Co-Designing Instruction in Virtual Learning Environments Using AI

by

Aishwarya Ganesh

A Project submitted to the
School of Graduate and Postdoctoral Studies in partial
fulfillment of the requirements for the degree of

Master of Education in **Education**

Faculty of Education

University of Ontario Institute of Technology (Ontario Tech University)

Oshawa, Ontario, Canada

December 2023

PROJECT REVIEW INFORMATION

Submitted by: **Aishwarya Ganesh**

Master of Education in **Education**

Project/Major Paper title: Co-Designing Instruction in Virtual Learning Environments Using AI

The Project was approved on December 8, 2023, by the following review committee:

Review Committee:

Research Supervisor	DR. WILLIAM HUNTER
Co-supervisor	N/A
Second Reader	DR. ALYSON KING

The above review committee determined that the Project is acceptable in form and content and that a satisfactory knowledge of the field was covered by the work submitted. A copy of the Certificate of Approval is available from the School of Graduate and Postdoctoral Studies.

ABSTRACT

The literature was explored to determine how artificial intelligence (AI) systems and algorithms are currently being used in the co-design of learning within virtual learning environments. Through the analysis of literature, the study aims to retrieve multiple methods of AI assistance to ease or uplift the educator's role in online learning design. The study determined a variety of themes that determine methods of AI use in online instruction, such as prediction, providing feedback, adaptive learning, and providing visualization of student data on learning management systems (LMS). The study also determined the importance of a repository of various student data input in AI algorithms, and the collaboration of educators and experts in the process of using AI systems. The key implications suggest the importance of bridging feedback immediacy and formative approaches to improving student performance in online environments. Furthermore, the study also determines the changing roles of stakeholders in the education process. Finally, it also suggests the potential to create a multifaceted AI system and an effective LMS that supports such features.

Word limit: **15,604 words Masters Project**

Keywords: artificial intelligence; instructional design; learning management systems; data; education

AUTHOR'S DECLARATION

I hereby declare that this project consists of original work of which I have authored. This is a true copy of the work, including any required final revisions, as accepted by my committee.

I authorize the University of Ontario Institute of Technology (Ontario Tech University) to lend this work to other institutions or individuals for the purpose of scholarly research. I further authorize the University of Ontario Institute of Technology (Ontario Tech University) to reproduce this work by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my work may be made electronically available to the public.

Aishwarya Ganesh

ACKNOWLEDGEMENTS

I would like to acknowledge the support of my supervisor, Dr. William Hunter who has not only been insightful in their role and helped me be a better researcher, but also have been patient, kind, and would continually articulate their belief in me. I would like to also give acknowledgement to my family and friends who have been encouraging and uplifting at all steps of my education and career and have been the root of my confidence and perseverance.

STATEMENT OF CONTRIBUTIONS

I hereby certify that I am the sole author of this work and that no part of this work has been published or submitted for publication. I have used standard referencing practices to acknowledge ideas, research techniques, or other materials that belong to others. Furthermore, I hereby certify that I am the sole source of the creative works and/or inventive knowledge described in this document.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2

LIST OF FIGURES

CHAPTER 1

LIST OF ABBREVIATIONS AND SYMBOLS

AAI/KBS	Artificial Augmented Intelligence/Knowledge-Based System
AES	Adaptive E-learning System
AI	Artificial Intelligence
AIEd	Artificial Intelligence in Education
ANN	Artificial Neural Network
AR-AI	Augmented Reality - Artificial Intelligence
cGPA	Cumulative Grade Point Average
DRFLO	Dynamic Recommendation of Filtered Learning Objects
EC	Evolutionary Computation
EWS	Early Warning System
GAR	Gradual-At-Risk
GP	Genetic Programming
ITS	Intelligent Tutoring Systems
LA	Learning Analytics
LMS	Learning Management System
LO	Learning Object
MDI	Misconception Detection and Identification
RiPPLE	Recommendation in Personalized Peer-Learning Environment
SME	Subject Matter Experts
SWS	Scholarly Writing Software

Chapter 1. Introduction

1.1 Virtual Learning Environments

When prompted with "Why are internet and digital technologies crucial to learning in the 21st century?" the artificial intelligence program Chat-GPT generated text which suggested that Internet and digital technologies possess vast potential in providing information access, opportunity for social and global connectivity, and improving creativity and innovation (OpenAI, 2023). Using such technologies to create learning solutions that heighten learner performance and professional development is known as electronic learning or e-learning (Kavitha & Lohani, 2019). A factor in changing educational contexts is the growing use of information technologies and the digitization of learning environments to create online learning environments (Gopo, 2022). Online communities hold additional learning challenges and have been increasingly normalized in a world reeling from the recent COVID-19 pandemic. In virtual learning environments (VLE), the constraints of time and geography are blurred to accommodate a variety of circumstances, including distance and part-time education (Kavitha & Lohani, 2019). How instruction is designed in virtual environments is vital to teaching and learning and can determine student success, measured through retention and motivation (Bedregal-Alpaca et al., 2022).

1.1.1 Learning Management Systems

Learning Management Systems (LMS) are significant facets of virtual learning. This type of software manages how learners and instructors interact with one another and how the learner interacts with the learning materials on the cloud-based system (David, 2013). Some learning materials in LMS classrooms or courses use learning objects (LOs). LOs can be used (and re-used) by designers within learning systems to uplift their course design through smaller, modular training content with a singular objective (Davey, 2023). Examples of this content could be assessment pieces such as quizzes, content such as videos, and interactive pieces such as games or simulations (Davey, 2023). LOs are interoperable among LMSs, stored as metadata and extractable from LO repositories through “tags” and keywords (Davey, 2023). Recent expectations of LMS include self-monitoring options, gamification aspects, notification of missing or pending tasks, instructional interventions displaying student strengths and weaknesses, and predicting student achievements (Sahin & Yurdugül, 2022). Using e-learning principles, artificial intelligence (AI) technologies can potentially elevate virtual learning environments to improve student learning performance and abilities (Ouyang et al., 2023; Sayed et al., 2022).

1.2 Uses of Artificial Intelligence in Education in Previous Studies

1.2.1 Artificial Intelligence

Artificial intelligence (AI) is “the development of machines that have some level of intelligence, with the ability to perform human-like functions, including cognitive, learning, decision-making, and adapting to the environment” (Chen et al., 2020, p.75267). The intelligence of these machines comes from the capacity to mimic human abilities, such as in cognitive problem solving, pattern recognition, and adapting content based on these inputs (Chen et al., 2020), as well as the competence to replace human intelligence characteristics of, for instance, “visual perception, speech recognition, decision-making, and translation between languages" (Chassignol et al., 2018, p.17). It is valuable to note that AI functions to make predictions through the harnessing of data and input into specific algorithms (Alexander et al., 2019), suggesting that decision-making and predictions are dependent on computational analogies of quantitative datasets. Moreover, growing maturation of AI technologies (Alexander et al., 2019) determines that natural intelligence of humans seems to not be wholly imitable yet.

1.2.2 Artificial Intelligence in Education

Artificial Intelligence in Education (AIEd) has encompassed the use of learning analytics (LA) (Sayed et al., 2022), assessment and evaluation (Zhao et al., 2023), eye tracking and facial interpretation analysis (Seo et al., 2021; Meikleham & Hugo, 2020), intelligent tutoring systems (ITS), detecting behaviours, automation of feedback based on student assessment, educational robots, and others (Xu & Ouyang, 2022). The capacities of AIEd can thus potentially help to facilitate the teaching and learning processes. The uses of such technologies include immediate support and personalized usage by educators, writing support such as grammar checking and suggested phrasing, research support such as analysis and summary, generating visual media and audio support, and support with administrative and technical tasks to improve efficiency (Chan & Hu, 2023). The willingness to use and leverage such technologies positively correlates with increased use (Xu & Ouyang, 2022). Furthermore, educators may benefit from incorporating AI technologies that allow students to receive immediate, just-in-time support and feedback when the teacher is unavailable (Seo et al., 2021; Meikleham & Hugo, 2020). The concept of AI, however, is vast, and exploring its various underpinnings can help to visualize a deeper understanding of its uses in education.

1.2.3 Machine Learning

One such underpinning is machine learning, a facet of AI technologies that uses computational algorithms to generate data patterns (Ng et al., 2022). These patterns are used to train and optimize intelligent systems without continued human intervention (Ng et al., 2022; Srivastava et al., 2019). Various examples of machine learning software have been implemented, particularly in post-secondary settings, such as adaptive learning systems. Adaptive learning systems are dynamic; using student learning profiles, retrieved through learning style data and assessments, learning experiences tailored to individual needs can be generated (Khosravi et al. 2020). These experiences are modified based on students' interactions with the system, allowing students to move at a pace that suits their abilities (Khosravi et al., 2020). Adaptive systems use a repository of available data to cultivate a learning environment based on student needs and can continually adapt the learning based on interactions with students during a learning experience (Khosravi et al., 2020). Such reinforcement learning mechanisms are embedded within systems using algorithms that receive input from student data or previously existing data (i.e., previous courses or training and sample data) and generate an output of customized learning trajectories (Sayed et al., 2022). Feedback-based learning networks help not only demonstrate improved student learning capacities (Sayed et al., 2022; Demszky et al., 2023), but also improve instructional efficiency in the classroom, particularly online (Meikleham & Hugo, 2020).

1.2.4 Deep Learning and Neural Networks

A more complex version of machine learning is deep learning, which is a network model that uses many “layers” of input or information to strengthen its understanding of specific input data (Ng et al., 2022). The model presents itself as a series of stacked, successive layers from which specific features, such as age, gender, or ethnicity, are extracted (Ng et al., 2022; Chen et al., 2020). The model delves deeper by uncovering new or hidden layers from initial input layers that may further influence the understanding of the data and continually improves its efficiency as new data is added and updated (Ng et al., 2022). The complex network system mimics the human brain's functionality and is called a neural network model (Ng et al., 2022). Understanding the complexities of AI can help determine how such models fit into educational programming from the educator’s standpoint.

1.2.5 AIEd for Improved Education Design

In a study by Chocarro et al. (2016), perceptions of AI chatbots by educators have seen a positive correlation between the ease of use and usefulness of the technology and the intention to use the technology. That investigation suggests no greater intention to use AI chatbots if less mental effort is applied; instead, the tool's efficiency demonstrates its effectiveness and useability (Chocarro et al., 2023). AI machines and robots in education have also improved instructional practices and strategies (Chen et al., 2020). AIEd can foster engagement, improve the qualities of pedagogical tools used in online learning environments, foster academic integrity and content personalization, and improve uptake and retention, among others (Chen et al., 2020). Many studies demonstrate the appeal of AIEd as an aid in student learning

(Kavitha & Lohani, 2019; Sayed et al., 2022; Osakwe et al., 2022; Demszky et al., 2023; Shoufan, 2023; Chocarro et al., 2023). However, as the pedagogical contexts shift focus to learner-centrism (Debattista, 2018), educators must be included in the conversation as learners of professional development and instructional design and as equivalent stakeholders of the teaching-learning process. Considering the utility of valuable AI tools and technology, further exploration can provide educators, specifically in virtual environments, to improve the efficiency and effectiveness of their instructional design.

