Large Language Models and Educational Gamification: A Literature Review

KATHLEEN KELLY, Colorado School of Mines

This literature review aims to show the intersection between two popular pedagogies: large language models (LLMs) in education and gamification in education. Sixteen relevant papers were used in this review, all published in the current year. While there is much work being done with LLMs in education as well as with gamification in education, this review reveals that the intersection of the two is a small intersection with a limited amount of research therein, especially when focusing specifically on computer science education. This gap paves the way for future work in the realm of LLMs and gamification as a way to enhance student motivation in computer science.

Additional Key Words and Phrases: gamification, education, higher education, knowledge, engagement, satisfaction, games, computer science, gamification computer science, augmented reality, serious games, digital game-based learning, extended reality, feedback, debugging, teacher utterances, large language models

ACM Reference Format:

Kathleen Kelly. 2025. Large Language Models and Educational Gamification: A Literature Review. 1, 1 (June 2025), 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Instructors are always looking for pedagogical techniques to improve student engagement and motivation in the classroom. One such technique is the addition of gamification to the classroom, while another newer technique is adding large language models (LLMs) as a source of classroom material. In this literature review, the goal is to research the intersection of LLMs and gamification of educational content. Specifically, three research questions will be explored: 1) what are the primary mechanisms through which LLMs can be utilized for delivering educational content; 2) how do gamification techniques impact student engagement and educational outcomes; and 3) what is the state of the art in leveraging LLMs for educational gamification. While searching for evidence to support each of these questions in this review, some common themes as well as limitations are summarized and discussed.

Author's address: Kathleen Kelly, kathleenmariekelly@mines.edu, Colorado School of Mines, Golden, CO.

XXXX-XXXX/2025/6-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 LARGE LANGUAGE MODELS IN EDUCATION

Many professors are exploring ways to use LLMs to deliver educational content. To research this area, a literature search was done focusing on LLMs in computer science education that were published in the current year. The search was narrowed down to six papers that highlight novel research in computer science education (see Table 1). The findings included themes of LLMs providing computer science program explanations, LLMs doing automated debugging, and LLMs acting as teachers in the form of chatbot conversations. From these findings, the primary ways in which LLMs are being utilized to deliver educational content in programming classrooms are by adding additional descriptions of the programming problems and by conversing with students as a teacher.

In a recent SIGSCE paper, MacNeil et al. [8] used an **LLM to help students receive explanations of programming code**. Instead of the instructor having to do extensive up-front work to generate explanations and hints, the LLM was able to support formative assessment. LLM output fell into three types of explanations: line-by-line, summaries, and concepts, which were then added to the students' E-Book. MacNeil et al. conducted the study during a web software development course at a university in Finland (N = 116). The researchers investigated undergraduate student engagement and the impact of the three LLM explanations. The explanations were pre-generated and added to the course E-Book. Students could optionally view them and give feedback on their usefulness. Participating students reported that the LLM explanations were useful. Explanations viewed for longer periods of time were associated with longer code snippets. Most students viewed explanations for the first few code snippets, then tapered off until they reached the last chapter with more challenging code. Students viewed the line-by-line explanations most frequently, but then rated the line-by-line explanations as less useful than the summary explanations. The authors concluded that the explanations generated by the LLM were correct but at times were too wordy. Researchers also concluded that the explanations were helpful to the students, especially in a format where the student could choose whether or not to view the explanation. *Researchers suggested future work on the LLM prompts, adding personalization, and letting the students provide their own explanations as input to the model. MacNeil et al. recommended extending the study to include live explanations responsive to student questions.*

Phung et al. [13] used **LLMs to generate feedback for students working through syntax errors in Python programming assignments**. To help introductory programming students with syntax errors, the researchers studied the use of LLMs to scale the human tutor in error explanations. The goal was to use a student's program containing errors as the model's input, generating output of the fixed program and explanations. Phung et al. created novel

Table 1. LLMs in Education

Author	Title	Key Findings
MacNeil et al.	Experiences from Using Code Explanations...	LLM explanations of code samples
Phung et al.	Generating High-Precision Feedback...	LLM generation of code fixes with explanations
Kang et al.	Explainable Automated Debugging...	LLM generation of bug fix and explanation
Matelsky et al.	A large language model-assisted education tool...	LLM feedback on open-ended quiz questions
Vasselli et al.	NAISTeacher: A Prompt and Rerank Approach...	LLM acting as a teacher chatbot
Hicke et al.	Assessing the efficacy of large language models...	LLM as a "knowledgeable teacher"

software, PYFIXV, with a "tunable precision parameter" to give returned feedback control to the teachers. The feedback generation contained a run-time validation mechanism to determine relevance and helpfulness (see Figure 1). A dataset of 480 distinct Python programs were used by the LLM, along with three Python human experts for program annotation. The research was novel in its attempt to take a buggy program and generate both a fixed program and natural language explanations. *Researchers encouraged future work with more recent LLMs, including syntax and semantic error analysis, and using the software in actual classrooms for real-world studies.*

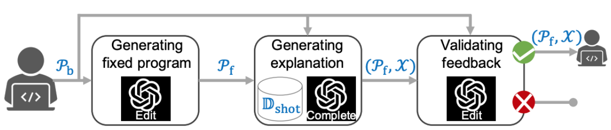

Fig. 1. Three Stages of PYFIXV Feedback Generation [13]

