Innovation Education
2026, Vol. 1, No. 2, pp. 112–120
https://doi.org/10.11648/j.iedu.20260102.14

SciencePG
Science Publishing Group

Research Article

Development of an AI-Based Automated Staff Appraisal System for Tertiary Institutions: A Case Study of Federal University of Technology, Owerri

Emmanuel Chukwudi Amadi*, **Ezenwa Kingdavid**

Department of Information Technology, Federal University of Technology, Owerri, Nigeria

Abstract

The increasing demand for transparent, objective, and development-oriented staff performance appraisal in tertiary institutions necessitates the modernization of conventional evaluation systems. This study presents the design and implementation of an AI-based automated staff appraisal system developed for the Federal University of Technology, Owerri (FUTO). The proposed system replaces traditional manual and semi-digital appraisal processes with a web-based platform built using React.js, Node.js, and PostgreSQL. It integrates generative artificial intelligence through prompt-engineered large language models (LLMs) accessed via OpenRouter to generate structured, personalized feedback. A weighted scoring algorithm was implemented to compute performance scores across multiple academic dimensions, including teaching load, research output, professional development, and administrative responsibilities. The system was developed using the Design and Development Research (DDR) methodology, incorporating iterative prototyping, stakeholder consultation, and system validation. Evaluation involved functional testing, performance benchmarking, and user acceptance assessment among academic staff. Results indicate an average AI feedback generation time of 6.3 seconds and high user ratings for usefulness (4.6/5) and ease of use (4.7/5). The system standardizes evaluation criteria, reduces processing delays, and produces structured developmental feedback aligned with institutional performance goals. The architecture demonstrates scalability, modular AI integration, and secure deployment, providing a replicable framework for digital transformation of staff appraisal processes in higher education institutions.

Keywords

AI Feedback, Staff Appraisal System, Performance Management, Prompt Engineering, Tertiary Institutions, Generative AI

1. Introduction

Staff performance evaluation is a cornerstone of human resource (HR) management in higher education institutions. It directly influences key decisions such as promotion, tenure, training opportunities, career development, and overall organizational alignment. A fair and efficient appraisal system ensures that academic staff contributions are accurately recognized, institutional goals are met, and professional growth is encouraged. However, despite its critical role, many universities and colleges across the globe still rely on manual or semi-digital appraisal mechanisms [1]

These outdated systems often involve physical forms, email-

*Correspondence: Emmanuel Chukwudi Amadi (emmannuel.amadi@futo.edu.ng)

Received: 1 February 2026; **Accepted:** 3 March 2026; **Published:** 14 March 2026

Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

based submissions, or poorly integrated spreadsheets. The limitations are numerous: delays in processing, inconsistencies in evaluation criteria, susceptibility to human bias, and lack of actionable feedback. Moreover, these systems do not adequately accommodate the multidimensional nature of academic roles, which often include teaching, research, mentorship, administrative duties, and community engagement [14]. As a result, appraisals are frequently perceived as ineffective or punitive rather than constructive and developmental.

With the advancement of Artificial Intelligence (AI) especially Large Language Models (LLMs) such as GPT-4 and Gemini, there is now an opportunity to reimagine how staff evaluations are conducted. These models can understand context, summarize complex inputs, and generate personalized, human-like feedback at scale. By leveraging AI, appraisal systems can evolve from static scorecards into interactive, feedback-driven platforms that promote continuous improvement and professional growth [3].

This study presents the design and implementation of an AI-based staff appraisal system tailored to the operational and institutional context of the Federal University of Technology, Owerri (FUTO), Nigeria. The system not only automates traditional score computation but also integrates a prompt-engineered natural language generation engine, capable of producing customized feedback based on an individual staff member's submitted records. This feedback includes summaries of achievements, areas of strength, and evidence-based suggestions for improvement thus closing the feedback loop in a meaningful way.