1.2.6 Concerns with AIEd

While many researchers see AIEd as positive for students, others critique AIEd in possessing several challenges (Chan & Hu, 2023; Seo et al., 2021). With the emergence of AI software, such as ChatGPT, learners have noted the importance of validating information generated from generative AI technologies (Shoufan, 2023; Chan & Hu, 2023). In a study of generative AI technologies, student participants note that their incomprehension of the technology's functional complexities leads to further concerns with how the information is generated (Chan & Hu, 2023). Thus, a lack of transparency of the processes used by the technology to make decisions and generate output can make generative AI technologies less trusting (Chan & Hu, 2023). Sharing information through AI messaging systems was also seen as a concern for those fearing security risks in which private information is stored and used for improvements (Chan & Hu, 2023). Furthermore, with learning analytics in online learning, the holding of digital records of individuals may also be a concern. Learning analytics in online learning is the process in which learner data, such as their background, behaviours, and progress, is collected (through interactions with the LMS) and automatically analyzed by the AI LA technology (Wilson et al., 2017). As AI LA technologies function by mining large amounts of data to create generalized behavioural patterns, residual traces of student data are left in learning analytic software as the machine engages in multiple iterations, using previous data and new input to update its understanding (Wilson et al., 2017).

A critical aspect of AI is that it can improve academic integrity such as through plagiarism detection, grading, and providing feedback for improvement, through applications such as Grammarly and Turnitin. (Chen et al., 2020) Interestingly, other AI technologies such as generative AI are criticized for inducing plagiarism (Chan & Hu, 2023). Challenges to creative and critical thinking skills, replacement or job takeover by AI technologies, inaccessibility to technologies in specific groups and communities, and irregularity of AI usage policies in leveraging or navigating the technology are all noted drawbacks with the use of certain generative AI technologies (Chan & Hu, 2023). In addition, in circumstances in which AI technologies are used to assess student work, there are concerns about having a machine be responsible for deciding student grades and the inability of students to dispute their marks or grades (Seo et al., 2021).

1.3 Research Focus

While the challenges of using AI technologies in the education system can be multifold, the challenges encourage research and a more in-depth understanding of how AI must be meaningfully incorporated into learning design. This paper explores the potential of relevant and ethical AI applications in educational contexts, specifically virtual learning environments. These technologies have innumerable uses for learners, however, educators may benefit from the applications in the planning process of online learning, as well. It is valuable to explore how AI technologies can be used to co-design the learning material, instead of solely the educator, to highlight the active role of the technologies in the design process and demonstrate educators' continued agency in the design of the materials.

This paper thus aims to investigate the following: “What exists in the literature that indicates how educators use ethical AI-based applications to co-design learning materials in virtual learning environments?” The research question aims to comprehensively document how educators can implement various artificial intelligence technologies into their design in virtual learning environments and how this may change the educator's capacity in the teaching-learning process. These methods can potentially generate a means for future instructional designers and educators to utilize similar strategies using AI technology in their classroom and design and perhaps circumvent potential aversion to the changes in educational climates through intelligent technologies.

Chapter 2. Methodology

2.1 Search Criteria

My literature search began in June 2023 and contains the most updated results as of October 2023. I used the following conditions in the search: peer-reviewed journals, studies from 2016 to 2023 (approximately 5 years) and written in the English language. Furthermore, I included only full-text articles indicating primary research and journal articles. It is important to emphasize the value of including students at all educational levels as data participants yet actively exclude them in the keyword search terms. Firstly, this paper aims to observe the uses of AI technologies by teachers in instructional design. Thus, keywords about students and learners were excluded from the search string to keep the focus more on educator use of the AI technologies and reduce research results that focused on the use of AI technologies by students in learning and for completing assessments.

However, this is not to say that students cannot be research participants. Studies demonstrating positive student performance through an AI-infused curricular design could potentially testify to the success of using AI methods of designing instruction, hence research that used students (at all educational levels i.e., from K-12 to graduate-level post-secondary) as participants were included. Secondly, AI is an emergent and fast-evolving theme in education, thus, studies were open to both students and educators as participants in order to have a variety and greater availability of data to analyze and construct meaning. A deeper inspection of the resultant papers could help keep the focus solely on the educator's use of AI technologies rather than the students’ use of them in their learning. Table 1 describes the search process employed in this study.

Table 1

Literature inclusion criteria

Literature Type	**Inclusion:** Empirical studies based on qualitative, quantitative, or mixed methods data, perspective-based literature, technical reports, conference papers, working papers **Exclusion**: Theses, dissertations, literature reviews, commentaries, and theoretical literature
Publication Source	**Inclusion**: Peer-reviewed scholarly journals, conference papers and proceedings **Exclusion**: Book chapters, books, speeches and presentations
Participants	**Inclusion**: (a) K-12 students of any age, gender, learning ability, geographical location, and socioeconomic status (b) Post-secondary students of any age, gender, learning ability, geographical location, and socioeconomic status (c) Teachers or educators at any institutional level and of any age, gender, and geographic location **Exclusion:**
Keywords	**Inclusion**: (“artificial intelligence” OR “machine learning” OR chatgpt OR “chat gpt”) AND (e-learning OR elearning OR “online learning” OR “virtual learning”) AND (“instructional design” OR “curriculum design” OR “teaching design” OR “e-learning design” OR co-design*) AND (teachers OR educators OR “teaching assistants” OR “student assistants” OR instructor) **Exclusion:** “students” OR “K-12 students” OR “post-secondary students” OR “learners” OR “K-12 learners” OR “post-secondary learners”

Various databases were used to complete the search, such as Education via EBSCO, ERIC via ProQuest, PsychINFO via ProQuest, and Web of Science. The search string, (“artificial intelligence” OR “machine learning” OR chatgpt OR “chat gpt") AND (e-learning OR elearning OR "online learning" OR "virtual learning") AND ("instructional design" OR "curriculum design" OR "teaching design" OR "e-learning design" OR co-design*) AND (teachers OR educators OR "teaching assistants" OR "student assistants" OR instructor) was inputted into Education via EBSCO, producing 13 results. The same exact search string was put into ERIC via ProQuest, garnering 6 results. The same exact search string produced only 1 result in PsycINFO via ProQuest and 15 results in Web of Science.

As teachers and educators are bound to be linked to the concepts of curriculum and instruction, it was decided to remove (teachers OR educators OR "teaching assistants" OR "student assistants" OR instructor) to observe the availability of other studies otherwise omitted with the search terms. The new string, ("artificial intelligence" OR "machine learning" OR chatgpt OR "chat gpt") AND (e-learning OR elearning OR "online learning" OR "virtual learning") AND ("instructional design" OR "curriculum design" OR "teaching design" OR "e-learning design" OR co-design*), was plugged into the same databases.

With similar exclusion and inclusion criteria and conditions of English language papers, peer-reviewed journals and studies only between 2016-2023, the Education via EBSCO database produced 24 results, the ERIC via ProQuest database produced 16 results, the PsycINFO via ProQuest database produced 3 results, and the Web of Science database produced 29 results.

The total number of papers produced was 108 papers from both searches. The titles of these searches were added to an Excel Sheets document. Duplicates were removed by accessing the "Data" tab in the toolbar, selecting "Data cleanup," and then “Remove duplicates.” A manual check through the papers to ensure all duplicates were removed was also conducted. A few discrepancies could render the automated duplicate removal tool ineffective (i.e., the difference between two titles due to a punctuation mark). This yielded 64 original papers. The titles of these papers were assessed for relevance to particularly the uses of AI in education, of which 53 were found relevant. From these results, 27 articles were selected based on the abstract description and omitting literature reviews. Through a skim of the 27 papers, 10 were chosen based on skimming the study’s methodology and results. This process for eliminating papers is described in Table 2.

Table 2

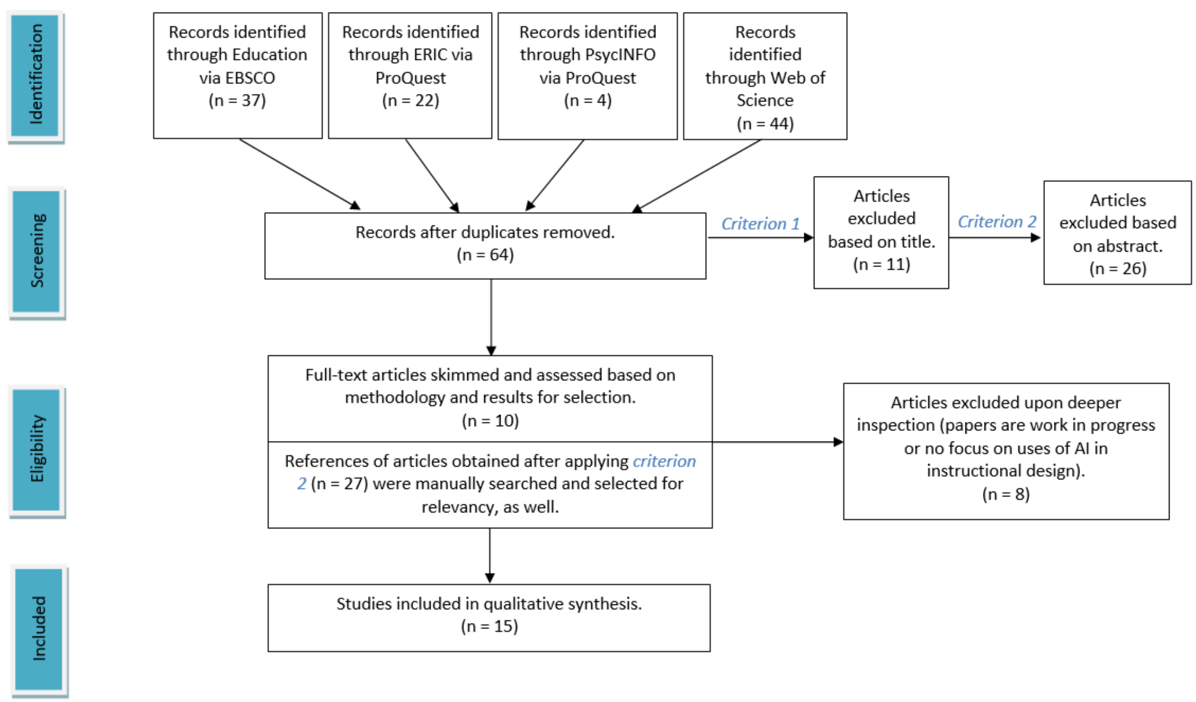
Coding Process

Coding Level	Exclusion Description
Inclusion based on the title	Titles demonstrate the use of AI technology in general or a specific AI model or algorithm in the design of classroom or course instruction and assessment.
Inclusion based on abstract	Abstract summarizes the uses of AI technologies in the research study and the resultant effects of using the technology in their study (findings).
Inclusion based on a skim of methodology and results	Methodology and results depict how the technologies are used to benefit educators in the design of their programs.
Inclusion based on the reading of full-text articles	The aim of adding the articles to the present research study is to find how AI technologies support instructional design (centering educators, teachers, and instructors).

Moreover, the reference lists of the 27 articles were also explored for potential in the present research, and 13 more papers were obtained using the same coding methodologies (i.e., title, abstract, and methodology/findings assessment). The final number of papers accepted thus amounted to 23. However, upon reading the articles in their entirety, and thus a closer and deeper inspection of these 23 papers, 7 were excluded due to their diverging focus from the present study or demonstrated ongoing research. The final number of papers amounted to 15. This search criteria are summarized in a flow diagram in Figure 1.

