REINFORCEMENT LEARNING IN EDUCATION 4.0: OPEN APPLICATIONS AND DEPLOYMENT CHALLENGES

Delali Kwasi Dake

Department of ICT Education, University of Education, Winneba, Ghana

ABSTRACT

Education 4.0 involves adopting technology in teaching and learning to drive innovation and growth across academic institutions. Artificial Intelligence and Machine Learning are frontrunners in Education 4.0, having already impacted diverse sectors globally. Since the COVID-19 pandemic, the conventional method of teaching and learning has become unpopular among institutions and is currently being replaced with intelligent educational data pattern identification and online learning. The teacher-centredpedagogical paradigm has significantly shifted to a learner-centredpedagogywith the emergenceof Education 4.0. Reinforcement Learning has been deployed successfully in diverse sectors, and the educational domain should not be an exemption. This survey discusses Reinforcement Learning, a feedback-based machine learning technique, with application modules in the academicfield. Each module is analysed for the state-action-reward implementation policies with relevant features that define individual use cases. The survey primarily examined the classroom, admission, e-learning, library and game development modules. In addition, the survey heightened the foreseeable challenges in the real-world deployment of Reinforcement Learning in educational institutions.

KEYWORDS

Education 4.0, Reinforcement Learning, Smart Campus, Intelligent Objects, Artificial Intelligence, Machine Learning

1. INTRODUCTION

Education which involves teaching and learning, has seen growth in varying pedagogy and technological integration in recent years[1]. As higher educational institutions (HEIs) prioritised a learner-centred approach[2, 3] to pedagogy and align institutional policies with the fourth industrial revolution (Industry 4.0), all disruptive technologies must be fully exploited. Disruptive technologies are emerging innovations with cutting-edge features to drive new markets and applications [4]. Artificial Intelligence (AI), a focal point of this review, is a key driver in the fourth industrial revolution, and its integration transcends diverse application domains, including education.

HEIs has seen a staggering increase in learner enrollment globally. The UNESCO (2020) report shows that student enrollment doubled in the last two decades, and the gross enrollment rate increased to 40% from 19% between 2000 and 2020. The quality of education, even as access to HEIs rises, needs to be protected. The high quality expected is usually seen when educational resources are sufficient, class sizes are minimal, learners' supervision is intensive, the student-teacher ratio is standard and governmental policies are progressive [5]. The struggle to achieve quality education amidst enrollment increases was exacerbated further by the COVID-19 pandemic[6].In March 2020, 1.5 billion students globally were affected by the temporal closure of schools (UNESCO, 2020). At the time of closure, COVID-19's policy measures to prevent the spread of the virus, including social and physical distancing, were strictly enforced [7].During the

pandemic's peak, in-person teaching and learning were discontinued, and most HEIs hastily migrated to eLearning platforms[8].The devastating pandemic exposed HEIs to technological and infrastructural deficits but provided an opportunity for HEIs to adopt proactive policies and integrate disruptive technologies in the educational ecosystem.

The following outlines guide the survey:

- A brief discussion of RL algorithms
- Open application of RL in education
- Deployment challenges of RL in education

2. ARTIFICIAL INTELLIGENCE, INDUSTRY 4.0 AND EDUCATION 4.0

Human intelligence forms the foundation for Artificial Intelligence (AI) and Machine Learning (ML), where algorithms perform intelligent tasks with features that mimic human cognitive functions[9, 10].While AI involves computers' ability to emulate human thought, ML uses algorithms to learn from data and automatically expose hidden patterns[11]. Industry 4.0 and Education 4.0 has AI and ML as primary technologies but with diverse application deployments. Industry 4.0 enables interconnected computers, intelligent machines, and smart objects to communicate and effect data-driven decisions with little or no human involvement[12, 13]. Industry 4.0 are cyber-physical systems with disruptive technologies characterised by higher productivity, efficiency, analytics and automation [13]. In the smart manufacturing value chain, the lifecycle of products and supply networks are vertically and horizontally integrated with modern control systems, sensors and technology.

Education 4.0, which is essential to this study, is closely related to Industry 4.0. Education 4.0 is a new paradigm which redefines teaching and learning to match the needs and technological advancement in Industry 4.0. The innovation-producing education, Education 4.0, transitioned from Education 1.0 as the download education to Education 2.0 as the open-access education and Education 3.0 as the knowledge-producing education[12, 14]. Education 4.0 contains 21st-century skills with collaboration, innovation, creativity, communication, critical thinking and problem-solving as learning strategies [15]. The deployment of technology, especially 3D printing, AI, Virtual and Augmented Reality (VAR) and the Internet of Things (IoT) in Education 4.0 will form the bases for the smart classrooms of the future. The AI aspect remains pivotal to this study, specifically ML in Education 4.0.

3. MACHINE LEARNING

In ML, there are three main categories of algorithms: Supervised, Unsupervised and Reinforcement Learning[16]. Supervised learning uses a labelled dataset where the predictive output of the training data is tagged with the correct answer [16, 17]. Successful prediction is only possible in supervised implementations when enough data is trained and tested for respective algorithms. In contrast, unsupervised learning uses unlabelled dataset to deduce clusters and expose hidden similarity patterns [16, 17].The clusters formed after unsupervised implementations become predictive labels for classification. Reinforcement Learning, the focal point of this review, involves actions exerted in an environment with rewards as guiding references to the agents[16 -18].

3.1. Reinforcement Learning

In Reinforcement Learning (RL) model, software agents play a central role in changing the state of the environment from s to s' using action a and receive a reward, r. As illustrated in Figure 1, the agent uses the trial and error principle with exploitation-exploration trade-off to select actions for future or immediate rewards [19, 20].

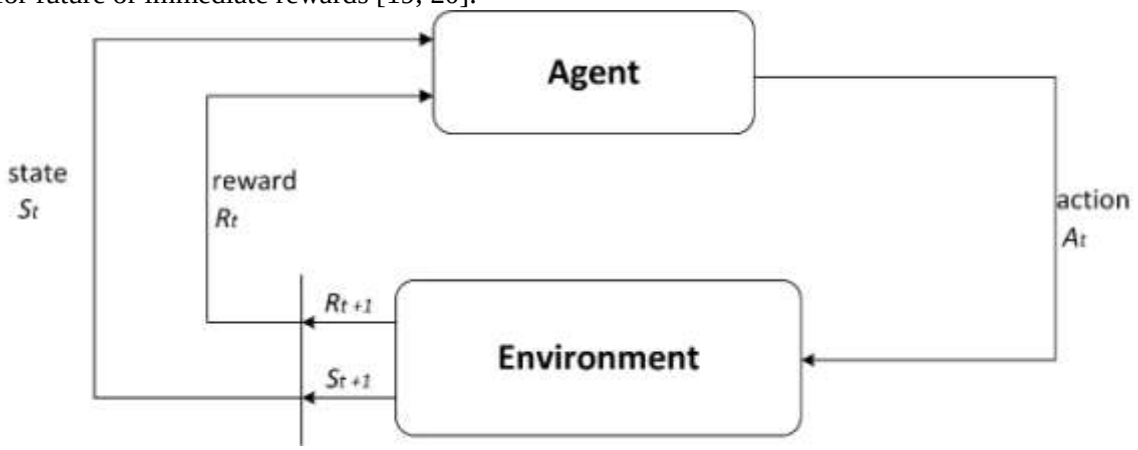

Figure 1. Reinforcement Learning [18]

In modelling the state-action-reward of a RL agent and its policy π, the Markov Decision Process (MDP) for decision-theoretic planning de facto standard formalism is utilised as shown in Algorithm 1 [21].