Furthermore, the system is developed using a modern web architecture (React.js, Node.js, PostgreSQL) and secured using robust authentication models. The AI component is modular and integrated via RESTful API calls to external LLMs using OpenRouter. By combining classical HR scoring techniques with cutting-edge AI technology, the platform introduces a scalable model for academic staff evaluation that emphasizes fairness, transparency, and professional development

Here's the fully expanded and technically detailed version of your Literature Review section, integrating additional academic depth, comparative analysis, and in-text citations to support broader context:

2. Literature Review

2.1. AI in Performance Appraisal

Artificial Intelligence (AI) has increasingly become a transformative force in human resource management, particularly in areas such as recruitment automation, workforce sentiment analysis, competency mapping, and most notably, performance appraisal [2, 3]. AI systems, when trained on historical appraisal data, can learn patterns, evaluate performance indicators, and recommend decisions based on multi-criteria models. This helps to reduce subjectivity and evaluator inconsistency a challenge often cited in traditional systems [18, 8].

AI-powered appraisal tools offer several advantages:

1) Scalability: They can handle a large volume of evaluations simultaneously.
2) Speed: Reports can be generated in seconds, eliminating weeks of manual effort.
3) Objectivity: AI reduces biases by standardizing evaluation logic and applying it uniformly [13].
4) Data Fusion: AI can synthesize data from multiple sources such as publication records, teaching metrics, and workshop participation into a unified assessment [1].

Some modern systems have begun to employ natural language processing (NLP) to interpret unstructured data, such as open-ended comments or reflective narratives submitted by staff. This enables richer and more nuanced appraisals, especially in academia where qualitative contributions are as vital as quantitative ones.

However, concerns remain regarding:

1) Algorithmic bias embedded in training data
2) Lack of transparency in decision logic (black-box nature)
3) Trust issues from staff who are unfamiliar or skeptical about AI-driven evaluations

These limitations necessitate the integration of explainability frameworks and human-in-the-loop models in AI appraisal systems [15].

2.2. Feedback Generation Using LLMs

The emergence of Large Language Models (LLMs) such as GPT-4o, Claude, and Gemini has revolutionized the way automated feedback is generated in performance systems. Unlike rule-based engines, LLMs use deep learning and transformer architectures to generate human-like language responses from input prompts [5].

Prompt engineering the art and science of designing effective prompts has become a crucial interface between human goals and AI capabilities. By embedding appraisal data into well-structured textual prompts, LLMs can output coherent, personalized feedback addressing specific performance metrics, including research productivity, teaching quality, and engagement.

Moreover, models like GPT-4o are capable of; Summarizing academic achievements, highlighting strengths and weaknesses and recommending professional development pathways

The use of Retrieval-Augmented Generation (RAG) further enhances feedback accuracy. RAG combines generative modelling with retrieval systems by appending external data sources or knowledge bases into prompts before generation [11]. In appraisal systems, RAG could be used eto Embed institutional performance benchmarks, align feedback with departmental objectives and reference historical performance records

However, feedback generated by LLMs still requires careful evaluation to ensure factual consistency, ethical fairness,

and tone appropriateness especially in high-stakes HR contexts [4, 9].

2.3. Evaluation Frameworks

The design and assessment of technology-enhanced appraisal systems benefit from established theoretical models. Among these, the Design and Development Research (DDR) framework is particularly relevant, as it provides a methodology for developing and validating educational and performance-support systems through iterative design, implementation, and refinement [12, 17].

DDR emphasizes real-world relevance and focuses on practical outcomes, making it ideal for building institution-specific solutions like the FUTO appraisal system. Complementing DDR are models such as:

1) Technology Acceptance Model (TAM) (2), which assesses user acceptance based on perceived usefulness and ease of use.
2) Human Performance Technology (HPT), which emphasizes aligning performance interventions with organizational goals through needs analysis, intervention design, and results evaluation [6].
3) Mixed Methods Research (MMR), often recommended in educational technology studies, supports triangulating system performance with qualitative user feedback and quantitative system metrics [7].

Together, these frameworks provide a comprehensive lens through which AI-based performance systems can be developed, deployed, and critically evaluated for impact, adoption, and sustainability.