Figure 1.1

Search Criteria and Classification Flow Chart



Chapter 3. Discussion

3.1 Findings

3.1.1 Summary of Studies

In Huang et al. (2021), a comparison between AI technology-infused and common teaching designs demonstrates whether a "paradigm shift" in learning and teaching has occurred. The paradigm has four core principles: promised beliefs, unanimous values, "symbolic generalizations," and imitable examples (Huang et al., 2021, p. 79). The study involved 100 students divided into an experimental (AI) and a control group. Results showed that more students taught using AI technologies had scores in the upper bracket (80-90 percent) than students taught using common design methods (Huang et al., 2021). The experimental group also scored higher in all four categories of learning attitudes and had greater student satisfaction in teaching activities incorporating AI techniques (Huang et al., 2021).

In Jiao et al. (2022), a quantitative prediction model was developed, known as the Evolutionary Computation (EC) technique, to predict the learning performance of 35 graduate students in an online engineering course, Smart Marine Metastructures in the Spring 2020 cohort. The technique used a genetic programming (GP) AI mechanism. The innovative technique employs 8 input variables based on varied assessments and generates a predicted outcome in learning effectiveness (Jiao et al., 2022). The study deemed that learning effectiveness is affected by students' knowledge acquisition, class participation, and summative performance, in that order. The model also demonstrated that prerequisite knowledge and group discussion participation do not impact learning effectiveness (Jiao et al., 2022). The overall contribution of this study determines that student data, both quantitative and qualitative, can be numerically weighted to make quantitative relationships between learning variables (Jiao et al., 2022).

Ouyang et al. (2023) completed a similar study using the same engineering course as Jiao et al., (2022), however, they observed students in the Summer 2022 cohort. The researchers used the same GP-AI model; however, the study integrated a learning analytic (LA) approach to provide the prediction results to the students as visualized feedback (2023). This was followed by a comparative study between the integrated AI-LA approach (experimental) versus using traditional methods of learning (control) on student performance, engagement, and perception (Ouyang et al., 2023). The researchers determined that the experimental approach provided more social engagement opportunities in collaborative learning, higher cognitive engagement and group regulation, as well as facilitated deeper levels of perspectives on their learning (Ouyang et al., 2023). Furthermore, students in the experimental group had better-written outcomes and demonstrated better learning and feedback satisfaction, such as a more positive attitude and perception (Ouyang et al., 2023).

Learning performance in the form of cumulative Grade Point Average (cGPA) was predicted in a study of 1,000 undergraduate students over three intake years at University Q using an Artificial Neural Network (ANN) model (Lau et al., 2019). The study involves a quantitative collection of the students' course entrance results in five subjects, namely, Chinese, English, Math, Comprehensive Science, and an overall proficiency test (Lau et al., 2019). Unlike the other papers retrieved in this literature review, this study determined the effects of gender, family, and socioeconomic backgrounds of students on the cGPA. The study used statistical approaches to generate quantifiable data as 11 variable inputs into the system (Lau et al., 2019). The study showed a gradual improvement in students' cGPA over the three years. The model garnered an 84.8% prediction accuracy. Factors that affected cGPA through the prediction model included English exam results and the mother's occupation. Though females scored better than males, a large false-negative rating for males deems gender classification in the ANN model unreliable (Lau et al., 2019).

Alshammari et al. (2019) constructed an Adaptive E-learning System (AES) by evaluating two adaptations: student learning style (L) and knowledge level (K) (2019). The system provides students with learning materials or learning objects (LOs) that could be leveled as basic, intermediate, and advanced based on three experimental conditions: using only learning style, only knowledge level, or both learning style and knowledge level. The study tested 174 undergraduate students' immediate and delayed learning gain through a difference in pre-tests versus post-tests (immediate) and follow-up tests (delayed). Post-test and follow-up mean scores for the combined group (L and K) were greater than the mean scores of the other experimental groups. The immediate and delayed learning gain was also highest for L and K groups. However, data analysis methods reveal a smaller effect size of the data for immediate learning gain, suggesting that improved adaptive material is required to enhance short-term learning effects (Alshammari et al., 2019).

Furthermore, Chango et al. (2021) used a multimodal approach to testing and finding an algorithm that best predicts student performance and the attributes that contribute to it (2021). In this study, six algorithms were used in three experiments to test 40 undergraduate students in a course about the human circulatory system (Chango et al., 2021). Researchers collected student interaction data from MetaTutorES, an intelligent tutoring system, such as system log data, videos to track facial emotions and gaze data (monitors eye tracking of areas of interest on the screen) (Chango et al., 2021). These attributes determine students' potential for a pass or fail status. The researchers also gathered prior knowledge scores and learning performance scores. The study demonstrated that when an ensemble of algorithm models was applied to the data from selected attributes, the prediction of final student performance improved. In this case, the ensemble classification model that produced the best prediction score was the REPTree algorithm. Furthermore, the study also determined the best attributes for student performance prediction based on their recurring status in the algorithm rules (Chango et al., 2021).

Predictive learning models were also explored through an Early Warning System (EWS), which uses undergraduate students' current progress on a series of activities, visualized through a Green-Amber-Red risk evaluation, to indicate whether they are "at-risk" for failing the course (Baneres et al., 2019). The green indicates a lack of risk, the amber indicates the possibility of risk, the red indicates at-risk, and the black indicates a dropout status (Baneres et al., 2019). The predictive model demonstrated that for the final three assessments, those demonstrating at-risk (red) or dropout (black) status in the EWS failed the course, while those receiving non-at-risk (green) status passed. Furthermore, students reported through a survey that the tool helped them remain motivated, improved their mood, and was useful (Baneres et al., 2019).

Similarly, Franzoni et al. (2020) used visualization techniques through an LMS-embedded learning tool, MonitorView, to gather real-time, quantitative student data in 12 different courses. They used three functionalities: a thermometer bar indicative of resource access recency, a dimensional morphing metaphor indicative of access frequency and level of engagement, and tag cloud morphing based on user access and learning object popularity (Franzoni et al., 2020). A qualitative analysis was conducted on the perspectives of the 12 instructors based on useability and an overall evaluation of the tool. The system's useability and engagement improved with the addition of the metaphors. Moreover, using metaphors allowed educators to adapt their learning and teaching as they receive student analytic data while in the course to improve the quality of their instruction (Franzoni et al., 2020).

Osmanoglu et al. (2020) used their eCampus LMS to conduct sentiment analysis, the understanding of students' feelings about distance education, to improve learning design (2020). The study was conducted on distance education students at Anadolu University, using a data mining approach (collecting data) and applying eight machine learning algorithms to classify and thematically categorize 2421 student comments (Osmanoglu et al., 2020). The data were pre-processed for cleaning, comments made lowercase for uniformity, then spell checking, and finally, stop words, removing unnecessary special characters or objects (Osmanoglu et al., 2020). Various combinations of the three pre-processing mechanisms were applied, and the results demonstrated that the Logistic Regression algorithm produced the highest accuracy in gathering the sentiment value of student feedback, such as positive, negative, or neutral (Osmanoglu et al., 2020). Overall, the study demonstrates that using AI classification methods on online student data can help inform the practice of instructional designers (Osmanoglu et al., 2020).

In Darvishi et al. (2022), a four-part analysis of the effectiveness and reliability of an AI-based peer feedback system was tested. As with Osmanoglu et al. (2020), a comment analysis was done; however, in this study, students' comments were analyzed to determine their viability in critiquing student-created resources in a course (Darvishi et al., 2022). The Recommendation in Personalised Peer-Learning Environments (RiPPLE) learning tool, an adaptive system that inculcates a student-as-a-collaborator approach to learning by allowing students to upload relevant resources to course material, was used (Khosravi et al., 2019; Darvishi et al., 2022). The system, students, and the instructor have the opportunity to vet the resource. Assessment of the resource followed 4 steps: individual reviews, assigning grades, feedback on reviews, and instructor oversight. The study found that such combined models of peer feedback review through AI systems improved student-generated content quality and, thus, built more trustworthy peer assessment systems (Darvishi et al., 2022).

A potential method to improve and simplify the instructional design process is using AI technologies to retrieve LOs that are linked through specific prerequisite components. A study that compared LOs for such prerequisite components was done by Gasparetti et al. (2018). This study used multiple machine-learning approaches on datasets from three online learning platforms: CrowdComp, Udacity, and edX. The study's novelty comes from its machine-learning technique for finding links between the text content of LOs in Wikipedia articles through a categorical feature selection process (Gasparetti et al., 2018). The researchers note that the “metadata” of LOs, that is, their title, description, and keywords within the text content, can be created and managed when identifying features, helping instructional designers more efficiently and effectively design courses with LOs. Furthermore, using multiple features allows a greater potential to find links than basing the comparison on a single feature (Gasparetti et al., 2018).

Another study that explored the use of intelligent mechanisms to retrieve semantically similar LOs from repositories to construct or update a course was conducted by Tahir et al. (2022). Unlike Gasparetti et al. (2018), the researchers used an automated text feature extraction mechanism to retrieve LOs, known as the Dynamic Recommendation of Filtered LOs (DRFLO) (Tahir et al., 2022). The DRFLO system contains three levels: repository, decision, and presentation. The study compared these experimental conditions to a control, in which a common search engine API was used for LO retrieval. Between the experimental and control conditions, time spent on LO retrieval through the individual systems, task completion, number of relevant LOs derived per task, number of query submissions per task, and the number of LOs saved into a repository were compared (Tahir et al., 2022). The study noted that 93.5% of the LO repository was accessed and that the experimental group could find a concentrated number of relevant LOs quicker than the common search engine (Tahir et al., 2022). The task efficiency is also greater in the experimental group, in which the precision and recall of LOs are more accurate. The teacher feedback on this system indicates usability, its unproblematic integration, and continued future use (Tahir et al., 2022).

Crowe et al. (2017) attempted to develop a language-based model, specifically, an all-encompassing artificial augmented intelligence/knowledge-based system AAI/KBS) for scholarly writing software (SWS), which amalgamates various existing writing frameworks. In a qualitative study that interviewed 20 instructional technology and computer science subject matter experts (SME), it was found that IBM Watson's cloud-based application could be used as a prototype based on specific expectations of the proposed model (i.e., a review of grammar and meanings, detection of writing patterns, correlating and adapting content and citations, and alerts for partial answers on discussion questions) (Crowe et al., 2017). Watson is an artificial augmented intelligence and knowledge-based system (AAI/KBS) created by the company, IBM, and is primarily known for participating successfully in the game show, Jeopardy, in 2011. Its creation is intended to promote cognitive computing methods and is continuing to be developed and enhanced (Crowe et al., 2017). However, the study notes the gaps in existing models in demonstrating advanced AI capabilities. Thus, future iterations of the AAI/KBS model would be necessary to develop the optimal prototype that consists of the expected core services of the SWS (Crowe et al., 2017).

In a study on misconception detection and identification (MDI) in 40-grade school students, a dual algorithm machine-learning approach was used to infer students' "learning and forgetting process" (p.87) in English and French language learning acquisition and subsequently provide LOs that suited the student's learning trajectory (Troussas et al., 2019). Two algorithmic mechanisms were used. The Fuzzy String Search technique detects spelling mistakes caused by neglect and lack of knowledge, and the String Interpreting Resemblance algorithm identifies errors due to language confusion and inappropriate use of tense and auxiliary verbs. The latter is known as language transfer errors (Troussas et al., 2019). Using the MDI-based inference system, three learning sessions were conducted, followed by a questionnaire determining the students' experience with the model (Troussas et al., 2019). While the accuracy of error diagnosis through machine learning was 70%, the accuracy of the learning material through the fuzzy logic technique was 68%, and both were deemed "very accurate" by Troussas et al. (2019). The results of this study determined that the MDI-based inference system produced improved learning outcomes for the students, very accurate error diagnosis, can produce adaptive learning material, and hence gives students a learning experience that centers them (Troussas et al., 2019). Through interviews, 4 out of 5 instructors credit the design's ability to improve the learning process (Troussas et al., 2019).

