International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

Applications of Artificial Intelligence and Deep Learning in Education

Somnath Kar[1], Priya Sahoo[2], Supriya Panda[3]

[1]University Department of Physics, Kolhan University, Chaibasa, Jharkhand, India
Corresponding Author Email: *1somnathkar11[at]gmail.com,*

[2]University Department of Computer Applications (MCA), Kolhan University, Chaibasa, Jharkhand, India
Email: *priya27.sahoo[at]gmil.com,*
Email: *pandasupriya55[at]gmail.com*

Abstract: ***Artificial Intelligence (AI) and Deep Learning are changing how teaching and learning happen. This paper reviews the literature on AI and Deep Learning in education and presents three technically explicit case studies: predicting at-risk students from Virtual Learning Environment (VLE) interaction data in the style of the Open University Learning Analytics Dataset (OULAD); tracing a student's evolving knowledge state with a recurrent neural network (Deep Knowledge Tracing, DKT); and scoring free-text answers with transformer language models (BERT-family Automated Essay Scoring). Each case study specifies an illustrative dataset schema, the model architecture used in the literature, and representative published performance figures. The paper also reviews the documented benefits of these systems (time saved for teachers, faster feedback, wider access), the challenges that limit them in practice (cost, the digital divide, data privacy, algorithmic bias, over-reliance), and the ethical principles- fairness, transparency, consent, and human oversight- that should govern their use. The overall finding is that AI and Deep Learning are most valuable in education when designed as decision-support tools for teachers rather than as replacements for them.***

Keywords: Artificial Intelligence, Deep Learning, Learning Analytics, Knowledge Tracing, Automated Essay Scoring

1. Introduction

Traditional classroom teaching follows a one-to-many model in which a single teacher delivers the same pace and sequence of instruction to an entire class. This works well for some students and poorly for others, because learners differ in prior knowledge, pace, attention span, and preferred mode of learning. Artificial Intelligence is the branch of computer science concerned with building systems that carry out tasks normally associated with human intelligence- understanding language, recognising patterns, and making decisions- while Deep Learning is the subfield of AI in which layered neural networks learn directly from large volumes of data rather than from rules written by a programmer [1], [2]. Applied to student data rather than, say, shopping or image data, these same techniques increasingly power learning platforms.

Much of the introductory literature on "AI in education" is written at a survey level: it lists application areas- personalised learning, chatbots, automated grading- without showing how a system in any one of those areas is actually built, what data it consumes, or how its performance is measured. That level of description is useful as orientation, but it does not tell a computer applications student what algorithm to implement, what a training dataset looks like, or how well such a system can be expected to perform. This paper addresses that gap by combining a literature review with three focused, technically explicit case studies, each centred on a well-documented technique: predicting students at risk of failing or withdrawing using supervised machine learning on VLE interaction data; tracing an individual student's evolving mastery of skills using a recurrent neural network (the DKT model); and scoring free-text answers and essays using transformer-based language models, benchmarked by Quadratic Weighted Kappa (QWK) agreement with human raters.

2. Literature Survey

The term "Artificial Intelligence" was coined in 1956 at the Dartmouth Conference, where researchers first proposed that machine intelligence could be studied as a formal discipline. Progress was slow for decades because both computing power and usable data were scarce. Two developments after 2010 changed this: the availability of very large datasets generated by the internet, and the use of graphics processing units (GPUs) to train neural networks efficiently, together enabling the rapid rise of Deep Learning and its subsequent spread into domains such as education [3].

Work on Intelligent Tutoring Systems (ITS) began in the 1980s and 1990s and proved that a computer program could adapt the sequence and difficulty of material to an individual learner, producing measurable gains over static instruction in controlled studies, using relatively simple rule-based or Bayesian student models [4]. A landmark later contribution was Deep Knowledge Tracing (DKT), which modelled a student's sequence of correct and incorrect answers with a recurrent neural network and reported substantially better prediction of future performance than the earlier Bayesian Knowledge Tracing approach, with later summaries citing roughly 25–30% relative AUC improvement over classical baselines on benchmarks such as ASSISTments and Khan Academy data [5]. In parallel, transformer-based language models such as BERT were adapted for automated essay scoring [6], and gradient-boosted and tree-based classifiers were shown to be strong baselines for predicting student dropout from clickstream data, using datasets such as the Open University Learning Analytics Dataset (OULAD) [7].