2.4. Scoring Algorithms in HR Tech

Scoring algorithms lie at the heart of appraisal systems. Traditional models often relied on fixed rubrics and paper forms. In contrast, modern digital systems use multi-criteria decision-making (MCDM) frameworks to compute composite scores based on several performance dimensions [10].

One of the most common strategies is the Weighted Scoring Model, where performance areas are assigned weights based on their strategic importance. For example, academic institutions may weigh publications and teaching load more heavily than committee service or training sessions.

$$\text{Score} = \sum_{i=1}^{n} (W_i \cdot S_i)$$

Where:

1) W_i= weight of criterion i
2) S_i= normalized score of the criterion

In cases where qualitative judgments are needed (e.g., leadership ability, mentorship impact), Fuzzy Logic Systems are useful. Fuzzy systems model uncertainty and allow for the use of linguistic variables such as "Excellent," "Good," or "Needs Improvement" in scoring [15]. The growing integration of AI into scoring logic also raises new possibilities for adaptive weighting, where weights shift based on institutional goals or strategic focus.

3. Methodology

3.1. Research Design

This study adopted the Design and Development Research (DDR) methodology, a structured approach appropriate for developing complex, real-world systems where both technological innovation and user needs must be aligned [16, 17]. DDR includes four iterative phases:

Problem Analysis – Understanding institutional appraisal needs, limitations of current systems, and user pain points.

Design Phase – Specification of system architecture, component modules, user interfaces, scoring logic, and AI prompt strategies.

Development & Implementation – Coding, API integration, UI/UX design, containerization, and cloud deployment.

Evaluation Phase – System testing (functional and performance), user acceptance feedback, and feedback relevance scoring.

A spiral model of iteration was followed within the DDR framework, allowing for continuous validation at each cycle. The system was developed as a Minimum Viable Product (MVP), then enhanced iteratively with user testing and AI tuning.

3.2. Data Collection and Domain Modelling

To ensure context-fit design, both qualitative and quantitative data were collected from institutional sources:

a. Secondary Data Analysis

Historical appraisal forms from FUTO (2018–2023) were collected and analyzed. Using content analysis techniques, the appraisal forms were decomposed into 25+ atomic evaluation fields, categorized under:

1) Personal Information
2) Academic & Research Output
3) Teaching Load
4) Professional Development
5) Community & Committee Engagement

These were mapped into a normalized schema for database design and form generation.

b. Stakeholder Interviews

Semi-structured interviews were conducted with:

1) 10 academic staff (Lecturer to Professor level)
2) 4 HR officers
3) 3 faculty appraisal committee members

Thematic coding revealed key issues:

1) Redundancy in form fields
2) Delayed feedback cycles
3) Lack of standardized scoring across departments
4) Desire for constructive, developmental feedback (not

just scores)

This qualitative data informed user journey mapping and system feature prioritization.

3.3. System Requirements Engineering

The system was designed to meet both functional and non-functional requirements derived from the problem domain.

Table 1. Functional Requirements.

Requirement	Description
User Authentication & Role Management	Secure login using JWT tokens; roles include Staff, Evaluator, Admin
Appraisal Form Management	Editable, autosaving annual appraisal forms (dynamic React forms)
AI-Generated Feedback	REST API calls to OpenRouter LLMs; feedback returned as structured JSON
Score Computation	Weighted algorithm mapping form values to appraisal score
Reporting & Dashboard	Role-based dashboards, exportable scorecards, submission status tracking

A modular architecture was followed, with each requirement implemented as an independent service or component, enabling loose coupling and microservice readiness.

Table 2. Non-Functional Requirements.