Algorithm 1 Markov Decision Process (MDP)	
An MDP is a 5-tuple (S, A, P, R, γ), where;	
	S is a set of states
	A is a set of actions
	$P\ (s, a, s')$ is the probability that action a in state s at time t will lead to state $s\text{^}'$ at time $t+1$
	$R\ (s, a, s')$ is the immediate reward received after a transition from state s to s', due to action a
	γ is the discounted factor which is used to generate a discounted reward

3.1.1. Reinforcement Learning Algorithms

RL algorithms are classified into two main classes: model-based (indirect) and model-free (direct) methods. Further, we have single and multi-agent RL algorithms based on the number of agents [22, 23]. Model-based algorithms maintain the agent's transition dynamics by utilising a model that influences future state, action, and reward. Model-based RL algorithms struggle to achieve asymptotic application performances but are data-efficient[24]. Model-free algorithms use trial and error for optimal actions without a model. Model-free algorithms are common with Q-learning and policy optimisation approaches [25]. Even though model-free methods are data-dependent, they have low computational complexity [25].

The single-agent RL (SARL), as depicted in Figure 2, has only one agent interacting in the environment. Q-learning are off-policy and value-based algorithm that improves its policy in a discrete action space. Unlike Q-learning, the State-Action-Reward-State-Action (SARSA) are on-policy algorithms that use the same policy for acting and updating the value function during training [26, 27].

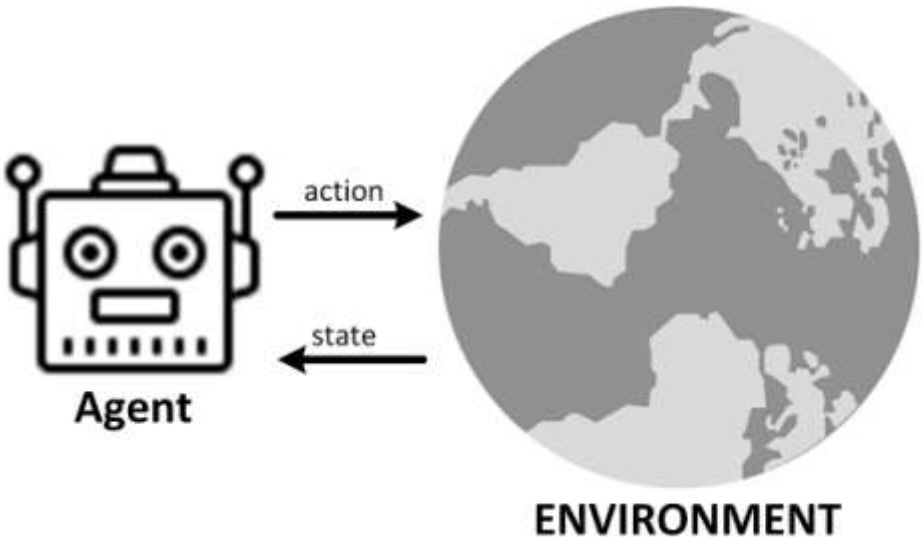

Figure 2. Single-agent RL

The Deep Q-Network (DQN) and the Deep Deterministic Policy Gradient (DDPG) SARL algorithms have become increasingly prominent due to artificial neural networks. The DQN is a combination of Q-learning and deep neural networks for discrete action space; whiles DDPG is for continuous action space [26, 27].

The multi-agent RL (MARL), as shown in Figure 3, has multiple agents interacting with the environment. In a MARL environment, the agents can be either cooperative or competitive. Cooperative agents help to improve group utility, whiles competitive agents are selfish [22, 23].

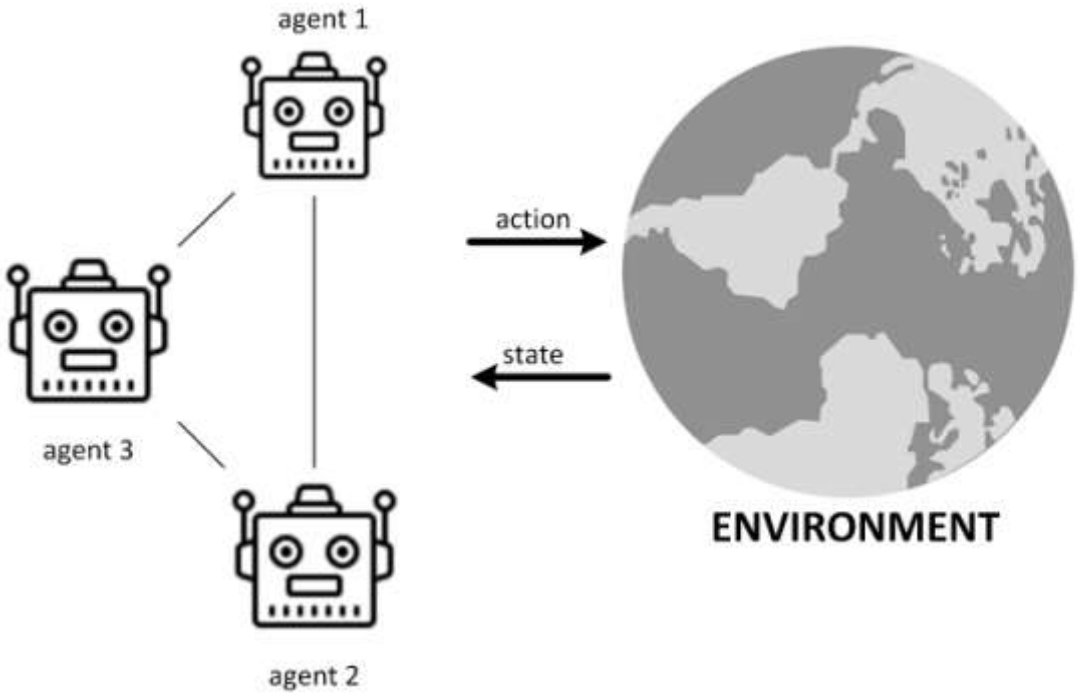

Figure 3. Multi-agent RL

The multi-agent Deep Deterministic Policy gradient (MADDPG) is the most dominant MARL algorithm that extends the DDPG algorithm for continuous action space [28]. The MADDPG is based on DDPG's actor-critic model, where the actor is the policy network and the critic the Q-value for training the actor network[29].

4. OPEN APPLICATION OF REINFORCEMENT LEARNING IN EDUCATION

This section focused on the relevant application integration of RL in education. The objective is to propose each module and examine relevant functionalities.