More recent work (2023–2026) has produced attention-based and memory-augmented knowledge-tracing models that outperform the original LSTM-based DKT on several

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

benchmarks, generative-AI tutors and copilots built on large language models, and a growing use of explainable-AI (XAI) techniques so that a model's risk score or grade is not a black box. Systematic reviews of this literature nonetheless identify recurring gaps: limited attention to students in low-connectivity settings, data-privacy and governance frameworks that have not kept pace with the volume of behavioural data now collected on students, published accuracy figures drawn mostly from benchmark datasets rather than long-running real-world deployments, and inconsistent reporting of model fairness across demographic subgroups [8]. This paper engages directly with these gaps: Section 5 is explicit about which figures come from public benchmark datasets versus which are illustrative, and Sections 5.6–5.7 treat the digital divide, privacy, and fairness gaps as first-class topics.

3. Problem Definition

Three research questions motivate this paper. RQ1: what categories of AI and Deep Learning techniques are currently used across the student lifecycle- from onboarding to assessment- and what published evidence supports their effectiveness? RQ2: for three representative techniques (dropout prediction, knowledge tracing, and automated essay scoring), what does a realistic input dataset look like, what is the model architecture, and what level of performance has been reported in the literature? RQ3: what benefits, limitations, and ethical safeguards should shape how an educational institution adopts these systems?

This paper is a literature-based study supplemented by illustrative technical case studies; it does not involve training new models on institutional data. Case-study datasets used below are either publicly documented benchmark datasets (described, not reproduced) or small illustrative samples constructed to match the schema of such datasets, clearly marked as illustrative where actual figures were not verified against a primary source. The study focuses on school- and college-level education, covering both classroom-based and online learning.

4. Methodology / Approach

A neural network is composed of small computational units ("neurons") arranged in layers. The input layer receives raw data- for example, the pixel values of a scanned answer sheet, or the words of a sentence. One or more hidden layers progressively combine these inputs into increasingly abstract features. The output layer produces the final prediction, such as a predicted grade or a correct/incorrect classification (Figure 1). The word "deep" refers to a network with many hidden layers; more layers generally allow the network to represent more complex relationships, at the cost of needing more data and computation to train [2].

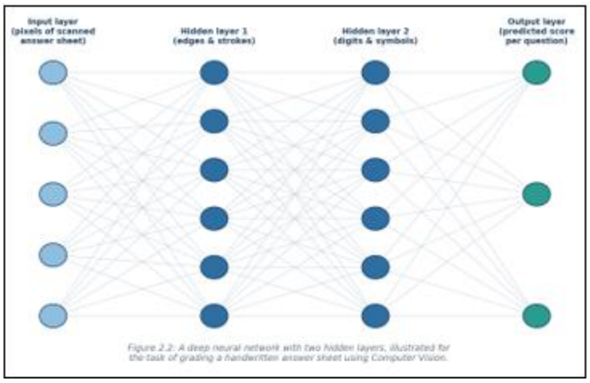

Figure 1: A layered neural network- input, hidden and output layers.

In conventional programming a developer writes explicit rules; in Machine Learning the developer instead supplies labelled examples, and the system infers the rules itself (Figure 2). The same principle underlies predicting whether a student is likely to drop out, or whether an essay deserves a particular grade: the system learns from many past, labelled examples rather than from hand-coded logic.

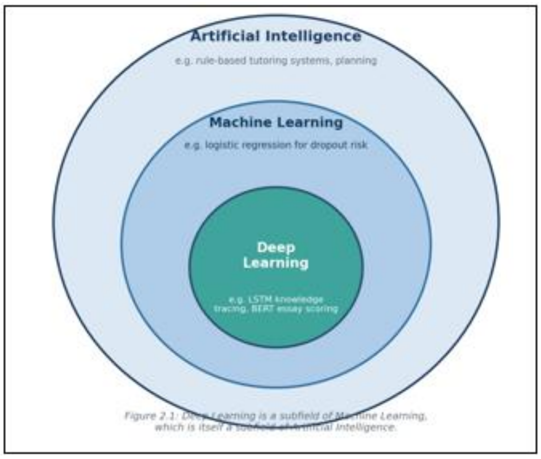

Figure 2: General machine-learning workflow- models learn statistical regularities from labelled examples.