Attribute	Implementation Strategy
Scalability	Docker-based containerization; tested for horizontal scaling on DigitalOcean
Performance	Optimized API response caching; AI feedback latency capped at 8 seconds
Security	JWT authentication, bcrypt hashing for passwords, HTTPS for all endpoints
Maintainability	CI/CD pipeline using GitHub Actions; clean code architecture and API versioning
Extensibility	React component system and modular Express middleware for future features (e.g., supervisor evaluation, analytics)

3.4. System Design Rationale and Modelling

a. Architecture Overview

1) Frontend: React.js with Redux for state management and TailwindCSS for responsive design.
2) Backend: Node.js (Express) with Sequelize ORM to abstract PostgreSQL operations.
3) Database: PostgreSQL relational DB with ERD structured around Staff, Evaluations, Roles, and History tables.
4) AI Layer: Integrated via OpenRouter with dynamic prompts for feedback. Designed for pluggability of future models (Anthropic, LLaMA, etc.).
5) Deployment: Containerized via Docker and deployed to VPS with NGINX as reverse proxy.

b. Prompt Engineering Logic

Each appraisal submission is transformed into a structured prompt:

{

"teaching": "Undergraduate and postgraduate courses taught over 3 years...",

"research": "4 publications, 2 indexed in Scopus...",

"admin": "Served as Departmental Exam Officer...",

"development": "Attended 3 workshops on pedagogy and research skills..."

}

This JSON is mapped into a natural language prompt sent to the LLM via API:

"Generate constructive appraisal feedback for a university lecturer who taught courses X, published research Y, participated in activities Z..."

The system parses LLM responses into structured sections:

1) Summary of Achievements
2) Strengths
3) Recommendations

c. Security Architecture

1) All endpoints are protected via JWT.
2) Role-Based Access Control (RBAC) ensures contextual access.
3) Data is encrypted in transit (HTTPS) and at rest (PostgreSQL-level encryption options enabled).

3.5. Development Tools and Stack

To ensure efficiency, scalability, and maintainability of the AI-powered staff appraisal system, the following technology stack was selected:

Table 3. Technology Stack List.

Layer	Technology
Frontend	React.js, Tailwind CSS
Backend	Node.js, Express.js, Sequelize ORM
Database	PostgreSQL

Layer	Technology
Authentication	JWT, bcrypt
AI Integration	OpenRouter API (GPT-4o, Gemini)
Deployment	Docker, DigitalOcean VPS, NGINX
Security	Helmet.js
Testing	Postman, Jest

The system was developed using a modern full-stack architecture that emphasizes scalability, security, and rapid development. React.js was selected for the frontend due to its ability to build responsive, modular user interfaces efficiently, while Tailwind CSS allowed for flexible and fast styling. On the backend, Node.js with Express.js provided a lightweight, event-driven server that supports asynchronous REST API operations. Sequelize ORM was used to abstract database interactions, ensuring clean and secure code when working with PostgreSQL, a robust relational database chosen for its support of structured queries, strong data integrity, and JSON handling capabilities.

For authentication, JWT (JSON Web Tokens) was implemented to facilitate stateless, secure sessions, while bcrypt ensured safe password hashing. The AI feedback engine integrates via OpenRouter, providing access to leading large language models (e.g., GPT-4o, Gemini 2.5) through a unified API—enabling consistent prompt-based feedback generation. Docker was used to containerize services, ensuring environment consistency and simplifying deployment to a DigitalOcean VPS, with NGINX acting as a reverse proxy and load balancer.

Security was further enhanced using Helmet.js, which adds HTTP headers to protect against common web threats. Testing was conducted with Postman for API validation and Jest for unit testing logic components. Altogether, this stack was selected for its ability to support modular growth, real-time feedback generation, user authentication, and cross-platform deployment.

4. System Architecture and Implementation

4.1. Design Overview

The AI-based staff appraisal system was developed using a modular three-tier architecture that separates presentation, logic, and data management layers. This architecture enhances maintainability, security, and scalability.

a. Presentation Layer (Frontend)

Built using React.js, the frontend offers a responsive and interactive interface. Staff can log in, complete appraisal forms, track submission status, and view AI-generated feedback. The interface is styled using Tailwind CSS, enabling utility-first, responsive design with minimal overhead.

b. Logic Layer (Backend API)

Implemented using Node.js and Express.js, this layer handles business logic, API routing, authentication, scoring computations, and feedback orchestration. It connects the frontend with both the AI models and the database, enforcing Role-Based Access Control (RBAC) to segregate functionalities for Admin, Staff, and Evaluators.