Like the MDI-based inference system, Xu et al. (2019) also used AI as an inference and detection system; however, in conjunction with augmented reality technologies in a study on two 4th-grade students in an English as a Second Language program. The study consisted of two groups, experimental and control, in which the experimental group used Augmented Reality - Artificial Intelligence (AR-AI) technology to complete activities on buildings in their communities and their functions to improve their English learning. The study used mobile devices embedded with the Optical Character Recognition (OCR) AI technology which allows for scanning handwriting using recognition software. The technology provides immediate feedback on the written accuracy (Xu et al., 2019). The students used the mobile device with handwriting recognition to scan cards, which gave them the position of the building, giving them opportunities to apply their spatial learning by attaching the word card correctly onto a physical map. The study, however, focused more on the uses of Augmented Reality technologies, the uses of AI by the students rather than in design, and was conducted in a physical environment (rather than online). However, in an educator questionnaire, the teachers mentioned that the text recognition function in the mobile device made the instructional preparation process more convenient and efficient (Xu et al., 2019).

3.1.2 Common Themes

3.1.2.1 Adaptive Systems

A theme in some of the papers demonstrated the adaptive nature of intelligent systems. The system adapts students' learning trajectories based on how students respond to the LOs; depending on the responses, the system will omit irrelevant LOs (Alshammari et al., 2019). For instance, if the student progresses more at the basic level, it will omit LOs at the intermediate or advanced level. As the student interacts with the system, these iterations will serve as learning until an optimal level is reached, causing the system to discontinue the creation of learning paths (Alshammari et al., 2019).

RiPPLE is another tool that uses adaptive learning through understanding students’ skills and learning styles (Darvishi et al., 2022). The tool’s function demonstrates how “AI-driven learning analytics” can enable the development of a multistep, trustworthy peer feedback system of learning resources. Students can contribute by creating learning resources, after which AI mechanisms are used to assess the resource's reliability through various inputs such as, peer reviews, assessor reliability, and instructor feedback (Darvishi et al., 2022). The reliability of the resource is then improved, or disproved, as the system receives information from peer assessors (Darvishi et al., 2022).

In the study by Troussas et al. (2019), adaptive material is provided based on students' misconceptions about foreign language acquisition. When students make language errors, categorized as transfer or non-transfer errors, the system provides material that helps them improve the skill to determine the "knowledge learned" variable based on their concept performance (i.e., unknown, unsatisfactory known, known, learned) (Troussas et al., 2019). For instance, one student in the study, Phillip, began with 'unknown' as his knowledge level of all the English and French concepts at the first interaction with the system (Troussas et al., 2019). After completing exercises in simple present, his first chosen topic in English, he had to answer 10 additional questions on the same level based on his scores at his next interaction to improve his knowledge level (Troussas et al., 2019).

3.1.2.2 Predictive Systems

Baneres et al. (2019) used the Gradual At-Risk (GAR) model, a predictive system using a set of sub-models, each assigned for an assessment activity (four in total). The system predicts the scores required to pass a course (binary pass/fail variable) based on students' performance in previous activities. Using green, amber, red and black signals, a second general warning level is provided to inform students of their progress (Baneres et al., 2019). Lau et al. (2019) used an ANN-based predictive model to determine student cGPA using factual data such as gender, socioeconomic, geographic, and educational backgrounds, as the researchers believed that cGPA is affected by a variety of factors in a complex manner, and not just through quantitative precursors such as student grades. The mechanisms for this predictive model, in terms of the algorithms used, are complex for this line of study. However, the model demonstrated a reliable prediction accuracy through various statistical analyses (Lau et al., 2019).

In the MDI inference model, after one participant's, Phillip, 12th and final interaction with the system, the system determined Phillip's knowledge level for each English concept (Troussas et al., 2019). However, even though Phillip did not attempt French concepts, the model predicted what his French concept knowledge levels will be (Troussas et al., 2019). Through collaboration with 15 foreign language teachers adept in their understanding of French and English language acquisition, Troussas et al. (2019) developed a model of interdependency between concepts in English and similar concepts in French (i.e., simple present in English with le présent in French, simple past in English with le passé composé in French, etc.). Therefore, with such an inference system, Phillip's number of exercises to complete will be altered from the initial 20 to a different number if his knowledge level is predicted to be other than unknown (Troussas et al., 2019).

In Jiao et al. (2022), the GP-AI model demonstrated that student learning effectiveness is affected by students' learning acquisition or learning gain, their class participation level, and their major summative performance (in this case, a collaborative literature review) in that order. This predictive system demonstrates how quantitative data can support theoretical frameworks to understand students' learning performance and the variables that contribute to it (Jiao et al., 2022). Similarly, Ouyang et al. (2023) also used the GP model to predict student performance to provide students and instructors with a visual representation of the predictions to improve learning and teaching. In this predictive modeling approach, the researchers used similar input variables of prerequisite knowledge, discussion participation frequency, summative procedural performance, and newer variables such as depth of discussion (Ouyang et al., 2023). The predictive model, coupled with a learning analytic feedback system, demonstrated improved student learning outcomes (Ouyang et al., 2023).

3.1.2.3 Feedback Systems

Predictive modeling is purposeful for educators to monitor and for students to observe and maintain their progress. Prediction and feedback have been integrated in some studies to demonstrate AI systems and devices' programming capacity to provide a formative mode of assessment. The AES, for instance, gave students instant feedback based on their interactions with LOs (Alshammari et al., 2019). The system provides recommendations and supplementary information if students do not progress in a specific domain. The feedback shows the set of failed quiz questions, the topics the student should revisit, and the order in which they revisit them to improve their learning and knowledge (Alshammari et al., 2019). This mechanism aims to enhance and augment the students’ use of learning materials and fix misconceptions.

Another example of immediate feedback prompts occurred in the RiPPLE system, which provided students with constructive messages in response to their feedback on a peer’s resource (Darvishi et al., 2022). The system’s algorithms “flag” the participants' comments when submitted to modify and improve textual feedback quality. This comes in the form of a self-monitoring checklist consisting of the comments' alignment with a rubric, specificity to the resource, providing meaningful improvements and being of suggestive nature, and the use of constructive language, as well as an automatic quality control prompt which compares the moderated comment with the previously submitted comment (Darvishi et al., 2022). Though the opportunity to modify their comment was used by only 35.5% of the flagged individuals, students within this percentage provided lengthier comments and they received more "likes" from peers, indicating the helpfulness of their comments. In all four categories of feedback quality, the group using the RiPPLE system for peer feedback produced overall better feedback (Darvishi et al., 2022).

In Ouyang et al. (2023), the feedback from the experimental AI-LA model consisted of four sections: the instructor feedback, the performance prediction scores, a visualization of this process (i.e., a bar graph with their assessment progress against their classmates), and any further learning suggestions. The students demonstrated a positive reception to the feedback; however, they were also critical of the specificity of the feedback. Unlike the AES, which provided students with specific areas of improvement and sections of the course to revisit (Alshammari et al., 2019), the students' feedback perception from Ouyang et al. (2023) required more detail regarding the qualitative and technical aspects of their written work. Furthermore, compared to the control group, more students checked their feedback, believed it was timely, and validated its ability to instigate collaboration, such as asking peers with higher scores for advice (Ouyang et al., 2023). Feedback for tasks was also provided through educator-pre-created feedback prompts for the students in Baneres et al. (2019). These were inputted and administered by the EWS model, providing feedback in the form of next steps or further recommendations when the student has the amber, red, or black light, and encouragement messages when the student is at the green light (Baneres et al., 2019).

Another feedback-oriented system was used in Troussas et al. (2019) using an MDI inference model. For non-language transfer errors, such as spelling mistakes due to lack of knowledge or negligence, a message would be provided to students to be careful. However, feedback would be more specific for language transfer errors, providing students personalized feedback on their errors and presenting them with similar questions to improve their learning (Troussas et al., 2019). Xu et al. (2019) also used a language-processing AI system in which text recognition led to the production of instant feedback or prompts for students when they would match the buildings with their appropriate descriptions. Moreover, the feedback allowed students to complete subsequent tasks more efficiently. Positive feedback would be provided once all the pairing tasks were completed (Xu et al., 2019).

Compared to studies that show AI systems providing students with feedback, Franzoni et al. (2020) study explored how student analytic data in real-time serves as feedback for educators and course instructors. The visualization metaphors indirectly served as a valuable source of feedback on the frequently used learning elements, the least used learning elements, and learning elements frequently used in proximity to major assessments such as assignments and exams (Franzoni et al., 2020). The frequency of learning material used can help instructors determine which to improve, if less used, and what level of complexity the students function at, based on the most used (Franzoni et al., 2020).

3.1.2.4 Student Data as Input

Alshammari et al. (2019) suggested that students benefit the most from the adaptive learning mechanisms when both knowledge level and learning style are considered. Thus, features that demonstrate the two adaptations together (as opposed to singularly) help students gain learning in both immediate and delayed testing conditions (Alshammari et al., 2019). Chango et al. (2021) set out to discover the best variables to help educators better understand student performance potential. The attributes that were most recurring in contributing to prediction scores in the algorithms were ITS log data, such as student abilities to summarize their learning and interaction with learning content for a length longer than 15 seconds, gaze data, specifically attention to avatars and images/graphics relevant to the content, and the surprise emotion from video data (Chango et al., 2021). These findings corroborate the potential for multimodal learning analytic (MLA) data to support student performance prediction and improve potential opportunities for instructors and teachers to act on student learning methods.

Huang et al. (2021) used multiple measurement parameters to determine student cognition and learning ability to improve learning and teaching efficiency. These include academic performance, learning attitudes, emotional experiences, self-cognition, and behavioural tendencies (Huang et al., 2021). The study also measured student satisfaction in activities incorporating AI teaching design through learning interest and enthusiasm, personal acceptance, functional mastery, and improving comprehension and utilization of code blocks (Huang et al., 2021).

In Lau et al. (2019), students' English entrance exam results contributed to a better cGPA score, suggesting the importance of English proficiency. Furthermore, the study also notes the influence of maternal occupation and the lack of impact of paternal occupation on the student's cGPA (Lau et al., 2019). The authors suggest that learning motivation could be an implicit effect of a mother's successful career. A limitation of this study is the inability to classify gender accurately as an attribute or effect on cGPA due to the imbalance of male and female participants (Lau et al., 2019).

Jiao et al. (2022) and Ouyang et al. (2023) explored the determinants of improved student performance using predictive AI models. In Jiao et al. (2022), five input types and eight variables were used in the model: student prerequisite knowledge, discussion participation frequency (in-class and group discussion), procedural performance (write-ups, discussion, and presentations), summative performance (final write-up), and knowledge acquisition (student self-evaluation). Ouyang et al. (2023) used similar input variables; however, they did not include student performances and knowledge acquisition, focusing more on collaborative opportunities such as discussion participation frequency and depth. Both measured learning effectiveness (final learning performance) as the output for the model (Jiao et al., 2022; Ouyang et al., 2023).

When developing an effective instructional design for a course, Jiao et al. (2022) determined that indicators of good or poor student performances in online environments are affected by student self-evaluations, class participation, and instructor evaluation of summative assessments. Hence, the EC model in online learning environments helped quantify, predict, and equate student performance input and output attributes (Jiao et al., 2022). Similarly, Ouyang et al. (2023) also indicate the importance of student self-evaluation or reflection as a critical input to their GP-LA model in their study on student performance, engagement, and perception.