Five broad technique families recur across educational applications: Natural Language Processing (NLP) for essay scoring and chatbots; Computer Vision for grading handwritten work and monitoring online exams; predictive/learning analytics for forecasting outcomes such as dropout; speech recognition for language learning and accessibility; and sequential models such as LSTMs and Transformers for modelling behaviour that unfolds over time, which underlies knowledge tracing. Table 1 summarises the mapping used in this study between application area and typical technique.

Table 1: Application areas and typical techniques

Application area	Typical technique(s)
Personalised / adaptive learning	Collaborative filtering; sequential models (DKT)
Intelligent Tutoring Systems	Bayesian student models; NLP answer understanding
Automatic grading	Transformers (essays); CNNs (handwriting)
Chatbots / virtual assistants	Retrieval-augmented / fine-tuned LLMs

Language learning	Speech recognition; spaced-repetition regression
Dropout / performance prediction	Logistic Regression; Random Forest; XGBoost
Proctoring / virtual classrooms	Computer Vision; anomaly detection

To ground this taxonomy, three techniques were selected for detailed case study on the basis of (a) having a well-documented public benchmark dataset, (b) a specific, named model architecture reported in the literature, and (c) published, literature-grounded performance figures: dropout prediction from VLE interaction data, Deep Knowledge Tracing, and transformer-based Automated Essay Scoring. For each, this paper follows the same three-part method: define an illustrative dataset schema based on the corresponding public benchmark, describe the modelling pipeline reported in the literature, and report representative performance figures drawn from published studies. Results are presented in Section 5.

5. Results and Discussion

Case Study A- Predicting At-Risk Students from VLE Data

The Open University Learning Analytics Dataset (OULAD) is a widely used, publicly documented benchmark that combines demographics, assessment scores, and daily VLE click summaries for 32,593 students across 22 course presentations [7]. It is used here only to illustrate schema and reported findings; Table 2 is an illustrative sample constructed to match that schema, not a reproduction of the actual dataset.

Table 2: Illustrative student records (schema after OULAD)

Student ID	VLE clicks (wk1–4)	Avg. score (wk1–4)	Outcome
S1001	512	78	Pass
S1002	96	41	Withdrawn
S1004	54	38	Fail
S1005	640	88	Distinction
S1007	42	29	Withdrawn

A typical pipeline extracts weekly aggregate features (submission timing, VLE click counts, running average score) and trains a supervised classifier to predict the final outcome, often collapsed to an at-risk vs not-at-risk binary label. Comparative studies on OULAD-style data commonly evaluate Logistic Regression, Support Vector Machines, Random Forest, Gradient Boosting, and XGBoost, reporting ROC-AUC as the primary metric because classes are imbalanced. Published comparative studies consistently rank ensemble tree methods above simpler linear models, with Gradient Boosting and XGBoost variants reported to reach ROC-AUC in roughly the 0.91–0.93 range [7], [8]. Figure 3 presents representative, illustrative values consistent with this reported ranking.

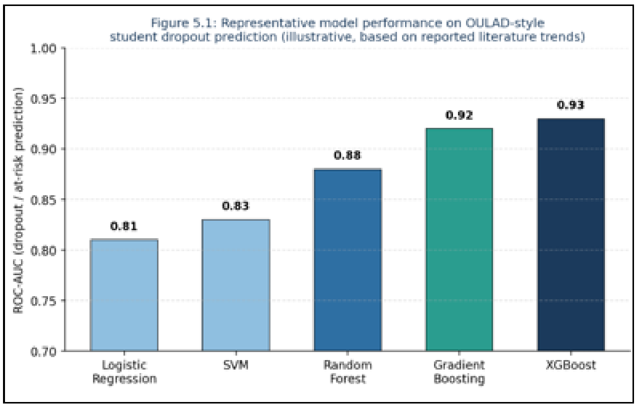

Figure 3: Illustrative comparison of classifier ROC-AUC for dropout/result prediction on OULAD-style data

A model like this is only useful if its output triggers a timely, human-led intervention- an advisor call, a nudge email, or peer mentoring- rather than being used punitively.