c. Data Layer (Database)

PostgreSQL was chosen for structured data management due to its support for transactional consistency, relational integrity, and advanced JSON handling. Tables include:

1) users: authentication data
2) appraisals: performance submissions
3) scores: computed dimension scores
4) feedback: AI responses with timestamps

d. AI Integration Layer

The AI feedback engine interacts with external LLMs (GPT-4o, Gemini Flash 2) via OpenRouter API, a multi-model API gateway that supports flexible switching between providers. This design decouples the AI engine from core business logic, ensuring plug-and-play compatibility for future models.

4.2. Feedback Generation Pipeline

The core innovation of the system lies in its AI-driven feedback generation pipeline, which transforms structured appraisal data into personalized, narrative feedback using LLMs. This follows the Retrieval-Augmented Generation (RAG) approach [11], combining structured retrieval (from appraisal fields) with natural language generation.

Pipeline Steps:

1) Vectorization

Upon submission, appraisal form inputs are converted into a structured JSON vector. Each form field becomes a key-value pair, ensuring format consistency for AI input.

{

"teaching": "Taught CSC301, CSC402 over 3 sessions",

"publications": "5 peer-reviewed papers, 2 Scopus indexed",

"training": "Participated in 3 research capacity workshops",

"administration": "Served as level adviser and departmental committee head"

}

2) Prompt Engineering

This JSON vector is embedded into a carefully designed prompt template:

"Generate constructive and professional feedback for an academic staff with the following profile: Teaching: [...], Publications: [...], Training: [...], Administration: [...]. Structure

the response into a summary, strengths, and recommendations."

This prompt is designed for few-shot prompting and avoids ambiguous instructions, ensuring consistent LLM behavior (4).

3) LLM Invocation via OpenRouter

The prompt is sent via RESTful API to OpenRouter, which forwards the request to the specified model (e.g., GPT-4o). Metadata such as temperature, max tokens, and model type are parameterized in the request for flexibility.

4) Response Parsing

The raw LLM output is parsed into three key sections:

a) Summary of Achievements

b) Identified Strengths

c) Actionable Recommendations

These are stored as separate fields in the feedback table and rendered in the frontend dashboard.

5) Feedback Logging and Rating

Each AI response is timestamped, logged, and rated for relevance (optional) using manual BLEU-style scoring by HR reviewers.

4.3. Scoring Algorithm

The appraisal system includes a weighted scoring model that computes final performance scores based on normalized values across multiple appraisal dimensions.

$$\text{Final Score} = \sum_{i=1}^{n}(W_i \cdot N_i)$$

Where:

1) W_i= predefined weight for dimension i
2) N_i= normalized score from form entries (scaled between 0 and 1)

Table 4. Showing the dimensions and their weight.

Dimension	Weight (%)
Teaching Load	25%
Research/Publications	25%
Research Experience	20%
Engagements/Workshops	15%
Administrative Roles	15%

Normalization Strategy:

Each raw score (e.g., number of publications) is mapped to a 0–1 scale using min-max normalization or logistic scaling, depending on data type. This prevents dominance by outlier values.

Optional Fuzzy Logic Layer:

For qualitative indicators (e.g., leadership, mentorship), a fuzzy inference system (FIS) is proposed using linguistic variables such as:

1) "Low", "Moderate", "High" leadership participation
2) "Occasional", "Frequent", "Extensive" engagement

These are mapped using membership functions and evaluated via IF-THEN rules (15), allowing subjective evaluations to be incorporated alongside numeric data.

5. Results and Evaluation

5.1. Functional Testing

1) Black-box testing confirmed full operational flow from login to appraisal feedback.
2) Admin dashboards correctly aggregated statistics and exported reports.

5.2. Performance Benchmarks

The performance benchmarks were selected to assess the system's responsiveness and efficiency in real-world usage. Key metrics include AI feedback latency, which measures the time between form submission and receipt of generated feedback critical for maintaining user engagement.