4.1. RL- Based Robots in Classroom

The traditional classroom is an in-person mode of teaching and learning where the instructor regulates the flow of knowledge and information[30]. The traditional classroom has a fixed space, usually with a formal arrangement of tables and chairs. This conventional space improves active learning where students are attentive during lesson delivery with proactive feedback sessions and greater class engagements[31]. Even though the traditional classroom approach to teaching and learning is recently underutilised, especially with the devastating COVID-19 pandemic, its benefits are largely significant. One beneficial aspect of a conventional classroom is the growth of interpersonal relationships. Secondly, some relevant courses, including aspects of medicine, engineering, agriculture and mathematics, are best taught in a traditional classroom setting. RL-based physical robots embedded with RL algorithms has interesting modules in a conventional classroom. The robot is either a learner robot or an instructor robot that replaces the teacher in the classroom. As shown in Figure 4, the learner robotsare personal agents fitted in the classroom for continuous monitoring of student's performance and behaviour. The robot using RL algorithms understands the learner and provides responsive solutions with detailed, intelligent analytics for the instructor, parents and school authorities. In a classroom, the state which forms

the environment varies from one student to another. As the robots perceive the state of a student, an action is recommended by the robots. The implemented action leads the student to a new state and the robot has a reward. The reward can be positive or negative. A positive reward indicates that, the action of the robot changes the state of the learner to a better one. A negative reward is a punishment which indicates a poor response from the student after the action was implemented.

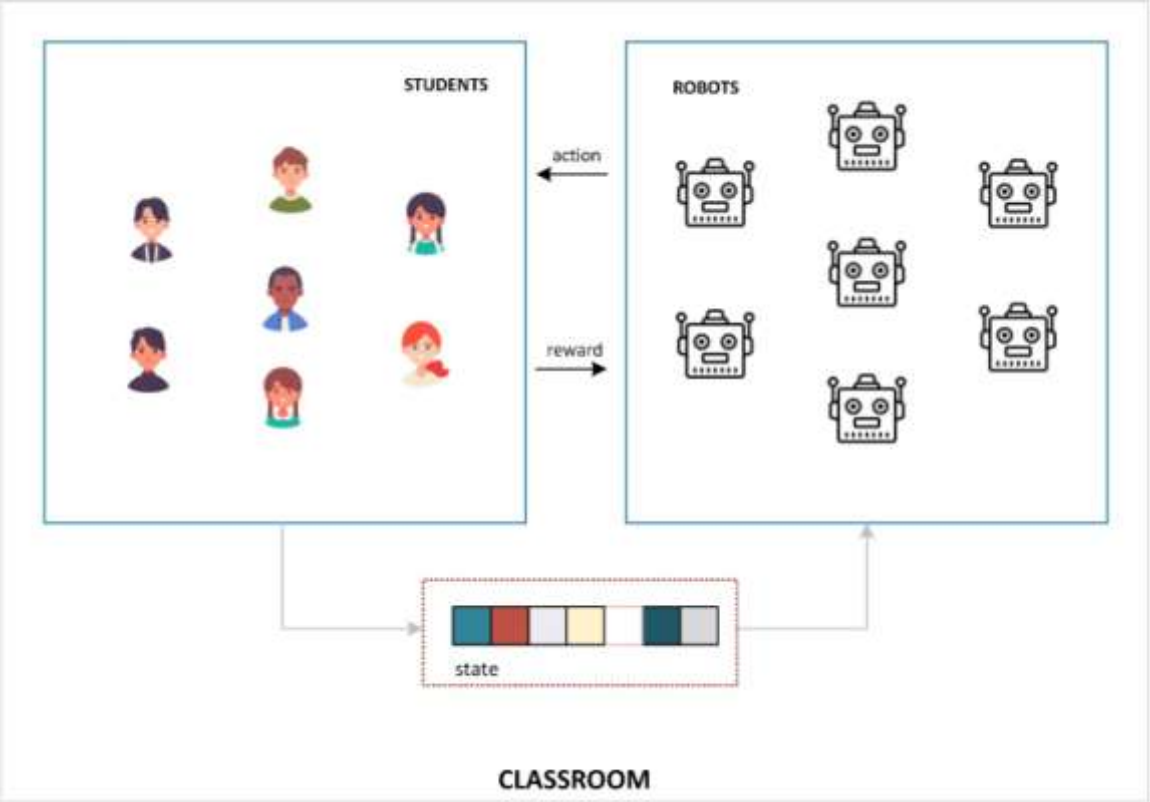

Figure 4. Learner robot in the classroom

Until the robot fully understands the learner using exploration and exploitation, the state of the student and the reward to the robot changes frequently. As depicted in equation 1, each student will have a state, s at every time slot t. A student's state in the classroom varies, as shown in Figure 5. Examples include sad, happy, angry, quiet, attentive, noisy, etc.

$$state\ space = \quad s_{1(t)}s_{2(t)}s_{3(t)} \ldots\ldots\ldots\ s_{n(t)} \qquad (1)$$

The action space determines the new state of the learner. In equation 2, the robot's action will change the student's state to a new one, s'. As illustrated in Figure 5, a robot's actions include video, animation, text, temperature, equation etc.

$$action\ space = \quad a_{1(t)}a_{2(t)}a_{3(t)} \ldots\ldots\ldots\ a_{n(t)} \qquad (2)$$

