. [CrossRef]
- Nasir, M. U., & Mchechesi, I. A. (2022). Geographical distance is the new hyperparameter: A case study of finding the optimal pre-trained language for English-isiZulu machine translation. *arXiv*, arXiv:2205.08621. [CrossRef]
- Ncube, M. M., & Ngulube, P. (2025). Leveraging learning analytics to personalise academic library services. *Journal of Academic Librarianship*, 51(6), 103160. [CrossRef]

- NLLB Team, Costa-jussà, M. R., Cross, J., Çelebi, O., Elbayad, M., Heffernan, K., Kalbassi, E., Lam, J., Licht, D., Maillard, J., Sun, A., Wang, S., Wenzek, G., Youngblood, A., Akula, B., Barrault, L., Mejia Gonzalez, G., Hansanti, P., . . . Wang, J. (2022). No language left behind: Scaling human-centered machine translation. *arXiv*, arXiv:2207.04672. [CrossRef]
- Olaogun, O. (2014, June 13). *Tutor.ng—A Startup with a new model for education in Nigeria*. How We Made It in Africa. Available online: <https://www.howwemadeitinafrica.com/tutor-ng-a-startup-with-a-new-model-for-education-in-nigeria/40459/> (accessed on 31 August 2025).
- Öncel, P., Flynn, L. E., Sonia, A. N., Barker, K. E., Lindsay, G. C., McClure, C. M., McNamara, D. S., & Allen, L. K. (2021, April 12–16). *Automatic student writing evaluation: Investigating the impact of individual differences on source-based writing* [Conference paper]. LAK21 Conference Proceedings: The Eleventh International Conference on Learning Analytics & Knowledge, ELECTR Network (pp. 620–625), Virtual. [CrossRef]
- Paik, J. H., Himelfarb, I., Yoo, S. H., Yoo, K., & Ha, H. Y. (2024, March 18–22). *The relationships among school engagement, students' emotions, and academic performance in an elementary online learning* [Conference paper]. 14th Annual International Conference on Learning Analytics and Knowledge (LAK 2024): Learning Analytics in the Age of Artificial Intelligence (pp. 219–230), Kyoto, Japan. [CrossRef]
- Parliament of South Africa. (2025, March 11). *Size and shape of the education sector in Gauteng. Ministerial briefing session on school infrastructure and equipment*. National Council of Provinces, Parliament of South Africa. Available online: https://www.parliament.gov.za/storage/app/media/Pages/2025/05-03-2025_NCOP_Ministerial_Briefing/2/Gauteng.pdf (accessed on 31 August 2025).
- Rai, H. K., Cavalcanti Barroso, A., Yates, L., Schneider, J., & Orrell, M. (2020). Involvement of people with dementia in the development of technology-based interventions: Narrative synthesis review and best practice guidelines. *Journal of Medical Internet Research*, 22(12), e17531. [CrossRef] [PubMed]
- Ravuri, A., Lourens, M., Aswini, S., Nijhawan, G., Zabibah, R. S., & Chandrashekar, R. (2023, December 1–3). *Improving personalized education: A machine learning method for flexible learning environments*. 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) (pp. 1715–1720), Gautam Buddha Nagar, India. [CrossRef]
- Rodríguez-Ortiz, M. A., Santana-Mancilla, P. C., & Anido-Rifón, L. E. (2025). Machine learning and generative AI in learning analytics for higher education: A systematic review of models, trends, and challenges. *Applied Sciences*, 15(15), 8679. [CrossRef]
- Sabiteka, M. (2025). Toward sustainable education: A contextualized model for educational technology adoption in developing countries. *Sustainability*, 17, 3592. [CrossRef]
- Samuel, J., Khanna, T., Esguerra, J., Sundar, S., Pelaez, A., & Bhuyan, S. S. (2025). The rise of artificial intelligence phobia! Unveiling news-driven spread of AI fear sentiment using ML, NLP, and LLMs. *IEEE Access*, 13, 125944–125969. [CrossRef]