Case Study B- Deep Knowledge Tracing (DKT)

Where Case Study A predicts a single end-of-course outcome, DKT tracks a student's mastery of many individual skills continuously, updating its estimate after every question answered. DKT represents each student's interaction history as a sequence of (question, correct/incorrect) pairs and feeds that sequence into a recurrent neural network, typically an LSTM (Figure 4); the network's hidden state, updated at every step, is a compact summary of what the model currently believes about the student's knowledge across all tracked skills [5].

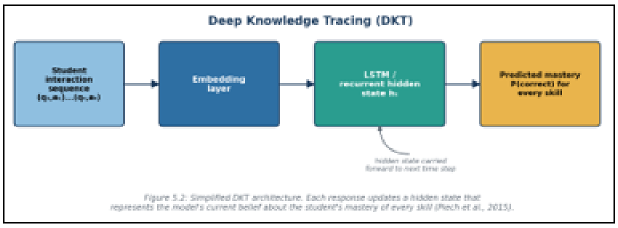

Figure 4: Deep Knowledge Tracing- a recurrent network updates a hidden knowledge-state after each answer

Table 3: Illustrative interaction sequence and evolving prediction

Step	Skill	Correct?	P (next fractions item)
1	Addition	1	0.52
2	Subtraction	1	0.58
3	Fractions	0	0.31
4	Fractions	1	0.44
5	Fractions	1	0.63

Piech et al. reported that DKT clearly outperformed the classical Bayesian Knowledge Tracing baseline on standard benchmarks such as ASSISTments 2009 and a Khan Academy dataset, with later summaries citing roughly 25–30% relative AUC improvement [5]. Because DKT infers relationships between skills automatically, it does not require an expert to hand-tag every question with the concept it tests. Subsequent variants- Dynamic Key-Value Memory Networks and Attentive/Self-Attentive Knowledge Tracing- use explicit memory or transformer-style attention and generally show further, incremental gains over the original LSTM-based DKT [8]. A trained knowledge-tracing model can drive an adaptive practice engine that presents, at every step, the question most likely to improve confidence in a weak skill.

Case Study C- Automated Essay Scoring (AES)
AES has moved through three generations: hand-crafted linguistic features with classical regression (1990s–2000s), recurrent/convolutional text encoders (2010s), and, most recently, pretrained transformer models such as BERT, RoBERTa and DeBERTa, fine-tuned on labelled essay datasets such as the Automated Student Assessment Prize (ASAP) corpus [6]. A transformer-based AES pipeline typically tokenises the essay, passes it through a pretrained encoder to obtain a contextual representation, and attaches a small regression/classification head to predict a score (Figure 5), compared against a human-assigned reference score using Quadratic Weighted Kappa (QWK), which penalises predictions more heavily the further they are from the true score.

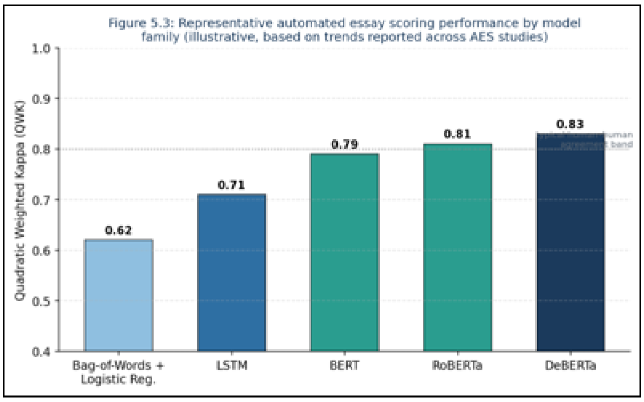

Figure 5: Transformer-based Automated Essay Scoring pipeline

Recent studies report that BERT-family models reach QWK in roughly the high-0.70s to low-0.80s range on standard AES benchmarks, and newer variants such as DeBERTa tend to edge out plain BERT and RoBERTa [9]. This performance is close enough to the QWK typically observed between two independent human raters that several studies describe fine-tuned transformer models as usable in a "second reader" role alongside a human grader rather than as a wholesale replacement [8]. Because grading is one of the most time-consuming parts of a teacher's job, even a second-reader system that flags disagreements can meaningfully reduce grading time while keeping a human in the loop.