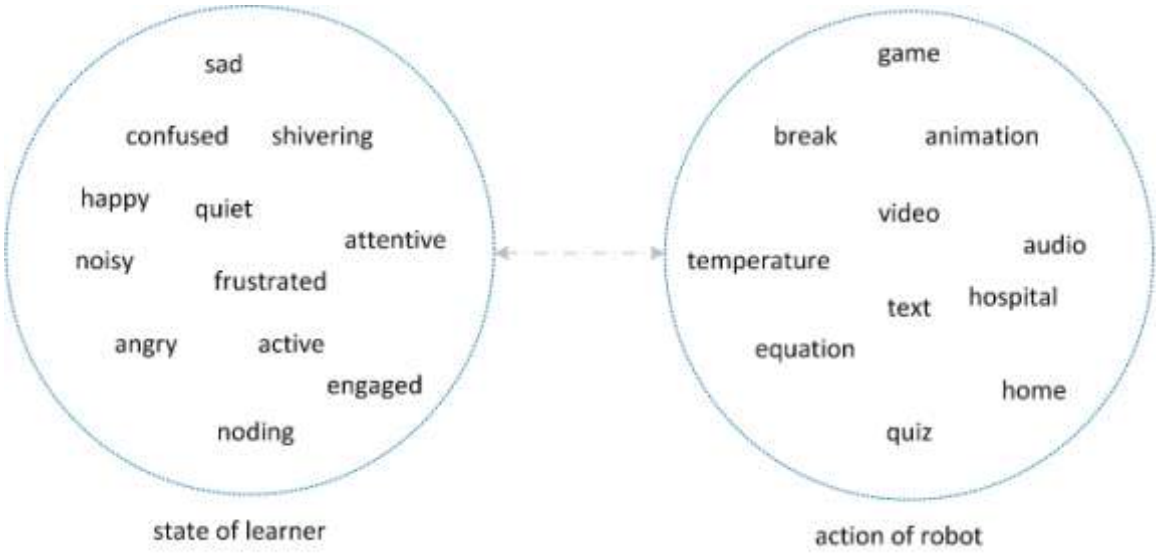

Figure 5. Examples of state-action activities

4.2. Educational Game Development for Students

Game development is one of the essential applications of Reinforcement Learning. The use of RL algorithms for game development gained popularity when DeepMind's AlphaGo-trained RL agent in 2016 defeated Lee Se-dol, the 18-times Go player champion in the world [32].In the educational context, Serious Games[33] are entertaining, engaging, and immersive, with embedded problem-solving design to improve students' learning strategies. The Serious Games modules have rewards based on challenges. In addition, such games support multiplayer participation, which enforces the concept of collaborative learning. Games played in a group are interesting, competitive, and engaging, accelerating the appreciation of specific educational content[34]. The primary attributes of single-agent and multi-agent RL algorithms, including state, action and reward, has influenced their usage in educational game programming[35].A RL technique in educational game development is even more gratifying since it uses unlabeled training data in its state-action-reward paradigm [35, 36].Dobrovsky et al.,[37] RL-based Serious Games framework, as shown in Figure 6, has fundamental elements in educational game development.

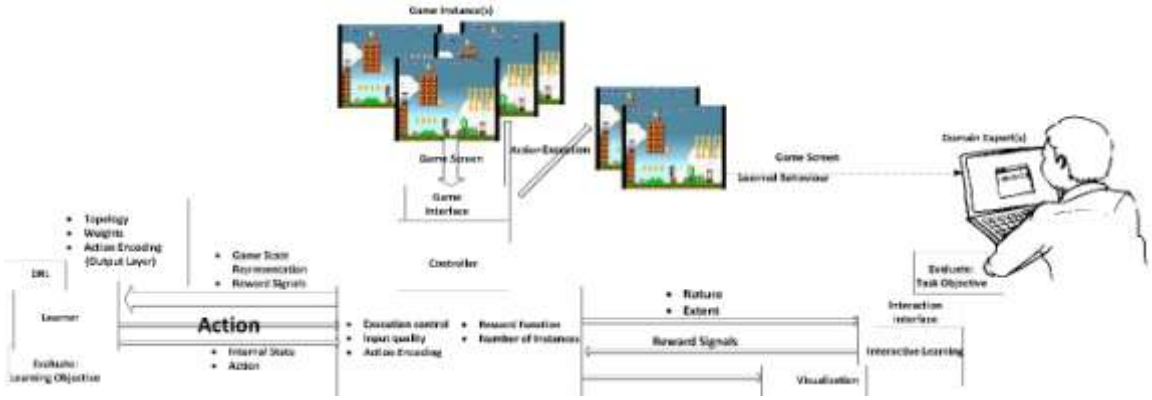

Figure 6. RL-based Serious Game Framework

The generic framework [37] offers default functionality and components for RL-based educational games. The framework's primary component is the controller, referred to as the game interface. The controller's major function is to receive learner actions and execute control. The domain expert implements the interaction and learning acquisition level and sets the learning objective. The game state of the player forms the learning environment, which is crucial in evaluating the learning objective of the game. The reward function of the learner is based on the actions taken and the positive development in changing the state, s, to a new state, s'.

4.3. Intelligent Admission Recommender System

Admissions are central to every educational institution. The objective of most institutions is to admit the best students into the appropriate programmes. Secondly, institutions are supposed to enrol the right number of students based on the availability of facilities and instructors. Furthermore, the admissions team should have firsthand information of the behaviour of students admitted for proactive counselling. The manual admission procedure is problematic, especially in a twenty-first century dominated by AI applications. In Education 4.0, technological integration and automation in learner admission are linked directly to Industry 4.0 to reduce undesirable unemployment figures[38]. The intelligent admission system, as illustrated in Figure 7, has an input component that forms the new students' state. The new students' state includes historical behaviour, grades and other relevant information. The agent that serves as the central engine of the framework receives the input and recommends a policy-informed action. The output modules guide the action recommended by the agent.

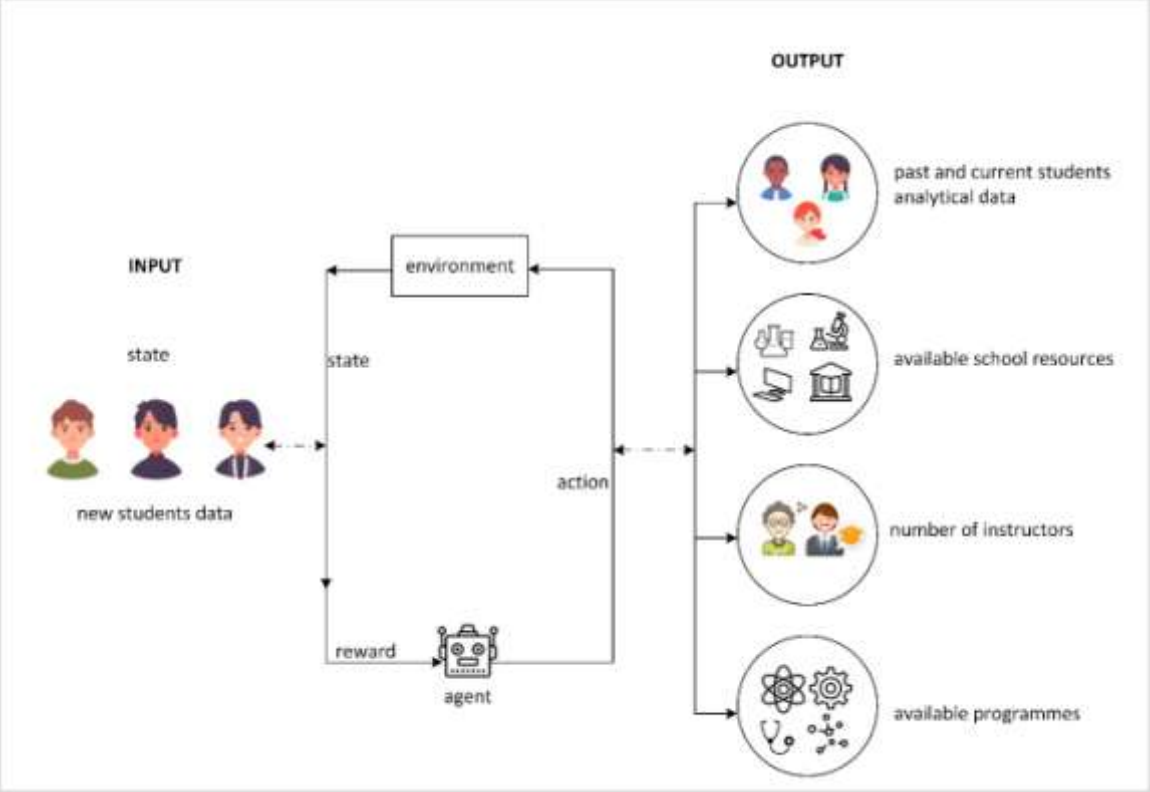

Figure 7. Intelligent Students Admission Framework

The output modules has four essential components. The past and current student analytical data module containsintelligent information about past and present students. This analytical module includes the profession of past students, grade performance, programmes studied, behaviour and

their admission data. The available school resources module informs the admission teamof the maximum number of studentswho can be admitted to a particular programme. The number of instructors module also helps the admission team to know the availability of academic staff in the institution for quality learner education. This helps enrol the right number of new students to different programmes. Available programme module guides in course availability and the number of applicants left. The agent receives a positive reward when the student is admitted into the right programme based on available analytics but a punishment when the recommendation by the agent results in admission problems.