Student perceptions are a key piece in the determinants of instructional design. Osmanoglu et al. (2020) used students' sentiment analysis as feedback on distance education courses to understand how to improve the future instructional design of these courses. Using machine learning algorithms applied to a series of pre-processed data, the study found that Logistic Regression was the algorithm that demonstrated the highest accuracy, 0.775, in terms of analyzing the sentiment of the feedback (Osmanoglu et al., 2020).

3.1.2.5 Instructional Preparation or Instructor Mediation

In a qualitative study on SMEs for the proposed AAI/KBS model, instructional designers played a crucial role in using intelligence systems in the online learning environment (Crowe et al., 2017). Instructional designers must understand the underlying mechanism within AI models. They must also advocate for adopting such intelligence applications of learning for use by distance learning educators while addressing solutions with the collaboration of SMEs and IT specialists (Crowe et al., 2017). By building the system's accuracy through multiple iterations and improvements under the guidance of designers and experts, the importance of human interaction in developing intelligent models is demonstrated (Crowe et al., 2017).

Vetting the platforms and data to determine their reliability is a theme observed in many studies. Darvishi et al. (2022) used instructor oversight as a key piece in determining peer review effectiveness. The RiPPLE platform's "spot-checking" algorithm uses specific metrics, both human-like and data-driven, to "flag" the resources with a quantitative "risk score" (Darvishi et al., 2022). The flagged risk score probes the instructor's intervention to check the resource manually. The platform demonstrated 68% of the resources required instructor intervention based on metrics such as the number of times the resource was reported for inaccuracy or misinformation, the resource's ineffectiveness based on downvotes, disagreement between assessors of the resources, and questionable distractors (e.g., a favourable or popular answer other than the author's) (Darvishi et al., 2022). While the questionable distractor metric received the most flags, low effectiveness showed the most revisions through instructor-mediated grade change for the resource, making it the most effective metric to elicit instructor intervention (Darvishi et al., 2022).

An instructional designer's importance in finding relevant textual similarities between LOs in Gasparetti et al. (2018) was noted. The study emphasized the potential of using a manual coding approach when irrelevant features are extracted from the machine-learning technique, such as a "dictionary" of terms or concepts to which the system will refer (Gasparetti et al., 2018). This would be created manually by domain experts and inputted into the model to improve the metadata retrieval and filter the ineligible learning materials (Gasparetti et al., 2018). This consideration further accentuates the notion of a vetting process through human intervention.

A manual coding of the oral and written work (students' procedural performance) in an engineering course was conducted in Jiao et al. (2022), in which the content was categorized based on three levels: superficial, medium, and deep knowledge. This coding scheme allowed researchers to weigh the procedural performance numerically (Jiao et al., 2022). Similarly, in Ouyang et al. (2023), the depth of students' discussions was also thematically categorized into superficial, medium, and deep knowledge levels and given a weighted score. Furthermore, Ouyang et al. (2023) also focused on researcher intervention in the data analysis stage to determine critical themes in the students' engagement, performance in the final assessment, and perception of the learning experience. Troussas et al. (2019) also used language domain experts to derive a series of "concept interdependencies" between French and English languages in their study (p.90). The influence of knowledge of each language upon the other and the effectual relationship between the two languages, known as "knowledge influence degrees," was determined by 15 foreign language teachers (Troussas et al., 2019, p.90). The number and difficulty levels of the exercises administered in the study were also defined by the teachers first (Troussas et al., 2019).

The proposed DRFLO system required teacher or user-generated data as inputs in order to be provided with the required LOs. At the repository level, the system derived information about the technical aspects of course building specific to the user's preference (i.e., format, subject, keywords, etc.), user logs, and the user's feedback on the LO (Tahir et al., 2022). In the decision layer, the system undergoes a multitude of machine-learning algorithms to extract features, map them based on user preference, and generate a ranked order of LOs (Tahir et al., 2022). The layer also involves the quantification of learning preference and a rating of the LO from a multitude of other teachers in the same course, known as collaborative filtering (Tahir et al., 2022). These factors help to generate a recommender engine that recommends LOs to the users. This recommendation system is shown in the final presentation layer (Tahir et al., 2022). User interactivity occurs with the dashboard (which consists of the course, the weekly lessons, and the learning objects); the user access logs (showing a plethora of quantitative data concerning the retrieved LOs used in the decision layer); and a query engine (in which the user can submit and re-submit queries till satisfaction is attained with the designed course LOs) (Tahir et al., 2022).

3.1.2.6 Visualization Models

In Baneres et al. (2019), feedback allowed students to visualize their progress. The first means of visualization is through a stack of progress bars for each activity. After each graded activity, the student is given a projected grade required in the next assessment to pass the course. The algorithm-based Gradual At-Risk (GAR) model was used to process this visualization (Baneres et al., 2019). Visualization also came in a general form through the Green-Amber-Red risk evaluation. Students and educators have dashboards depicting the warning indications based on the coloured signal. The visualization dashboards were also present for the course instructors, who had access to individual student progression for all four assessments and a graphical representation of the overall progression of the warning levels (Baneres et al., 2019).

Franzoni et al. (2020) used AI-based visual metaphors to help quantify and visualize student data. Of the three metaphors, the dimensional morphing metaphor shows the most access, followed by the tag cloud and the thermometer bar. Its useability, capacity to improve quality, and impact on course management received the highest ratings in the qualitative assessment (Franzoni et al., 2020). Though the thermometer bar provides a more organized appearance, teacher feedback suggests it to be less appealing and provides more effort to read and compare between LOs (Franzoni et al., 2020). In comparison, the tag cloud metaphor is more attention demanding. The dimensional morphing metaphor was most valued and effective according to quantitative and qualitative assessments due to its useability from simple visual cues instead of statistical data interpretation and analysis on the part of the educator (Franzoni et al., 2020).

In the study by Ouyang et al. (2023), comprehensive feedback was provided to the experimental groups after every group write-up submission. The feedback included a visual bar graph or chart to show the predictive data for the four “process-oriented data” variables or inputs as separately coloured bars for each student in the group (labeled on the x-axis), which allowed students to visualize their contributions against other group members (Ouyang et al., 2023).

3.2 Implications of Findings

This study aimed to determine how to use ethical AI-based applications to co-design learning materials within virtual learning settings. Within the thematic analysis of the research found in this study, it was determined that various factors corroborate successful uses of AI in the design of online educational environments.

The majority of the studies in this research used students as the sole or primary participants within AI-influenced instructional design conditions (Alshammari et al., 2019; Baneres et al., 2019; Chango et al., 2021; Franzoni et al., 2020; Huang et al., 2021; Jiao et al., 2022; Lau et al., 2019; Osmanoglu et al., 2020; Ouyang et al., 2023; Troussas et al., 2019; Xu et al., 2019). The studies demonstrated the potential for improved student performance and perception of learning. For instance, Phillip's effective language acquisition was coupled with improved efficiency in time to complete pending concepts (Troussas et al., 2019). Though student results are not the primary focus of this review, the percentage accuracy of error diagnosis and the accuracy of learning material delivery through these models demonstrate its potential in instructional design to elicit optimal student output (Troussas et al., 2019). Furthermore, student performance projection and trajectory prediction are also a large component of the AI systems used. The capacity for education to be predictive eases the manual load of educators to foresee student progress and use automated approaches to address the needs of the student in a timely, visual manner for student-initiated improvement.

AI-based applications may be used to generate student-performance-oriented feedback which could be pre-inputted by the educator or constructed based on the analysis of student data such as prior knowledge performance, students' discussion levels, student assessment, non-academic factors, and sentiments. The research also demonstrates that feedback generated immediately is valuable to student success. Furthermore, feedback immediacy allows students to "act" on the feedback and address the misconceptions through re-familiarization with academic concepts, receiving attempts from the system to continue working on the skill, seeking collaboration from peers for further tips and suggestions, or adopting better learning strategies for the following assessment.

Through feedback, the predictive, and visual nature of the models, the systems create a repository of student data which in online environments can be beneficial. In Baneres et al. (2019), the results suggest the potential for visual, predictive models to improve student learning behaviours through motivations, moods, and tools. The representations allowed early intervention on the part of the educators, allowing opportunities to provide intermittent guidance and feedback and increase the potential for improved student outcomes. Further, students' appreciation of this visualization in an online setting demonstrates the opportunity to bridge any gap of personalization.

However, AI models function through quantifiable data; hence, conversion of qualitative data or the normalization of quantitative data (i.e., conversion to a value between 0 and 1) is necessary. This requires the intervention of machine experts. Predictive models, feedback-generating models, adaptive models, visualization, and input of quantifiable, normalized student data demonstrates the importance of machine learning experts in the design of the instruction, in the administration of the instruction at an observational level and to vet the system if needed.

The potential of developing a highly efficient LMS for online learning environments through AI systems, specifically gathering appropriate and relevant LOs, demonstrates how AI systems can assist in the development of appropriate courses (Alshammari et al., 2019; Gasparetti et al., 2018; Tahir et al., 2022). In Tahir et al. (2022), teachers or users are at the helm of the study, and their efficiency in course building and updating is the valued output or intention for the proposed

DRFLO intelligent system. By using the query search feature on their dashboard, educators can continually interact with the system to retrieve the appropriate LOs to construct their course (Tahir et al., 2022).

Moreover, the EWS was a useful tool for educators as they were able to visibly observe student progress throughout the course using the dashboard (Baneres et al., 2019). In Franzoni et al. (2020), the three metaphors used for student analytic data demonstrated the opportunity for educators to use the functionality of the LMS better. For instance, when observing their instructor logs from the previous academic year, it was determined that only five out of the 12 instructors were using the Moodle log report utility function. In contrast, Franzoni et al. (2020) showed that four out of five teachers had shifted to a new metaphor, while one maintained the Moodle log report. These metaphors or features augment the learning in online, virtual environments, where technology promotes a more efficient means of creating and observing student data on LMS.

Moreover, some studies have demonstrated how LMS could be used in an improved method by additive AI features, such as through MonitorView's use of the three tools (Franzoni et al., 2020), the query search engine in the DRFLO intelligent system (Tahir et al., 2022), and the EWS system on the dashboard for both teachers and students (Baneres et al., 2019). Studies have also demonstrated the use of existing LMS to be analyzed using machine learning methods such as using quantified student data from the eCampus LMS to conduct a sentiment analysis (Osmanoglu et al., 2020) or using Genetic Programming AI model to input quantified data from the Blackboard LMS to predict the learning performance of students in an online environment (Jiao et al., 2022; Ouyang et al., 2023). Though they have set out to collect data based on a specific AI function (i.e., feedback, prediction, adaptive learning, LO retrieval, etc.), the studies can help to envision a multilayered approach to educational design with AI-based LMSs. These systems can work in conjunction with one another, working at each phase and/or section of the learning process to create a more complex system that benefits all education stakeholders.

3.3 Conclusions and Future Work

The overview of existing research on AI systems and machine-learning algorithms provides only a foundation for understanding the functionality of the systems from an educational perspective. As an educator, my limitations in the conceptual understanding of AI technologies and their diversity perhaps testify to the importance of collaborating with varied experts in order to truly understand how to augment educational design using AI. Although learning and teaching continue to involve teachers and students as stakeholders, AI system experts may also be valuable as educational models evolve. The inclusion of appropriate expert resources is vital to not only adopting more 21st-century learning methods but also improving opportunities for instructional design collaboration. Perhaps the new norm for instructional design will be the collaboration between educators and/or instructional designers, subject-matter experts, and AI engineers specific to the educational platforms, which may help to create an optimal AI-based educational system for online learning. It is valuable for future studies to highlight the importance of this collaboration.