Case Study D- Spaced Repetition (Duolingo)
Duolingo's Half-Life Regression (HLR) model estimates the "half-life" of a word or phrase in a learner's memory- the time after which the learner has roughly a 50% chance of still recalling it- using features such as word difficulty and past recall history. Trained on about 13 million learner-word practice records, HLR roughly halved the prediction error of the earlier Leitner and Pimsleur heuristics and improved next-day return rate in production A/B tests [10]. It is included here because it is one of the few AI-in-education systems with a large, published, real-world A/B test behind it, and because it shows that not every effective model needs to be a deep neural network.

6. Benefits

Knowledge tracing gives each student an individually sequenced learning path, something difficult for one teacher to sustain across a large class. Automating grading, attendance, and first-line question answering frees teacher time for mentoring and planning; industry analysis reports automated grading cutting essay-review time from roughly ten minutes to well under a minute per essay in some deployments, with a human still reviewing flagged cases. AI tutors and chatbots are available continuously, and automated grading/knowledge-tracing systems return feedback in seconds rather than days. Predictive early-warning systems let institutions direct advising resources toward students most likely to benefit, while text-to-speech, speech-to-text and captioning tools widen access for students with disabilities.

The scale of institutional investment is a useful, if imperfect, proxy for perceived value. Market-research estimates place the global AI-in-education market at USD 8.3 billion in 2025, growing to an estimated USD 11.4 billion in 2026 and a projected USD 57.2 billion by 2033 — a compound annual growth rate of roughly 25.9% [11]; other estimates report different absolute figures depending on scope but agree on the same broad direction of rapid, sustained growth [12], with adaptive assessment and grading cited as one of the fastest-growing segments (Figure 6).

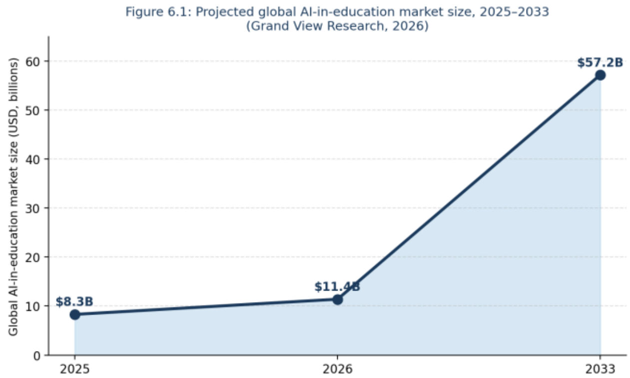

Figure 6: Reported growth trajectory of the global AI-in-education market, 2025–2033.

7. Challenges and Limitations

- High cost: building or licensing AI systems, and the infrastructure and staff training around them, is a real recurring cost that lower-resource institutions may struggle to sustain.
- Digital divide: predictive and adaptive systems assume a baseline of device access and connectivity; where that access is uneven, AI-driven personalisation risks widening rather than closing gaps between students.
- Data privacy and security: learning-analytics systems collect fine-grained behavioural data that must be properly secured and governed.
- Over-reliance: if students lean on AI-generated answers, or teachers defer entirely to a risk score without independent judgement, the tool substitutes for thinking rather than supporting it.
- Accuracy limits: none of the systems above is perfect — AUC below 1.0 and QWK below perfect agreement mean

a human review step remains necessary for high-stakes decisions.
- Bias: a model trained mostly on data from one demographic, language, or region can perform worse for students outside that distribution, and fairness metrics are inconsistently reported [8].
- Loss of human connection: teaching also involves motivation and relationship-building that current AI systems do not replicate.