4.4. Smart Library Management

Libraries are central to every educational institution. Primary, libraries support teaching and learning with a collection of books, journals, digital content and media that are accessible easily. However, the increasing number of students and changing pedagogy have necessitated the incorporation of technology into library management[39].The modern-day Internet of Things (IoT) enabled library consists of sensing technologies for object connectivity in tracking library facilities and generating big data for analytics. Book recommendation, seat management, stock control, security and user profiles in a modern-day library are integrated with AI. The intelligent library goes beyond the physical space of learning to online mediums and analytics that drive personalised learning[40]. RL agents in a library can be deployed as physical robots embedded with RL algorithms or as software agents on smart devices.

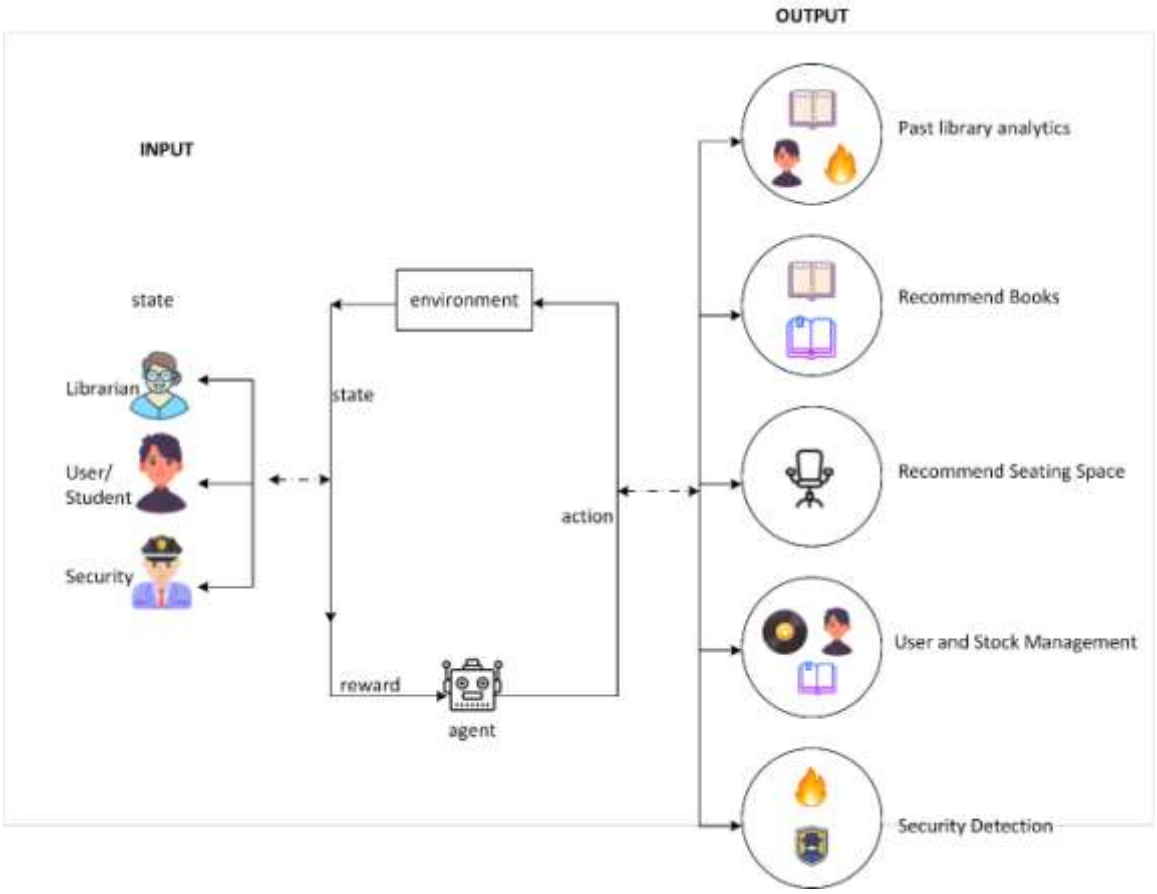

Figure 8. Intelligent Library Management

As illustrated in Figure 8, the RL-based intelligent library management consists of three primary users who form the state and are part of the environment. The librarian plays a vital role in managing library resources and library users. The emotional state of the librarian varies from happy to sad mood. In addition, the RL agent receives other relevant state information from the librarian concerning stock management, shelve management, behaviour and attendance. The RL agent recommends actions using past library analytics, user and stock management to determine the new state of the librarian. The effect of the proposed action determines the reward of the RL agent. The user module of the intelligent library system is fundamental in promoting learning. The library user has diverse states, including behavioural, preferential and academic. The RL agent recommends books to the user based on the past library analytics of the user. In addition, the seating space based on availability, user preference and behavior is also enforced as an action. The new state of the user after the recommended action varies from positive to negative feedback, guiding the RL agent on the next action. Security in the library pertains to theft and other dangers threatening library resources and users. The awareness and state of the security personnel is central to the environment. Past library analytics concerning security lapses and detection also forms a significant component in the library environment. The RL agent recommends actions to the security personnel based on past library analytics of user behaviour that leads to theft and security signals that should not be ignored. The security personnel enforces the recommended action with a changing state. The actions of the security personnel can be positive or negative. This informs the next recommendation by the RL agent.

4.5. Automation of E-learning Platforms

The COVID-19 pandemic has increased the use of electronic learning platforms globally[41, 42]Whether it's an online learning platform or a learning management system, incorporating intelligent analytics into previously intractable learner patterns is critical in modern-day education. The proliferation of the internet, affordable digital devices, user-friendly learning management systems (LMS) and multimedia has strengthened the adoption of e-learning platforms in educational institutions. The adoption of e-learning also provides Big Data sources, which are essential in developing an analytic engine that covers various aspects of the learning experience.Integrating RL-based software agents in e-learning platformsgoes beyond conventional analysis to learning perfection without class labels and feature selection. As illustrated in Figure 9, the RF-based intelligent e-learning system has two primary users, student and teacher.