Further, the growing capacity of technology in education must be wholly understood by educators. Thus, opportunities for collaboration may also evolve the knowledge and functionality of the educator in the online realm. The potential for educators to become more AI literate and build skills to help them monitor and understand the workings of these systems grows. This may encourage hierarchical educational policy changes, eliciting a more controlled, shared responsibility toward AI uses in educational environments. This research study aims to demonstrate how AI systems can be utilized by educators in online environments specifically to improve education by developing more efficient design processes that have the potential to augment student performance. The long-term acceptance of such AI technologies in instructional design can possibly alter potential preventative, resistive, or problem-based standpoints of AI in online education and make educators beneficiaries of education by improving the efficiency and effectiveness of their instructional abilities and materials.

REFERENCES

Alshammari, T. M., & Qtaish, A. (2019). Effective adaptive e-learning systems according to learning style and knowledge level. *Journal of Information Technology Education, 18,* 529–547. https://doi.org/10.28945/4459

Baneres, D., Rodriguez, M. E., & Serra, M. (2019). An early feedback prediction system for learners at-risk within a first-year higher education course. *IEEE Transactions on Learning Technologies, 12*(2), 249–263. https://doi.org/10.1109/TLT.2019.2912167

Bedregal-Alpaca, N., Tupacyupanqui-Jaén, D., Delgado-Barra, L., Guevara, K., & Laura-Ochoa, L. (2022). Instructional design for a virtual teaching-learning environment (VTLE): Process, structure and validation by experts. *Journal of Information Systems Engineering & Management, 7*(4). https://doi.org/10.55267/iadt.07.12535

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*(1), 43–18. https://doi.org/10.1186/s41239-023-00411-8

Chango, W., Cerezo, R., Sanchez-Santillan, M., Azevedo, R., & Romero, C. (2021). Improving prediction of students' performance in intelligent tutoring systems using attribute selection and ensembles of different multimodal data sources. *Journal of Computing in Higher Education, 33*(3), 614–634. https://doi.org/10.1007/s12528-021-09298-8

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, *8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Chocarro, R., Cortiñas, M., & Marcos-Matás, G. (2023). Teachers' attitudes towards chatbots in education: A technology acceptance model approach considering the effect of social language, bot proactiveness, and users' characteristics. *Educational Studies*, *49*(2), 295-313. https://doi.org/10.1080/03055698.2020.1850426

Crowe, D., LaPierre, M., & Kebritchi, M. (2017). Knowledge based artificial augmentation intelligence technology: Next step in academic instructional tools for distance learning. *TechTrends*, *61*(5), 494–506. https://doi.org/10.1007/s11528-017-0210-4

Darvishi, A., Khosravi, H., Sadiq, S., & Gašević, D. (2022). Incorporating AI and learning analytics to build trustworthy peer assessment systems. *British Journal of Educational Technology*, *53*(4), 844–875. https://doi.org/10.1111/bjet.13233

Davey K. (2023, September 1). *Learning objects.* docebo. https://www.docebo.com/glossary/learning-objects/

David, S. A. (2013). A critical understanding of learning management system. *Education India Journal: A Quarterly Refereed Journal of Dialogues on Education*, *2*(4), 4-12.

Debattista, M. (2018). A comprehensive rubric for instructional design in e-learning. *The International Journal of Information and Learning Technology*, *35*(2), 93-104. https://doi.org/10.1108/ijilt-09-2017-0092

Demszky, D., Liu, J., Hill, H. C., Jurafsky, D., & Piech, C. (2023). Can automated feedback improve teachers' uptake of student ideas? Evidence from a randomized controlled trial in a large-scale online course. *Educational Evaluation and Policy Analysis*. https://doi.org/10.3102/01623737231169270

Franzoni, V., Milani, A., Mengoni, P., & Piccinato, F. (2020). Artificial intelligence visual metaphors in e-learning interfaces for learning analytics. *Applied Sciences, 10*(20), 7195–. https://doi.org/10.3390/app10207195

Gasparetti, F., De Medio, C., Limongelli, C., Sciarrone, F., & Temperini, M. (2018). Prerequisites between learning objects: Automatic extraction based on a machine learning approach. *Telematics and Informatics*, *35*(3), 595–610. https://doi.org/10.1016/j.tele.2017.05.007

Gopo, C. (2022). The role of technology in the 21st century education of learners. *The Official Research Journal of Tagum City Division.*

Huang, J., Shen, G., & Ren, X. (2021). Connotation analysis and paradigm shift of teaching design under artificial intelligence technology. *International Journal of Emerging Technologies in Learning, 16*(5), 73–86. https://doi.org/10.3991/ijet.v16i05.20287

Jiao, P., Ouyang, F., Zhang, Q., & Alavi, A. H. (2022). Artificial intelligence-enabled prediction model of student academic performance in online engineering education. T*he Artificial Intelligence Review, 55*(8), 6321–6344. https://doi.org/10.1007/s10462-022-10155-y

Kavitha, V., & Lohani, R. (2019). A critical study on the use of artificial intelligence, e-Learning technology and tools to enhance the learners experience. *Cluster Computing*, *22*, 6985–6989. https://doi.org/10.1007/s10586-018-2017-2

Khosravi, H., Kitto, K., & Williams, J. J. (2019). RiPPLE: A crowdsourced adaptive platform for recommendation of learning activities. *Journal of Learning Analytics*, *6*(3), 91-105. https://doi.org/10.18608/jla.2019.63.12

Khosravi, H., Sadiq, S., & Gasevic, D. (2020). Development and adoption of an adaptive learning system: reflections and lessons learned. *SIGCSE '20: Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 58-64. https://doi.org/10.1145/3328778.3366900

Kim, N. J., & Kim, M. K. (2022). Teacher's perceptions of using an artificial intelligence-based educational tool for scientific writing. *Frontiers in Education*, *7*, 755914. https://doi.org/10.3389/feduc.2022.755914

Lau, E. T., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences*, *1*(9), 982–. https://doi.org/10.1007/s42452-019-0884-

Meikleham, A., & Hugo, R. (2020). Understanding informal feedback to improve online course design. *European Journal of Engineering Education*, *45*(1), 4–21. https://doi.org/10.1080/03043797.2018.1563051

Ng, D. T. K., Leung, J. K. L., Su, M. J., Yim, I. H. Y., Qiao, M. S., & Chu, S. K. W. (2022). *AI Literacy in K-16 Classrooms.* Springer International Publishing. https://doi.org/10.1007/978-3-031-18880-0

OpenAI. (2023). *ChatGPT* (Sept 25 version) [Large language model]. https://chat.openai.com/chat

Osakwe, I., Chen, G., Whitelock-Wainwright, A., Gašević, D., Pinheiro Cavalcanti, A., & Ferreira Mello, R. (2022). Towards automated content analysis of educational feedback: A multi-language study. *Computers and Education. Artificial Intelligence*, *3*, 100059–. https://doi.org/10.1016/j.caeai.2022.100059

Osmanoglu, U. O., Atak, O. N., Caglar, K., Kayhan, H., & Can, T. C. (2020). Sentiment analysis for distance education course materials: A machine learning approach. *Journal of Educational Technology and Online Learning, 3*(1), 31-48. http://dx.doi.org/10.31681/jetol.663733

Ouyang, F., Wu, M., Zheng, L., Zhang, L., & Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education, 20*(1), 4–4. https://doi.org/10.1186/s41239-022-00372-4

Sahin, M., & Yurdugül, H. (2022). Learners' needs in online learning environments and third generation learning management systems (LMS 3.0). *Technology, Knowledge and Learning, 27*(1), 33–48. https://doi.org/10.1007/s10758-020-09479-x

Sayed, W. S., Noeman, A. M., Abdellatif, A., Abdelrazek, M., Badawy, M. G., Hamed, A., & El-Tantawy, S. (2023). AI-based adaptive personalized content presentation and exercises navigation for an effective and engaging E-learning platform. *Multimedia Tools and Applications, 82*(3), 3303–3333. https://doi.org/10.1007/s11042-022-13076-8

Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education*, *18*(1), 54–54. https://doi.org/10.1186/s41239-021-00292-9

Shoufan, A. (2023). Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey. *IEEE Access*, *11*, 38805–38818. https://doi.org/10.1109/ACCESS.2023.3268224

Srivastava, A., Saini, S., & Gupta, D. (2019). Comparison of various machine learning techniques and its uses in different fields. *2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA), India*, 81–86. https://doi.org/10.1109/ICECA.2019.8822068

Tahir, S., Hafeez, Y., Abbas, M. A., Nawaz, A., & Hamid, B. (2022). Smart learning objects retrieval for e-learning with contextual recommendation based on collaborative filtering. *Education and Information Technologies, 27*(6), 8631–8668. https://doi.org/10.1007/s10639-022-10966-0

Troussas, C., Chrysafiadi, K., & Virvou, M. (2019). An intelligent adaptive fuzzy-based inference system for computer-assisted language learning. *Expert Systems with Applications*, *127*, 85–96. https://doi.org/10.1016/j.eswa.2019.03.003

Xu, J., He, S., Jiang, H., Yang, Y., & Cai, S. (2019). Design and implementation of an English lesson based on handwriting recognition and augmented reality in primary school. *International Conference e-Learning.* https://files.eric.ed.gov/fulltext/ED621582.pdf

Xu, W., & Ouyang, F. (2022). A systematic review of AI role in the educational system based on a proposed conceptual framework. *Education and Information Technologies*, *27*(3), 4195–4223. https://doi.org/10.1007/s10639-021-10774-y

Zhao, R., Zhuang, Y., Zou, D., Xie, Q., & Yu, P. L. H. (2023). AI-assisted automated scoring of picture-cued writing tasks for language assessment. *Education and Information Technologies*, *28*(6), 7031–7063. https://doi.org/10.1007/s10639-022-11473-y

APPENDICES

Appendix A. Literature Summary Chart

Studies/Authors/ Publication Year	Participants/Age, grades/gender/geo-locations	Data collection and analysis methods	Major results
Alshammari et al., 2019	- 174 undergraduate students (102 males, 72 females) - Information and Computer Science - College of Computer Science and Engineering, University of Hail, Saudi Arabia - 3 experimental groups: L, K, and L+K - 19-25 years	- Learning style (L) and knowledge level (K) measured using Adaptive E-learning System (AES) - AES displays learning objects (LOs) based on students' learning level - Learning style measured through ILS questionnaire; 4 dimensions with 11 questions each (each dimension for a learning style) -Learning gain: pre-, post- (immediate), and follow-up (delayed) tests with 22 questions each (MCQ) Cronbach's Alpha used to observe reliability of questionnaire and tests ANOVA	- Similar prior knowledge of all three groups - Post-test (immediate) mean score highest for L+K group, then K group, then L group - Follow-up tests' (delayed) score shows same pattern - Learning gain = Post-test score - pre-test score - Immediate learning gain highest for L+ K, then K, then L - Delayed learning gain followed the same pattern as immediate learning gain - Delayed learning gain and follow-up test had medium to larger effect size - Need for adaptive material to enhance short-term learning effects
Baneres et al., 2019	<u>Training set:</u> - 2016 Fall and 2017 Spring semester courses - 608 courses in total - 418 undergraduate and 190 Master's <u>Testing set:</u> - Computer Fundamentals	- Detecting at-risk students using a GAR model (Gradual at-risk) through Naive Bayes Classifier algorithm - Early Warning System (EWS): predictive analytic dashboard using green, amber, red for warning indication	- AA1: only yellow warning level given (low GAR model quality) - AA2, AA3, and AA4: students receiving red and black failed the course and most receiving green passed (for AA4, 100% of the students that received black failed) - Green signal students: Prediction software is useful, motivation to continue, positive mood - AA2, yellow signal: less motivation due to potential of failing