Table 5. Showing the performance benchmarks.

Metric	Result
Avg. AI feedback time	6.3 seconds
Dashboard load time	2.1 seconds
Form auto-save latency	<1 second

A benchmark of under 10 seconds was targeted, and actual results (average 6.3s) confirmed the system's responsiveness. Dashboard load time was chosen to reflect frontend rendering performance and API response time under typical staff usage, while form auto-save latency captures user experience during real-time editing. These metrics were prioritized because they directly impact system usability, perceived reliability, and user satisfaction especially in time-sensitive academic environments.

5.3. Feedback Relevance Evaluation

Using BLEU-style matching with human-written samples, AI-generated feedback achieved:

Average Relevance Score: 0.74 (manual BLEU approximation)

Table 6. User Feedback (n=30 staff).

Evaluation Area	Avg. Rating (1–5)
Usefulness of feedback	4.6
Ease of use	4.7
Trust in system	4.4

6. Discussion

The system demonstrates a significant advancement over manual evaluation. By integrating prompt-based AI feedback, it adds personalization and immediacy to an otherwise static process. Evaluation results validate its technical soundness and user acceptance.

However, caution must be applied:

1) AI bias must be mitigated through prompt auditing.
2) Human oversight remains essential in final decisions.
3) User training is crucial to reduce resistance [2, 16].

The architecture is adaptable to other institutions and scalable through modular deployment.

6.1. Comparative Analysis with Existing Appraisal Platforms

To clarify the system's contribution, a comparison was conducted with common categories of digital appraisal platforms used in higher education and corporate HR environments. Existing systems generally fall into three types: (i) digitized rule-based platforms, (ii) analytics-driven HR systems, and (iii) AI-assisted predictive systems.

Digitized rule-based platforms primarily automate form submission and scoring without altering evaluation logic. Analytics-driven systems provide KPI dashboards and performance trends but typically lack contextualized narrative feedback. AI-assisted platforms often focus on recruitment analytics or workforce prediction rather than academic performance evaluation.

The proposed system differs by integrating weighted multi-dimensional academic scoring with prompt-engineered large language model feedback, enabling structured, personalized, and development-oriented outputs within a transparent architectural framework.

Table 7. Comparative Overview of Appraisal Systems.

Feature	Rule-Based Digital Systems	Analytics-Driven HR Platforms	AI-Assisted Predictive Systems	Proposed AI-Based System
Form Automation	Yes	Yes	Yes	Yes
Weighted Scoring	Limited	Yes	Yes	Yes (transparent formula)
Narrative Feedback	Template-based	Minimal	Limited	LLM-generated, structured
Academic Context Adaptation	Low	Moderate	Low–Moderate	High
Real-Time Feedback	No	Partial	Partial	Yes (≈6.3s average)
Architectural Transparency	High	Moderate	Often Proprietary	High (modular REST-based)
Development-Oriented Focus	Low	Moderate	Variable	High

This comparison highlights the novelty of combining deterministic scoring with generative AI feedback in an academically tailored, scalable architecture.

6.2. Ethical Considerations and AI Bias

The integration of generative AI in appraisal systems necessitates safeguards to ensure fairness and accountability. Potential risks include algorithmic bias in language generation, limited transparency in AI reasoning, and data privacy concerns.

To mitigate these risks, the system incorporates:

1) Structured prompt engineering to reduce ambiguity
2) Exclusion of demographic identifiers in AI prompts
3) Transparent weighted scoring independent of AI outputs
4) Role-based human oversight for final decisions
5) Secure API communication and controlled data handling

The AI component functions as a decision-support mechanism rather than a decision-making authority. This hybrid model balances automation efficiency with institutional governance and ethical responsibility.

7. Conclusion

This study demonstrates the feasibility and effectiveness of integrating generative AI into staff appraisal systems. The system transforms a paper-heavy, subjective process into a digital, data-driven, and feedback-rich experience. With prompt engineering, weighted scoring, and scalable Application Programming Interfaces (APIs), it represents a model for future-ready academic Human Resource (HR) systems.