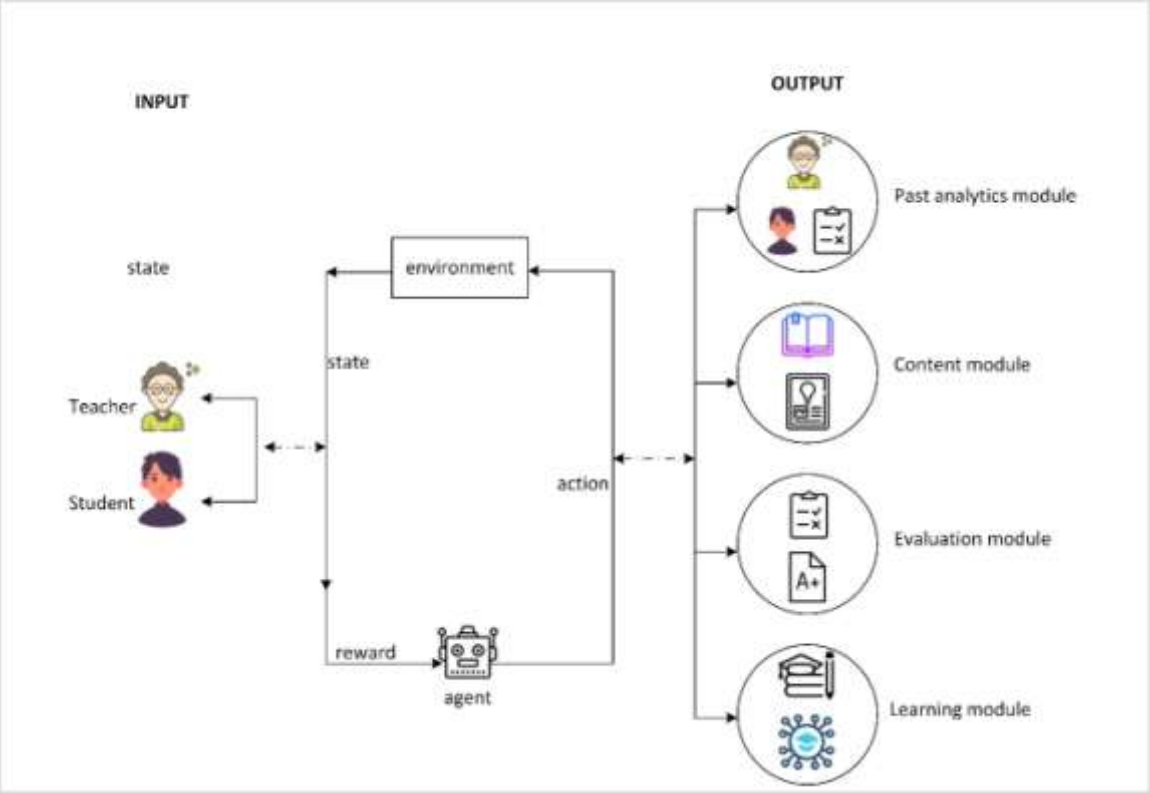

Figure 9. Intelligent E-learning System

The student has diverse input states based on interaction with the e-learning system. The state of the student and the teacher in an e-learning system changes significantly after interacting with the system. As teachers set up educational content that seeks to cover the curriculum, the need for intelligent analytics is relevant for personalised learning and collaborative studies. The modern-day instructor relies on analytics from learner interactions with the e-learning modules for reflective practice.The agent's action in an intelligent E-learning system is based on four modules: the past analytics module, the learning module, the evaluation module and the content module. The content module includes multimedia elements such as text, animation, video, audio and images. The agent's action on content selection depends on the past analytics module and the state of the learner and teacher. The evaluation module contains assessment strategies for specific learners based on the current state and the past analytics module. The evaluation engine determines learners preferred means of assessment to improve strengths and suggest areas of improvement. The agent receives a positive reward when the suggested evaluation leads to performance and skills enhancement. The learning module can be teacher-centred, learner-centred and collaborative based on the past analytics modules and the learner's current state. The agent's goal is to recommend the best learner module for respective students and create analytics based on rewards and punishment from learner performances. The RF-based intelligent e-learning system agent proposes actions based on the four modules after receiving state inputs from the learner and teacher. Implementing the agent's action based on the modules leads to an improved or worst state for the learner and teacher.The agent receives a reward or punishment which informs the next action using the exploration and exploitation RL mechanism.

5. DEPLOYMENT CHALLENGES OF RL IN EDUCATION

Reinforcement Learning, as discussed, has application deployments that will advance Education 4.0 with intelligent AI integrations and modules. However, RL implementation has foreseeable challenges that need solutions generally and in the context of education.

5.1. Availability of Data

Reinforcement Learning algorithms perform better with large data samples [43, 44]. Exploration in RL increases the chances of higher agents' reward, but this trade-off with the exploitation mechanism is data intensive. Furthermore, RL agents learn incrementally with data until mastery before proposing intelligent and efficient actions. In an uncertain environment with sequential decision enforcement, RL algorithms are data-dependent and rely mainly on the state space data definitions. In educational institutions, effective data aggregation remains a problem, especially in developing countries[45]. Aside the volumes of data expected, the variety of data requires well structured record keeping to enable state-action-reward policy definitions with the right RL algorithm. The velocity of data generation also requires best data storage architectures to avoid missing values that affects the performance of the RL algorithms.

5.2. Partially Observable Environment

The non-stationary nature of objects in some detectable environments has made it impossible for RL agents to observe some state space features fully. Partially Observable Markov Decision Process (POMDP) is a crucial challenge in modelling RL environment and, if not addressed, can lead to agents performing poorly in the environment[46].In POMDP, the agents' observation $x \in X$ is separate from the state with $O(x \mid s)$ as observation function with the probability of observing x in the environment s [43]. Even though fully observable environments are formalised as MDPs, most environments are partially observable, where it is difficult for an agent to obtain knowledge of the environment's full state or transition probabilities. In an educational setting, the agent cannot fully observe the environment. In RL-based robots in a classroom module, for instance, there are no observations on the mental state of the learners but rather sensors that pick the facial emotions. Addressing POMDP in RL is an ongoing research challenge, with several proposed models stacking observation as full state.

5.3. Curse of Dimensionality

In RL modelling, the state and action space representation grows exponentially when exploring optimal control in high-dimensional spaces, and the phenomenon is described as Bellman's curse of dimensionality[47]. The curse of dimensionality is unavoidable in an environment with diverse and growing features and data. As the dimensionality increases, more data points are needed to model good performances of the environment, and the situation is considered a curse. The curse of dimensionality has negative implications on memory and predominantly affects the effective deployment of RL applications[48]. In an educational setting, the environment monitored has various features and data that increase the state and action representation of the RL agent. The curse of dimensionality is an ongoing research difficulty in RL, and efficient solutions are required before real-world application deployments.

5.4. Coordination in Multi-Agents

Whether cooperative or competitive, multi-agents in RL have significant advantages, including faster task execution, performance improvement, resource sharing and intrigue competition.

Aside from scalability that leads to the curse of dimensionality [23], coordination among agents for cooperative and competitive tasks is problematic. In multi-agent systems (MASs), the goal is for the agents to become autonomous in task execution. The independent agents in MASs are also expected to communicate and share ideas with other agents to achieve a goal. Even with centralised training and decentralised execution, improper agent coordination can lead to the selfish depletion of MASs resources and the crashing of MAS applications.