Studies/Authors/ Publication Year	Participants/Age, grades/gender/geo-locations	Data collection and analysis methods	Major results
	undergraduate course (first year) -2018 Spring -537 students	- Educator dashboard for student progression - 4 activities (AA1, AA2, etc.) - 5-point Likert scale questionnaire for student opinion on EWS - Survey detecting the reasons to fail or for non-submissions - Two-sample Wilcoxon tests	- Red/black colours: usefulness, motivation, and mood are lower, usefulness is larger than 50%, reduced continuity as the course progresses - AA4, red signal: higher values for usefulness, motivation, and mood; high motivation due to "last chance" - Survey: Failure mostly due to inability to calculate workload, family commitments, and professional commitments
Chango et al., 2021	40 undergraduate students (17 males and 23 females) Mean age is 23.58 years Public university, Northern Spain First-years, various programs -Students learning about circulatory system (for this study)	- 3 experiments to find best algorithm for highest prediction value - Predicting student performance from multimodal data (testing algorithms for best prediction) using WEKA data mining tool <u>Data collection:</u> - Learning data from MetaTutorES logs - Videos (facial emotions) using Microsoft API classifying in 8 emotion categories or classes - Eye tracking and gaze data focus on areas of interest (AOI) - Pret-test MCQ (prior knowledge) - Post-test MCQ (performance value score from 0-10 about learning) K-fold cross validation Area under ROC curve (AOC)	- 3 experiments to find best algorithm for highest prediction value; could not find a single-best one - According to Accuracy and AUC scores, REPTree algorithm has the best prediction value though - Variables and attributes recurring in the three experiments (IF-THEN rules) were: summarizing of content strategies and coordinating information sources such as between image and text (ITS logs), AOI2 attention on instructional avatars and AOI3 attention to images/graphics of learning content (eye tracking/gaze data), and "surprised" facial emotion (video) - Values of PASS/FAIL dependent on value of AOC/ACC

Studies/Authors/ Publication Year	Participants/Age, grades/gender/geo-locations	Data collection and analysis methods	Major results
		ACC (accuracy measure)	
Crowe et al., 2017	- 20 Instructional Technology and Computer Science subject matter experts (SMEs)	- Qualitative study that determines how the combination of a variety of existing tools can be amalgamated into an all-encompassing artificial augmented intelligence/knowledge-based system (AAI/KBS) - Developing a prototype for SWS (scholarly writing system or software) based on Watson Cloud intelligence software for distance learning - Interviews of SMEs - Collection of documentation - 4 research questions asked of 20 participants NVivo software (Determining attributes/nodes and transcription)	<u>Expectations of AAI/KBS model:</u> - Augmentation to Microsoft Word to be reviewed through semantics and syntactics - Writing content patterns detections - Altering content based on citations - Answering of questions or checking meaningful discussion on the topic <u>Gaps when producing the software:</u> - Pattern recognition through provision of a multitude of examples as opposed to only a couple - Machine becomes intelligent through iterations and can produce results much quicker - SMEs help to ensure that the input is not obscure or ambiguous so that Watson doesn't provide false-positives - ID may need to know the underlying mechanisms of the AAI/KBS to be able to collaborate with SMEs and the IT specialists

Studies/Authors/ Publication Year	Participants/Age, grades/gender/geo-locations	Data collection and analysis methods	Major results
Darvishi et al., 2022		- RiPPLE learning tool is an adaptive learning system that modifies learning level or instruction based on student skill and learning style - Student-created resources added to the system has to go through a peer-assessment procedure - 4 processes to assess the resource: Individual reviews, Assigning grades, Feedback on reviews, Instructor oversight	
	374 participants 2 undergraduate courses: The Brain and Behavioural Sciences (n=234) and Introduction to Information Systems (n=140) control group (n=187)	<u>Individual Reviews:</u> Experimental group used the self-regulation checklist paired with AI-generated prompts from 3 separate "quality control functions" Manual coding of 10% of comments based on textual feedback quality (alignment, specificity, suggestion, constructive) Chi-square test Cohen's Kappa Coefficient Cramer's V	Training and self-regulation checklist benefited half the students Longer comments by experimental group; not making a correlation between length and quality Experimental group receives more "likes," from peers indicative of helpfulness In all four categories of textual feedback quality, experimental group has a higher percentage than control

Studies/Authors/ Publication Year	Participants/Age, grades/gender/geo-locations	Data collection and analysis methods	Major results
	- Data from 10 courses from University of Queensland (1 semester) in 2021 - 2837 undergraduate students	<u>Assigning Grades</u> - Assessment of quality of student-generated resources using 1-5 Rating scale (3 is min. for approval, 5 is considered outstanding) - 4 models used to assess reliability of assessor (Expectation-maximisation, Trust propagation, Comment length, and relatedness) -AUC, ACC	- Majority of the students are "easy graders" → instructor oversight is required to assess peer assessments - A combined model (i.e., length and relatedness) demonstrates more success than when using the features alone - It could depict better effort and critical thinking on the part of the assessor - Best AUC and ACC outcome is through a combination of the three models (trust propagation, length, and relatedness)
	- 1348 users	<u>Feedback on Reviews</u> - Assessors and assessees will evaluate reviews and ratings separately by liking or disliking the comments - Participants can decide whether they agree/disagree with the outcome; participants can also reconsider - Anonymous comments can be offered additionally about the process Cohen's Kappa Coefficient	- Through an iterative process, developed a codebook of feedback data that resulted in 5 major themes: peers, themselves, outcome, system, and assessed resource - Primary topic of feedback for assessor and assessee is resource - Secondary: Assessee gave themselves feedback but assessors "focused on peers" (p.858) - 80% of students in agreement with outcome - Opportunity for discussion on peer contributions, changes on results, resource quality and system functionality
	- 1328 cases of instructors reviewing flagged items - data from 10 different courses	<u>Instructor oversight</u> - RiPPLE has spot-checking algorithm using human- and data-driven metrics to flag the resources with a risk score (0 → 5) with 5 being the highest risk	- 32% clearing of flags through spot-checking (no instructor intervention required) - 68% efficacy of algorithm to demonstrate the need for instructor oversight - 42% got a new grade based on revision - Questionable distractors and assessors' disagreement are the most used metrics for flagging

Studies/Authors/ Publication Year	Participants/Age, grades/gender/geo-locations	Data collection and analysis methods	Major results
		- Instructor intervention is based on the risk score	- Metric efficacy: low effectiveness actually has a 89.3% revision rate, followed by users' reports at 81.9%
Franzoni et al., 2020	- University of Perugia - 12 courses, different bachelor degrees, 12 course instructors - Each course had 90-120 students - Age range: mostly 18-25 years	- Quantitative analysis of instructor and student access logs using MonitorView in Moodle LMS - Comparing means of data analytics before 3 visualization metaphors are added (traditional log reports) to after (control vs. experimental) - 3 metaphors for visualization: thermometer bar, dimensional morphing, and tag cloud morphing - Qualitative analysis of experience: instructor questionnaire (useability) and overall evaluation of tool using a 5-point Likert scale	- Dimensional morphing metaphor is prevailing among the 3 - Thermometer bar easy to read and organized but didn't have the appeal of the other metaphors according to teacher feedback; instead the tag cloud metaphor did better - Length of the thermometer bar requires more effort in observation - Valuable indirect feedback through the use of metaphors which allow instructors to adjust and plan along the way as they receive information about what is more or less accessed ("refine strategy [and]... basing decisions on student engagement" (p.17) - Cannot control for the use of external tools besides the LMS's learning objects to share and retrieve some course information (limitation of access log information)
Gasparetti et al., 2018	N/A	- Comparing two LOs' "text content" to find a prerequiste component through Natural Language Processing (NLP) - Using a unique machine-learning approach to find most important attributes	- What affects the identification process (sensitivity) is when LOs are characterized by 2 characteristics - When LO's are short it can be challenging to make relationships with other LO's - Sometimes irrelevant results are retained or extracted affecting the accuracy of the system

Studies/Authors/ Publication Year	Participants/Age, grades/gender/geo-locations	Data collection and analysis methods	Major results
		- Datasets of different domains from 3 platforms (CrowdComp, Udacity, edX) - Random LO pairs sampled and experts will gauge whether there are potential prerequisites - Performance assessment is valued at 5 measures: precision, recall, F1-measure, accuracy, and area under curve (AUC) - Paired t-test to compare approaches	- Approaches that use multiple features can determine prerequisites better than those using only a single feature - The paper highlights that a manual annotation step may be needed (through a domain expert such as an instructional designer) in which there is a “dictionary” to which the system makes a reference and discards any ineligible LO
Huang et al., 2021	- 100 students - Teaching design varies for each group - Experimental group uses AI technology whereas the control group uses more “common teaching design methods”	- Testing whether AI has made a “paradigm shift” in teaching design - Paradigm shift explained by 4 aspects: promised beliefs, unanimous values, “symbolic generalizations”, and imitable examples - Seeing how AI teaching design “verifies” its achievements - Measuring academic performance, learning attitude, emotional experience, self-cognition, and behavioural tendency, and teaching satisfaction - Standard deviation - Significant difference index value	- Experimental group students with scores of 80 or more are significantly greater in proportion than those in the control group - The scores of experimental group students higher in learning attitude, emotional experience, self-cognition, and behavioural tendency than in control groups - Application to teaching activities measured through student satisfaction (i.e., interest and enthusiasm, acceptance, mastery of material, and deepening of understanding and application); greater than in control group

Studies/Authors/ Publication Year	Participants/Age, grades/gender/geo-locations	Data collection and analysis methods	Major results
Jiao et al., 2022	- 35 graduate students in an ocean engineering course, "Smart Marine Metastructures" - Ocean College, Zhejiang University, China - 22-27 years of age - 11 females and 24 males	- Quantitative prediction model to predict learning performance of students in online environments - Optimizing design based on analysis of input variable contributions to academic performance - Blackboard for LMS and DingTalk (like Zoom in China) - Peer reflections and students' self-reflection act as assessments as well - Questionnaire evaluating student pre-course knowledge with a 5-point scale - 5 input variables in genetic programming (GP) model (prerequisite knowledge, participation frequency, procedural performance, summative performance, knowledge acquisition) and 1 output variable (learning effectiveness)	- Learning effectiveness is affected by knowledge acquisition, followed by class participation and then summative performance (in that order according to the GP model) - Prerequisite knowledge doesn't play a role - Instructional design of online courses should consider student self-evaluation, discussion participation and instructor's summative evaluation → indicators of good or poor performance - Group discussion participation not an indicator of student performance - Students' self-evaluation the most critical
Lau et al., 2019	- 1000 students (275 female and 810 male) - Undergraduate students - University Q (China) - Intake year between 2011 to 2013	- Artificial Neural Network (ANN) modelling to predict student CGPA using data from their course entrance results and their socioeconomic background - Quantitative and factual data collection	- Gradual improvement of students' CGPA (2011 to 2013) - Female students score better than male students - No significant different in the CGPA of students belonging to different settlement types (i.e., urban vs. rural) - Students scoring well in their English exam have better CGPA - Effect of mother's occupation (motivation) but not father's