- Semi, M. M. A., Uddin, M. B., Sultana, S., Tamanna, M., Uddin, A., & Ahmed, K. R. (2025). AI-driven education: Integrating machine learning and NLP to transform child learning systems. *International Journal of Advanced Computer Science and Applications*, 16(6), 16–22. [CrossRef]
- Sethosa, M. R., Rananga, S., & Mbooi, M. S. (2025, May 28–30). *Hate speech detection in isiZulu: A semantic approach*. 2025 IST-Africa Conference (IST-Africa) (pp. 1–9), Nairobi, Kenya. [CrossRef]
- Shaghaghi, N., Ghosh, S., & Kapoor, R. (2021, October 19–23). *ClassRoute: An English to Punjabi educational video translation pipeline for supporting Punjabi mother-tongue education*. 2021 IEEE Global Humanitarian Technology Conference (GHTC) (pp. 342–348), Seattle, WA, USA. [CrossRef]
- Shaik, T., Tao, X., Li, Y., Dann, C., McDonald, J., Redmond, P., & Galligan, L. (2022). A review of the trends and challenges in adopting natural language processing methods for education feedback analysis. *IEEE Access*, 10, 56720–56739. [CrossRef]
- Simon, P. D., & Zeng, L. M. (2024). Behind the scenes of adaptive learning: A scoping review of teachers' perspectives on the use of adaptive learning technologies. *Education Sciences*, 14, 1413. [CrossRef]
- Simpson, M. A., Norberg, K. A., & Fancsali, S. E. (2024). Replicating an “astonishing regularity in student learning rate”. In *2024 international conference on educational data mining (EDM 2024)*, Atlanta, GA, USA, July 14–17 (pp. 40–47). Educational Data Mining Society. Available online: <https://www.educationaldatamining.org/edm2024/proceedings/2024.EDM-short-papers.40/> (accessed on 31 August 2025).
- Singh, H., Mohammad, S., Yaseen, A., Molawade, M., Mohite, S. G., Jadhav, V., & Jadhav, R. (2024, April 25–27). *Multilingual education through optical character recognition (OCR) and AI*. 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSociCon) (pp. 1–6), Pune, India. [CrossRef]
- South African Government. (2021). *POPIA Compliance guidelines for EdTech*. DCDT.
- Statistics South Africa. (2024). *A profile of education enrolment, attainment and progression in South Africa*. Education and Child Statistics Directorate. Available online: <https://www.statssa.gov.za/publications/Report-03-01-81/PresentationUpdatedFinal.pdf> (accessed on 31 August 2025).
- Storey, V. A., & Wagner, A. (2024). Integrating Artificial Intelligence (AI) into adult education: Opportunities, challenges, and future directions. *International Journal of Adult Education and Technology-IJAET*, 15, 345921. [CrossRef]

- Strielkowski, W. (2025). AI-Driven adaptive learning for sustainable educational development. *Sustainability*, 17, 1234. [CrossRef]
- Sun, X., Wong, A., & Urban, A. (2024, March 18–20). *Breaking language barriers: The power of machine translation in online learning*. 2024 12th International Conference on Information and Education Technology (ICIET) (pp. 298–301), Yamaguchi, Japan. [CrossRef]
- Sun, Z., Yu, H., Song, X., Liu, R., Yang, Y., & Zhou, D. (2020). MobileBERT: A compact task-agnostic BERT for resource-limited devices. *arXiv*, arXiv:2004.02984.
- Suresh Manic, K., Al-Bemani, A. S., Nizamudin, A. A., Balaji, G., & Amal, A. A. (2024). Optimizing academic journey for high schoolers in Oman. *Procedia Computer Science*, 235, 2716–2729. [CrossRef]
- Tajik, O., Golzar, J., & Noor, S. (2024). Purposive sampling. *International Journal of Education & Language Studies*, 2, 1–9. [CrossRef]
- Tan, P.-N., Steinbach, M., & Kumar, V. (2018). *Introduction to data mining* (2nd ed., pp. 209–215). Pearson.