6. CONCLUSION

Education 4.0 is driving derivatives to transform the future of teaching and learning with advanced and emerging technologies. Coupled with the tragic COVID-19 pandemic, educational reforms and policies have tilted to technological integration that will discontinue conventional educational applications. Even though traditional educationhas significant advantages, the need for intelligent analytics and predictive education cannot be over-emphasised. Supervised and unsupervised ML algorithm has been discussed, and applications are deduced in literature to cover diverse domain aspects of education. The challenge, however, is the discussion and implementation of Reinforcement Learning, which involves incremental learning of agents in an environment to the mastery level.

Reinforcement Learning leads to autonomy in educational applications and modules where personalised, collaborative and reflective practices are deduced with agents mastering their environment. The survey discussed critical deployment areas of RL learning in education with diagrammatic modules that identify state-action-reward policies of the agents. The modules discussed include RL agents in the classroom, admission, game development, library and e-learning system. These modules are central in Education 4.0 and relevant to all stakeholders in education. Aside from the application modules, the anticipated challenges to successful RL algorithm implementation in educational institutions were also considered.

REFERENCES

[1] L. M. Castro Benavides, J. A. Tamayo Arias, M. D. Arango Serna, J. W. Branch Bedoya, and D. Burgos, "Digital Transformation in Higher Education Institutions: A Systematic Literature Review," *Sensors (Basel).*, vol. 20, no. 11, pp. 1–22, 2020, doi: 10.3390/s20113291.

[2] N. Bremner, "ResearchSPAce," pp. 53–64, 2019.

[3] A. Mikroyannidis, J. Domingue, M. Bachler, and K. Quick, "A Learner-Centred Approach for Lifelong Learning Powered by the Blockchain," *EdMedia + Innov. Learn. 2018*, pp. 1388–1393, 2018, [Online]. Available: http://uk.businessinsider.com/santander-has-20-25-use-cases-for-bitcoins-blockchain-technology-everyday-banking-2015-6%0Ahttps://www.semanticscholar.org/paper/A-Learner-Centred-Approach-for-Lifelong-Learning-by-Domingue-Bachler/ce096256873ab23915eb39312de.

[4] D. Majumdar, P. K. Banerji, and S. Chakrabarti, "Disruptive technology and disruptive innovation: ignore at your peril!," *Technol. Anal. Strateg. Manag.*, vol. 30, no. 11, pp. 1247–1255, 2018, doi: 10.1080/09537325.2018.1523384.

[5] H. Akoto-Baako, P. J. Heeralal, and B. Kissi-Abrokwah, "Concept of Increase Enrolment: Its effect on teachers in Ghana," *Mediterr. J. Soc. Sci.*, vol. 12, no. 6, p. 167, 2021, doi: 10.36941/mjss-2021-0066.

[6] A. Serrano Mamolar, P. Salvá-García, E. Chirivella-Perez, Z. Pervez, J. M. Alcaraz Calero, and Q. Wang, "Autonomic protection of multi-tenant 5G mobile networks against UDP flooding DDoS attacks," *J. Netw. Comput. Appl.*, vol. 145, no. November 2018, p. 102416, 2019, doi: 10.1016/j.jnca.2019.102416.

[7] A. Mirahmadizadeh *et al.*, "Evaluation of students' attitude and emotions towards the sudden closure of schools during the COVID-19 pandemic: a cross-sectional study," *BMC Psychol.*, vol. 8, no. 1, pp. 1–7, 2020, doi: 10.1186/s40359-020-00500-7.

[8] A. M. Müller, C. Goh, L. Z. Lim, and X. Gao, "Covid-19 emergency elearning and beyond:

Experiences and perspectives of university educators," *Educ. Sci.*, vol. 11, no. 1, pp. 1–15, 2021, doi: 10.3390/educsci11010019.

[9] M. O. Riedl, "Human-centered artificial intelligence and machine learning," *Hum. Behav. Emerg. Technol.*, vol. 1, no. 1, pp. 33–36, 2019, doi: 10.1002/hbe2.117.

[10] V. Kuleto *et al.*, "Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions," *Sustain.*, vol. 13, no. 18, pp. 1–16, 2021, doi: 10.3390/su131810424.

[11] M. van der Schaar *et al.*, "How artificial intelligence and machine learning can help healthcare systems respond to COVID-19," *Mach. Learn.*, vol. 110, no. 1, pp. 1–14, 2021, doi: 10.1007/s10994-020-05928-x.

[12] M. Ghobakhloo, "Industry 4.0, digitization, and opportunities for sustainability," *J. Clean. Prod.*, vol. 252, p. 119869, 2020, doi: 10.1016/j.jclepro.2019.119869.

[13] C. Bai, P. Dallasega, G. Orzes, and J. Sarkis, "Industry 4.0 technologies assessment: A sustainability perspective," *Int. J. Prod. Econ.*, vol. 229, p. 107776, 2020, doi: 10.1016/j.ijpe.2020.107776.

[14] A. H. Anaelka, "Education 4.0 Made Simple: Ideas For Teaching," *Int. J. Educ. Lit. Stud.*, vol. 6, no. 3, p. 92, 2018, [Online]. Available: https://journals.aiac.org.au/index.php/IJELS/article/view/4616.

[15] R. Kasih, N. Hanafi, and M. Amin, "Education 4.0 and the 21st Century Skills: A Case Study of Smartphone Use in English Classes," vol. 465, no. Access 2019, pp. 48–51, 2020, doi: 10.2991/assehr.k.200827.013.

[16] M. Batta, "Machine Learning Algorithms - A Review ," *Int. J. Sci. Res. (IJ*, vol. 9, no. 1, pp. 381-undefined, 2020, doi: 10.21275/ART20203995.

[17] R. Saravanan and P. Sujatha, "Algorithms : A Perspective of Supervised Learning Approaches in Data Classification," *2018 Second Int. Conf. Intell. Comput. Control Syst.*, no. Iciccs, pp. 945–949, 2018, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8663155.

[18] D. K. Dake, J. D. Gadze, G. S. Klogo, and H. Nunoo-mensah, "Multi-Agent Reinforcement Learning Framework in SDN-IoT for Transient Load Detection and Prevention," 2021.

[19] M. A. Yasin, W. A. M. Al-Ashwal, A. M. Shire, S. A. Hamzah, and K. N. Ramli, "Tri-band planar inverted F-antenna (PIFA) for GSM bands and bluetooth applications," *ARPN J. Eng. Appl. Sci.*, vol. 10, no. 19, pp. 8740–8744, 2015.

[20] E. Asiain, J. B. Clempner, and A. S. Poznyak, "Controller exploitation-exploration reinforcement learning architecture for computing near-optimal policies," *Soft Comput.*, vol. 23, no. 11, pp. 3591–3604, 2019, doi: 10.1007/s00500-018-3225-7.

[21] M. Van Otterlo and M. Wiering, "Reinforcement learning and markov decision processes," *Adapt. Learn. Optim.*, vol. 12, pp. 3–42, 2012, doi: 10.1007/978-3-642-27645-3_1.

[22] D. K. Dake, G. S. Klogo, J. D. Gadze, and H. Nunoo-Mensah, "Traffic Engineering in Software-defined Networks using Reinforcement Learning: A Review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 5, pp. 330–345, 2021, doi: 10.14569/IJACSA.2021.0120541.