Studies/Authors/ Publication Year	Participants/Age, grades/gender/geo-locations	Data collection and analysis methods	Major results
		- Two sample T-tests and ANOVA (between gender and background of the student to their CGPA) - Pearson correlation coefficients to measure the relationship between the 5 entrance examination subjects (Chinese, English, Math, Comprehensive Science and Proficiency Test) and CGPA	- ANN has a overall "good prediction accuracy of 84.8%" (p.8) - Area under curve (AUC) value is 0.86 (1 is perfect) hence demonstrating "sufficient success" - Poor classification according to gender potentially due to sample imbalance (more males to females)
Osmanoglu et al., 2020	- Distance education students at the Open Education Faculty at Anadolu University	- Machine learning algorithms tested using data gathered from eCampus LMS platform to understand student feelings to improve design - Triple Likert scale (3 points) to "tag" comments about the eCampus application (bad, suggestive, and good) - 2421 comments analyzed (70% used for training, 30% for testing) - Pre-processing of data before inputting into 8 algorithms → clean text conversion to lower case, spell-checker, clearing of stop words (special/unnecessary characters) (individually, in pairs, all three, none) - Logistic Regression analyses	- Logistic Algorithm produces most accuracy after clean text and spell-checker is applied (0.775) - No significant difference in the success rate after CT and SC corrections than before applying the operations indicating the success of the algorithm in processing the comments - Helpful to improve the "social dimension" of LMS; providing better learning experiences - Creating a model without the correction process as it takes time

Studies/Authors/ Publication Year	Participants/Age, grades/gender/geo-locations	Data collection and analysis methods	Major results
Ouyang et al., 2023	- 8-week online graduate level course - Course: "Smart Marine Structure" - Online using DingTalk platform - China - 62 students (43 Master's and 19 Doctoral students) - 7 groups in control and 8 groups for experimental	- Using a Genetic programming (GP) AI model to predict student performance, engagement, and perception - Used LA visualization to give feedback - Control: Manual instructor feedback about written work - Model input variables: prerequisite knowledge, discussion participation frequency, discussion depth, and write-up performance - Model output variable: learning effectiveness (performance) - Social network analysis, content analysis and lag sequential analysis, writing assessment (rubrics and t-tests), thematic analysis - Cohen's Kappa	Experimental and control groups had similar outlook towards online collaborative experiences All students appreciated timely feedback <u>Experimental group (Integrated AI-LA approach):</u> - Demonstrated overall better written performances - More cognitive engagement and group regulation - Maintain more social engagement within groups - 28 students had a positive attitude about their feedback while 12 students thought they should have feedback providing *how* they can improve their written work - More reflective on their collaborative learning experience (i.e., making self-reflections, having more communication, making improvements) <u>Control group:</u> - 25 students checked their feedback - Fewer students expressed a positive attitude; more students required a more thorough feedback or suggestions for improvement

Studies/Authors/ Publication Year	Participants/Age, grades/gender/geo-locations	Data collection and analysis methods	Major results
Tahir et al., 2022	- 2 groups: Control group will create a new course/update old course manually whereas Experimental group will give the task to create a new course/update old course - 15 instructors (mix of lecturers and assistant professors) - Experts (teaching same course for 2 semesters prior) and novice	- Using an intelligent Dynamic Recommendation of Filtered LOs (DRFLO), a 3-layer system, in comparison to common search engine API to create/update a course through LO retrieval Measure through: - Using DRFLO prototype (measuring number of LO's retrieved) - Using statistical factors to compare time spent on LO retrieval, task completion, number of LOs derived per task, number of queries submitted/resubmitted by course designer per task, and number of LOs saved and used - Using machine-learning measures like precision, recall, F1-measure, accuracy by inputting dataset from various universities and course domains - Using qualitative analysis through Likert scale (1-5 rating) on DRFLO usefulness - T-test, Cohen's K, Cronbach alpha, SUS matrix	- 93.5% of records successfully accessed - Control group took longer to retrieve LOs; Experimental group found relevant LOs quicker through DRFLO system - Experimental group received "dense" number of LOs for developing course topics - DRFLO application saves more time when creating a new course and updating an old course - Task completion rate higher in experimental group, increased precision with system improves experimental group efficiency - Precision and recall of LO retrieval greater in DRFLO system; 7% more accurate - Continued future use, easy-to-use, lack of complexity, lower prior knowledge to use system, well-integrated, and confidence in conducting tasks all reported in majority w.r.t. DRFLO system - Positive feedback from participants in the features' useability

Studies/Authors/ Publication Year	Participants/Age, grades/gender/geo -locations	Data collection and analysis methods	Major results
Troussas et al., 2019	- 40 students from private school using MDI version and 40 using conventional version (non MDI mechanism) - Various grades - Learning English and French as foreign languages - 5 school teachers for assistance	- 3 learning sessions for Speech Language Acquisition (SLA) using the MDI software followed by a scale-based questionnaire containing 13 questions with ratings of 1-10 - software uses an "error diagnosis using machine learning"mechanism called MDI (misconception detection and identification) - MDI uses a "Fuzzy String searching" (non language transfer errors) and "String Interpreting Resemblance" (language transfer errors) concepts to infer a student's "learning and forgetting process" (i.e., what type of misconception it is and what knowledge is lacking) - Using information to provide appropriate feedback -Two-sample t-test	- Effectiveness of the system (learning outcome improvement): 63% very effective - Efficiency in use of time: 60% very efficient - Accuracy of error diagnosis (machine): 70% very accurate - Accuracy of learning material delivery (fuzzy logic technique): 68% very accurate - User satisfaction: 65% very satisfied - Easiness of use: 55% very easy - MDI produces better learning outcomes, error diagnosis, provide adaptive delivery of learning material, and overall experience, and provides more student-centred learning experience (conventional does not) - Students need to learn the mechanism first - Both have friendly user interface for engagement - 4/5 instructors in interview noted the design's benefit in learning and education process

Studies/Authors/ Publication Year	Participants/Age, grades/gender/geo-locations	Data collection and analysis methods	Major results
Xu et al., 2019	- Two 4th grade classes - ESL students - 35 students in experimental group - 32 in the control group	- Use of "mobile-based handwriting recognition AR application Way to the Buildings" (pg 174) - Buildings and Description matching: Students match images of buildings with the correct description of the building purpose (i.e., drugstore is to buy medication and other items) using the application—instant feedback for pairing tasks and positive feedback for completing 9 pairing tasks. - Finding the Way: scanning word card for the building using AR/AI tech and finding it on the map through verbal instructions - Questionnaires for students on the useability and acceptance of tools - Interviews of teachers and students on the useability and acceptance of the tool - T-test: for performance effectiveness, for learning attitude	- Handwriting recognition AR application improves Primary school students' knowledge - Improved retention in comparison to the control group - Questionnaire of students demonstrates keenness to use technology and its potential; hence a positive reception/attitude - Text recognition function in the application (using AI methods) made instructional preparation more convenient for the teachers

Appendix B. Acronym Definitions

Acronym	Full form	Definition
AAI/KBS	Artificial Augmented Intelligence/	IBM's Watson is an example of an AAI/KBS which, in Crowe et al (2017), is based on using cognitive computing to "understand natural language processing,

	Knowledge-Based System	[and] adapts, learns, generates, and evaluates hypotheses" (p. 495). This system's use of an "augmented" AI approach comes from trying to keep the human computer interaction intact, and not replace educator roles (Crowe et al., 2017).
AES	Adaptive E-learning System	Used in Alshammari et al. (2019), this online system uses learning style and knowledge level of students to give them the appropriately leveled learning materials (basic, intermediate, or advanced).
AI	Artificial Intelligence	The development of machines that have some level of intelligence, with the ability to perform human-like functions, including cognitive, learning, decision-making, and adapting to the environment" (Chen et al., 2020, p.75267).
AIEd	Artificial Intelligence in Education	The uses of AI in education are vast and can be but isn't limited to learning analytics (Sayed et al., 2022), assessment and evaluation (Zhao et al., 2023), eye tracking and facial interpretation (Seo et al., 2021; Meikleham & Hugo, 2020), intelligent tutoring systems, feedback automation, robots, etc. (Xu & Ouyang, 2022). These facilities are intended to assist the learning and teaching process.
ANN	Artificial Neural Network	AI modelling tool that imitates the system of the human brain's neurons to create system that allows the system to complete a task, which can be improved through multiple trials and iterations (Lau et al., 2019).
AR-AI	Augmented Reality - Artificial Intelligence	A combined technology used by Xu et al. (2019) using an AI feature which allows for handwriting scanning in one section of the study, and an AR function for a matching activity.
cGPA	Cumulative Grade Point Average	cGPA is a measure of a student's academic grade based on a unique point system that takes the average of all their course grades in their program till date.
DRFLO	Dynamic Recommendation of Filtered Learning Objects	An automated text feature extraction mechanism to retrieve LOs (Tahir et al., 2022). Based on semantic similarity, LOs are retrieved using intelligent mechanisms from a repository through a query search feature within the system (Tahir et al., 2022).
EC	Evolutionary Computation	A technique used in Jiao et al. (2022) which helps to predict quantitative relations between input and output variables from complex datasets. EC is a subdivision of

		AI which deals with algorithms that are continually evolving and optimizing (Jiao et al, 2022).
EWS	Early Warning System	Early identification of at-risk students using a visual system for teachers and students in which green-amber-red signals are given to determine students' chances of passing the course (Baneres et al., 2019). There are also email messaging systems personal to each student providing intervention recommendations (feedback) for improvement (Baneres et al., 2019).
GAR	Gradual-At-Risk	A model created in Baneres et al. (2019) which helps to detect at-risk students based on their grades.
GP	Genetic Programming	An AI algorithm model used in Jiao et al. (2022) and Ouyang et al. (2023) that uses quantified input from the program's Blackboard LMS to predict student performance.
ITS	Intelligent Tutoring Systems	Computer learning systems that are more personalized and focus on detection, modelling, tracing, and fostering self-regulated learning through human-like avatars that interact with students as "pedagogical agents" (Chango et al., 2021, p.615). MetaTutorES, used in Chango et al. 's (2021) study is one example of an ITS used in literature.
LA	Learning Analytics	The process in which the information or data about individuals is collected (i.e., background, behaviours, etc.) and analyzed by some algorithm to produce an output (i.e., feedback, prediction about student scores, etc.).
LMS	Learning Management System	Software or system that is used in virtual learning environments which manages learning and how students and educators interact with one another, and the learning materials uploaded on the cloud system (David, 2013).
LO	Learning Object	Learning materials uploaded on LMSs such as assessment pieces such as quizzes, content such as videos, and interactive pieces such as games or simulations (Davey, 2023). LOs can be "saved", tagged, and re-used on LMSs through search features which retrieve them from repositories based on keyword detection (Davey, 2023).
MDI	Misconception	Used in Troussas et al. (2019), MDI is an algorithm-

	Detection and Identification	based system which helps to detect language errors and provide information on students' language acquisition.
RiPPLE	Recommendation in Personalized Peer-Learning Environment	A comprehensive, systematic, and adaptive tool that follows a 4-section process to vet the resources uploaded by students to the learning system through feedback mechanisms (Darvishi et al., 2022). The tool analyzes the student-generated resources' validity by analyzing their peers' and instructor's feedback (Darvishi et al., 2022).
SME	Subject Matter Experts	Individuals who are experts or professionals in a specific subject or topic.
SWS	Scholarly Writing Software	An application that is embedded into written tools and word processing applications to help improve the written work by providing suggestions in real time, beyond the capacities of the word processing application and its add-ons (Crowe et al., 2017).