- Tang, X., & Chen, Y. (2024, March 4–5). *Adaptive education platform based on machine learning: A new way to improve the quality of higher education*. 2024 International Conference on Interactive Intelligent Systems and Techniques (IIST) (pp. 310–316), Bhubaneswar, India. [CrossRef]
- Tennakoon, S., Wickramaarachchi, T., Weerakotuwa, R., Sulochana, P., Karunasena, A., & Piyawardana, V. (2021). E-pod: E-learning system for improving student engagement in asynchronous mode. In A. Yousaf (Ed.), *7th international conference on engineering and emerging technologies (ICEET 2021), Istanbul, Turkey, October 27–28* (pp. 581–586). IEEE. [CrossRef]
- Thomas, D. R., Gupta, S., & Koedinger, K. R. (2023). Comparative analysis of learner-sourced human-graded and AI-generated responses for autograding online tutor lessons. In G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, O. C. Santos, & N. Wang (Eds.), *Artificial intelligence in education. Posters and late breaking results, workshops and tutorials, industry and innovation tracks, practitioners, doctoral consortium and blue sky, AIED 2023* (Vol. 1831, pp. 714–719). Springer.
- Tonja, A. L., Dossou, B. F., Ojo, J., Rajab, J., Thior, F., Wairagala, E. P., & Rosman, B. (2024). InkubaLM: A small language model for low-resource African languages. *arXiv*, arXiv:2408.17024. [CrossRef]
- Tsimba, G., Mugoniwa, B., & Mutembedza, A. N. (2022, May 16–20). *Equitable access to eLearning during COVID-19 pandemic and beyond: A comparative analysis between rural and urban schools in Zimbabwe*. 2022 IST-Africa Conference (IST-Africa) (pp. 1–8), Virtual. [CrossRef]
- Ujjain, S., Akash, S. V., Dhanush, I., & Thilak Kumar, V. (2025, March 27–28). *Enhancing education infrastructure through machine learning: A framework for classifying and prioritizing schools*. 2025 International Conference on Frontier Technologies and Solutions (ICFTS) (pp. 1–5), Chennai, India. [CrossRef]
- UNESCO. (2021). *Teacher readiness for digital learning: Global trends*. Available online: <https://unesdoc.unesco.org/ark:/48223/pf0000376702> (accessed on 1 August 2025).
- United Nations. (2025). *Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all*. United Nations sustainable development goals. Available online: <https://sdgs.un.org/goals/goal4> (accessed on 31 August 2025).
- Vashee, K. (2019). *Understanding MT quality: BLEU scores*. RWS Blog. Available online: <https://www.rws.com/blog/understanding-mt-quality-bleu-scores/> (accessed on 27 July 2025).
- Vega, K. A. G., & Tejeda, S. (2023, October 23–27). *Exploring the transformative potential of machine learning in engineering and STEM education: A comprehensive study on enhancing cognitive levels and learning insights*. 2023 World Engineering Education Forum—Global Engineering Deans Council (WEEF-GEDC) (pp. 1–9), Monterrey, Mexico. [CrossRef]
- Wang, E. C., Guan, X., Chen, X., & Ng, T. K. (2025). Assessing mindset states of secondary students using machine learning. *Computers & Education*, 2025, 105554.
- Wikipedia Contributors. (n.d.). *Xhosa language*. Wikipedia. Available online: https://en.wikipedia.org/wiki/Xhosa_language (accessed on 1 August 2025).
- Wikipedia Contributors. (2020). *English Wikipedia surpasses six million articles*. Wikipedia. Available online: https://en.wikipedia.org/wiki/English_Wikipedia (accessed on 1 August 2025).
- Yang, X., Ma, Y., & Pei, P. (2024, April 26–27). *English translation system assisted by machine learning and cloud computing data aggregation algorithm*. 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE) (pp. 1–5), Ballari, India. [CrossRef]
- Yilmaz, K., & Deniz, K. Z. (2024). Natural language processing and machine learning applications for assessment and evaluation in education. *Journal of Measurement and Evaluation in Education and Psychology*, 15(4), 421–445.
- Zawacki-Richter, O., & Loglo, F. S. (2023). Learning with digital media: A systematic review of students' use in African higher education. *Journal of Learning for Development*, 10, 1–23. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.