[23] L. Canese *et al.*, "Multi-agent reinforcement learning: A review of challenges and applications," *Appl. Sci.*, vol. 11, no. 11, 2021, doi: 10.3390/app11114948.

[24] A. S. Polydoros and L. Nalpantidis, "Survey of Model-Based Reinforcement Learning: Applications on Robotics," *J. Intell. Robot. Syst. Theory Appl.*, vol. 86, no. 2, pp. 153–173, 2017, doi: 10.1007/s10846-017-0468-y.

[25] T. Degris, P. M. Pilarski, and R. S. Sutton, "Model-Free reinforcement learning with continuous action in practice," *Proc. Am. Control Conf.*, pp. 2177–2182, 2012, doi: 10.1109/acc.2012.6315022.

[26] M. Hausknecht, P. Stone, and O. Mc, "On-Policy vs. Off-Policy Updates for Deep Reinforcement Learning," *Ijcai*, 2016.

[27] S. Fujimoto, H. Van Hoof, and D. Meger, "Addressing Function Approximation Error in Actor-Critic Methods," *35th Int. Conf. Mach. Learn. ICML 2018*, vol. 4, pp. 2587–2601, 2018.

[28] H. Qie, D. Shi, T. Shen, X. Xu, Y. Li, and L. Wang, "Joint Optimization of Multi-UAV Target Assignment and Path Planning Based on Multi-Agent Reinforcement Learning," *IEEE Access*, vol. 7, pp. 146264–146272, 2019, doi: 10.1109/ACCESS.2019.2943253.

[29] H. U. Sheikh and L. Boloni, "Multi-Agent Reinforcement Learning for Problems with Combined Individual and Team Reward," *Proc. Int. Jt. Conf. Neural Networks*, 2020, doi: 10.1109/IJCNN48605.2020.9206879.

[30] A. Hassan, N. Z. Abiddin, and S. K. Yew, "The Philosophy of Learning and Listening in Traditional

Classroom and Online Learning Approaches," *High. Educ. Stud.*, vol. 4, no. 2, pp. 19–28, 2014, doi: 10.5539/hes.v4n2p19.

[31] S. Hartikainen, H. Rintala, L. Pylväs, and P. Nokelainen, "The concept of active learning and the measurement of learning outcomes: A review of research in engineering higher education," *Educ. Sci.*, vol. 9, no. 4, pp. 9–12, 2019, doi: 10.3390/educsci9040276.

[32] A. G. Barto, P. S. Thomas, and R. S. Sutton, "Some recent applications of reinforcement learning," *Work. Adapt. Learn. Syst.*, p. 6, 2017, [Online]. Available: http://psthomas.com/papers/Barto2017.pdf.

[33] E. Gyimah, D. K. Dake, and M. Agbeko, "The Role of Computer Games in the Learning of Programming among Tertiary Students in Ghana," *African J. Appl. Res.*, vol. 4, no. 2, pp. 242–252, 2018.

[34] E. Sudarmilah, U. Fadlilah, H. Supriyono, F. Y. Al Irsyadi, Y. S. Nugroho, and A. Fatmawati, "A review: Is there any benefit in serious games?," *AIP Conf. Proc.*, vol. 1977, no. June 2018, 2018, doi: 10.1063/1.5042915.

[35] A. Dobrovsky, C. W. Wilczak, P. Hahn, M. Hofmann, and U. M. Borghoff, "Deep Reinforcement Learning in Serious Games: Analysis and Design of Deep Neural Network Architectures," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10672 LNCS, pp. 314–321, 2018, doi: 10.1007/978-3-319-74727-9_37.

[36] I. Szita, "Reinforcement learning in games," *Adapt. Learn. Optim.*, vol. 12, pp. 539–577, 2012, doi: 10.1007/978-3-642-27645-3_17.

[37] A. Dobrovsky, U. M. Borghoff, and M. Hofmann, "Applying and augmenting deep reinforcement learning in serious games through interaction," *Period. Polytech. Electr. Eng. Comput. Sci.*, vol. 61, no. 2, pp. 198–208, 2017, doi: 10.3311/PPee.10313.

[38] M. Tvaronavičienė, "Insights into global trends of capital flows' peculiarities: Emerging leadership of China," *Adm. si Manag. Public*, vol. 2019, no. 32, pp. 6–17, 2019, doi: 10.24818/amp/2019.32-01.

[39] A. Larrabee Sønderlund, E. Hughes, and J. Smith, "The efficacy of learning analytics interventions in higher education: A systematic review," *Br. J. Educ. Technol.*, vol. 50, no. 5, pp. 2594–2618, 2019, doi: 10.1111/bjet.12720.

[40] J. Liu, "Construction of Intelligent Library Service System from the Perspective of Artificial Intelligence," *Int. J. Front. Sociol.*, vol. 3, no. 1, pp. 44–51, 2021, doi: 10.25236/ijfs.2021.030106.

[41] A. Shahzad, R. Hassan, A. Y. Aremu, A. Hussain, and R. N. Lodhi, "Effects of COVID-19 in E-learning on higher education institution students: the group comparison between male and female," *Qual. Quant.*, vol. 55, no. 3, pp. 805–826, 2021, doi: 10.1007/s11135-020-01028-z.

[42] X. Xie, K. Siau, and F. F. H. Nah, "COVID-19 pandemic–online education in the new normal and the next normal," *J. Inf. Technol. Case Appl. Res.*, vol. 22, no. 3, pp. 175–187, 2020, doi: 10.1080/15228053.2020.1824884.

[43] G. Dulac-Arnold *et al.*, *Challenges of real-world reinforcement learning: definitions, benchmarks and analysis*, vol. 110, no. 9. Springer US, 2021.

[44] J. Oh *et al.*, "Discovering reinforcement learning algorithms," *Adv. Neural Inf. Process. Syst.*, vol. 2020-Decem, no. NeurIPS, 2020.

[45] B. Daniel, "Big Data and analytics in higher education: Opportunities and challenges," *Br. J. Educ. Technol.*, vol. 46, no. 5, pp. 904–920, 2015, doi: 10.1111/bjet.12230.

[46] M. T. J. Spaan, "Partially observable markov decision processes," *Adapt. Learn. Optim.*, vol. 12, pp. 387–414, 2012, doi: 10.1007/978-3-642-27645-3_12.

[47] A. Koppel, G. Warnell, E. Stump, P. Stone, and A. Ribeiro, "Policy evaluation in continuous mdps with efficient kernelized gradient temporal difference," *IEEE Trans. Automat. Contr.*, vol. 66, no. 4, pp. 1856–1863, 2021, doi: 10.1109/TAC.2020.3029315.

[48] L. C. Garaffa, M. Basso, A. A. Konzen, and E. P. de Freitas, "Reinforcement Learning for Mobile Robotics Exploration: A Survey," *IEEE Trans. Neural Networks Learn. Syst.*, pp. 1–15, 2021, doi: 10.1109/TNNLS.2021.3124466.

AUTHOR

Delali Kwasi Dake, PhD is a Senior Lecturer in the department of ICT Education at the University of Education, Winneba, Ghana. His research interests include educational data mining, artificial intelligence, sentiment analysis, software-defined networking, and Internet of Things.