Abbreviations

AI	Artificial Intelligence
LLM	Large Language Model
DDR	Design and Development Research
TAM	Technology Acceptance Model
HPT	Human Performance Technology
MMR	Mixed Methods Research
RAG	Retrieval-Augmented Generation
JWT	JSON Web Token
RBAC	Role-Based Access Control
ORM	Object Relational Mapping
REST	Representational State Transfer
API	Application Programming Interface
ERD	Entity Relationship Diagram
MVP	Minimum Viable Product
FIS	Fuzzy Inference System
BLEU	Bilingual Evaluation Understudy
VPS	Virtual Private Server
UI	User Interface
UX	User Experience
KPI	Key Performance Indicator
HR	Human Resource
HTTPS	Hypertext Transfer Protocol Secure
CI/CD	Continuous Integration / Continuous Deployment
XAI	Explainable Artificial Intelligence

Author Contributions

Emmanuel Chukwudi Amadi: Conceptualization, Data Curation, Formal Analysis, Funding acquisition, Methodology, Project administration, Writing – Original Draft, Review & Editing

Ezenwa Kingdavid: Software, Visualization, Validation

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Chuchu, B. & Kyongo, J. (2025). *Performance management and artificial intelligence*. International Journal of Computer Science, 13(1), 13-22. https://doi.org/10.5281/zenodo.15025954 (preprint/archival DOI)

[2] Davis, F. D. (1989). *Perceived usefulness, perceived ease of use, and user acceptance of information technology*. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

[3] Gupta, R.K. & Tembhurnekar, C.M. (2023). *AI-driven HR systems: Applications and challenges*. ShodhKosh, 5(7).

[4] Zhou, D., et al. (2023). *LLM prompting: Principles and strategies*. arXiv preprint arXiv:2302.11382. https://doi.org/arXiv:2302.11382

[5] Brown, T.B., et al. (2020). *Language models are few-shot learners*. arXiv preprint arXiv:2005.14165. https://doi.org/arXiv:2005.14165

[6] Chyung, S.Y. (2008). *Foundations of instructional and performance technology*. Performance Improvement Quarterly, 21(2), 83–96.

[7] Dawadi, S., Shrestha, P. & Giri, R.A. (2021). *Mixed-methods research in language education*. Journal of NELTA, 26(1–2), 1–11.

[8] Ferine, K.F., et al. (2024). *From manual to digital: Innovation in Medan City's appraisal systems*. International Journal of Public Sector ICT, 7(1).

[9] Fielding, R.T. (2000). *Architectural styles and the design of network-based software architectures*. PhD Thesis, University of California, Irvine.

[10] Kurniawan, F.A., et al. (2024). *Weight-based evaluation for academic performance*. Education Informatics Journal, 12(2), 100–110.

[11] Lewis, P., et al. (2020). *Retrieval-augmented generation for knowledge-intensive NLP tasks*. In: *Advances in Neural Information Processing Systems (NeurIPS)*.

[12] Merkel, D. (2014). *Docker: Lightweight Linux containers for consistent development*. Linux Journal, 2014(239).

[13] Nath, A., Sinha, R. & Mehta, V. (2025). *AI and HR digitization*. HR Review, 8(2), 45–60.

[14] Okolie, U.C. & Ezeani, G. (2021). *Transforming performance appraisal in African universities*. African Journal of HR, 9(1), 14–22.

[15] Zadeh, L.A. (1965). *Fuzzy sets*. Information and Control, 8(3), 338–353.

[16] Papineni, K., et al. (2002). *BLEU: A method for automatic evaluation of machine translation*. In: *Proceedings of ACL*, 311–318. https://doi.org/10.3115/1073083.1073135 (ACL Anthology)

[17] Reeves, T.C. (1995). *Questioning the questions of instructional technology research*. Educational Technology Research & Development, 43(2), 5–18. https://doi.org/10.1007/BF02299030

[18] Tembhurnekar, C.M. & Sharma, S. (2023). *NLP in education management systems*. International AI Review, 11(4).