8. Ethical Considerations

Systems should be tested across demographic subgroups before deployment, not only on an aggregate metric. Where a model produces a score or risk flag affecting a student, there should be an explainable basis for it. Institutions should obtain informed consent before collecting behavioural data and restrict its use to genuinely educational purposes. Consequential decisions- a final grade, academic probation, admission- should always retain a human decision-maker able to review and override a model's output, consistent with international guidance on the responsible use of AI in education [13]. Both staff and students need basic AI literacy so that AI functions as a support for learning rather than a shortcut around it.

9. Conclusion

Across the three case studies and the wider literature reviewed, the same pattern recurs: Deep-Learning-based methods (DKT, transformer-based AES, ensemble tree models for dropout prediction) consistently outperform their classical predecessors (Bayesian Knowledge Tracing, bag-of-words scoring, plain logistic regression) on standard benchmarks, but none reaches the reliability needed to remove a human reviewer from consequential decisions. The design principle recommended throughout this paper is to use AI to surface a well-informed first estimate — a risk flag, a predicted grade, a next practice question — and keep a teacher or advisor positioned to review, correct, and act on it. High costs, uneven device and connectivity access, privacy risk, and the possibility of biased or over-relied-upon models remain real constraints rather than incidental footnotes. Used carefully, transparently, and with clear human oversight, AI and Deep Learning have real, evidenced potential to make education more personalised, efficient, and inclusive.

10. Future Scope

- Finer-grained personalisation combining knowledge-tracing signals with affective/engagement signals.
- AI paired with VR/AR for immersive practice, such as simulated lab experiments.
- More natural, conversational tutors built on large language models with tighter grounding to reduce factual errors.
- Wider global access through lower-cost, lower-connectivity-tolerant tools for currently excluded schools.
- Lifelong-learning applications extending these techniques into adult upskilling and reskilling.
- Teacher-facing copilots for lesson planning and progress tracking, positioned as an assistant rather than a replacement.

References

[1] Russell, S., & Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th ed.). Pearson Education.
[2] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
[3] Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Center for Curriculum Redesign.
[4] Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence Unleashed: An Argument for AI in Education. Pearson.
[5] Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep Knowledge Tracing. Advances in Neural Information Processing Systems (NeurIPS) 28.
[6] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019.
[7] Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open University Learning Analytics dataset. Scientific Data, 4, 170171. https://doi.org/10.1038/sdata.2017.171
[8] Various authors. (2021–2026). Systematic reviews of machine learning and AI techniques and datasets in education, including studies on OULAD-based dropout prediction, ASSISTments/EdNet-based knowledge tracing, and BERT/DeBERTa-based automated essay scoring. IEEE Xplore, ACM Digital Library, SpringerLink, and arXiv.
[9] Cho, S., et al. (2024). Dual-scale BERT using multi-trait representations for holistic and trait-specific essay grading. ETRI Journal. https://doi.org/10.4218/etrij.2023-0324
[10] Settles, B., & Meeder, B. (2016). A Trainable Spaced Repetition Model for Language Learning. Proceedings of the 54th Annual Meeting of the ACL, 1848–1858.
[11] Grand View Research. (2026). AI in education market size, share and growth report, 2033. https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-education-market-report
[12] Precedence Research. (2026). AI in education market size to surge USD 136.79 Bn by 2035. https://www.precedenceresearch.com/ai-in-education-market
[13] UNESCO. (2021). Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development. UNESCO Publishing.

Author Profile

Dr. Somnath Kar is an Assistant Professor at University Department of Physics, Kolhan University, Chaibasa and also the Coordinator of University Department of MCA, Kolhan University, Chaibasa.

Priya Sahoo is pursuing the Master of Computer Applications (MCA) at Kolhan University, Chaibasa. This paper is drawn from her MCA thesis on Artificial

Intelligence and Deep Learning applications in education, prepared under the guidance of Dr. Somnath Kar

Supriya Panda is pursuing the Master of Computer Applications (MCA) at Kolhan University, Chaibasa. This paper is drawn from her MCA thesis on Artificial Intelligence and Deep Learning applications in education, prepared under the guidance of Dr. Somnath Kar.