Kang et al. [7] used **LLMs to perform automated debugging**, prompting LLMs to interact with buggy code and reach conclusions that provided explanations to students. The researchers proposed novel software, Automated Scientific Debugging (AutoSD). The input to the software was code containing errors and the output was code with suggestions to correct the errors. The output also generated a human explanation for fixing the bug, giving the student further feedback. Internally, AutoSD utilized an LLM to generate a hypothesis, a test for the error, and an execution of the debugger. Kang et al. performed their study on participants (N = 20) to rate the validity of the code fixes and explanations. The researchers described the "hypothesis formulation, then verification" process used by students as the inspiration for the use of the LLM (see Figure 2). The researchers concluded that the explanations provided by AutoSD were helpful in the use of concrete execution steps. *The researchers suggested adding more theory to the explanations and encouraged more research to ensure a trustworthy LLM that only generates correct bug patches.*

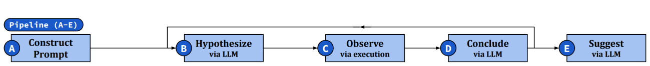

Fig. 2. Pipeline of AutoSD [7]

Matelsky et al. [10] used **LLMs to return feedback on open-ended quiz questions**. Open-ended questions are used to test student understanding but are tedious for instructors to grade effectively. Matelsky et al. created a novel open source tool, FreeText, using Python and an integration with Jupyter Notebook. Within the tool, questions, instructor criteria, and student responses were used as input to the LLM to generate rapid and personalized feedback. *The authors pointed out limitations of potential biases in the feedback and the need to ensure student awareness of the source of the feedback.* The researchers noted the importance of FreeText not being a replacement for human teachers but rather an additional tool to supplement teachers.

Vasselli et al. [15] used **LLMs to generate teacher responses in educational conversations**. Data was collected from actual conversations between teachers and students that predicted the next teacher utterance in a conversation. The conversations came from the Teacher-Student Chatroom Corpus with over 2,000 partial conversations between teachers and students learning English, containing 273 conversations. Prompts were sent to the LLM, then multiple response candidates were put through a post-processing step and ranked in order of most appropriate. The model also had to learn when to produce a new response that immediately followed a student response, also known as a continuation of a previous teacher response, therefore having to maintain context. The model was not allowed to give away an answer too quickly but was to encourage the student with hints. The goal was to increase student engagement and have the LLM respond like a human teacher. The study went through many iterations of prompts and ranking algorithms in order to get LLM responses that sounded the most realistic and accurate. *The researchers suggested future work to create a fully functional teacher chatbot to instruct and encourage students. The future work should also not rely so heavily on the quality and specificity of the LLM prompts.* The researchers noted the importance of taking biases into account such that the LLM responses do not contain any profanity or other biased comments.

Hicke et al. [5] used **LLMs to simulate the role of a "knowledgeable teacher"**, again using data from the Teacher-Student Chatroom Corpus for training. The research goals included generating conversations that would sound like a teacher that understands the student and helps improve student understanding of the course material. Hicke et al. used data collected from 102 chatroom conversations where teachers instructed students learning English as a second language. The researchers also performed supervised fine-tuning of the model along with reinforcement learning tuning. These additional steps contributed to AI conversations that were more context-aware and effective. Performance was compared using GPT-4, GPT-2, and DialoGPT to find the most pedagogically correct teacher utterances.

The researchers proposed the need for more extensive prompts to the model as well as better metrics to help in the ranking and reward training of the model.

LLMs are being utilized for delivering educational content in varying ways. In this section use cases were explored that showed LLMs being used as supplements to human teachers, giving the ability to scale teachers to better assist the much larger number of students. LLMs were also used to provide feedback and explanations to students while studying code samples or answering open-ended questions. *In each of the scenarios studied there was optimism in the use of LLMs but the recurring theme of a need for more extensive research. Researchers also shared concerns about the LLM having bias, replying with profanity, or otherwise being untrustworthy to be used without any human supervision.*

3 GAMIFICATION TECHNIQUES IN EDUCATION

Gamification is defined as adding game elements into a non-game context, such as into a computer science class. Many studies have been done to show that adding game design elements helps overall student motivation. To research this area, five papers were chosen to examine gamification techniques in the context of how they impact student engagement and educational outcomes (see Table 2). Authors in the studies looked at self-determination theory, student satisfaction and sentiment, and a student-centered approach. The findings showed that students have an increased sense of motivation and engagement when the learning outcomes include gaming elements and active participation in the creation of the game. From these findings, gamification techniques are impacting student engagement and educational outcomes with their use of personalization and student creativity.

Alsadoon [1] studied the **impact of gamification on student motivation and engagement** using various game design components in an online environment. The experiment involved first-year undergraduate male students (N = 97) at Saudi Electronic University in the fall of 2021. The group was divided such that one half used a gamified version of a learning management system, the other half used the conventional version with no game elements. The gamified version awarded points and badges as students completed activities, displaying scores on a leaderboard. Alsadoon utilized the GAFCC model in the game, the five elements of goal, access, feedback, challenges, and collaboration. Motivation and engagement instruments were used at the beginning and end of the semester to do the measurements and look for a difference between them. Alsadoon found that students were more motivated and engaged in the experimental gamified group where they were more encouraged to participate. The findings were linked to the self-determination theory, the theory of students feeling a sense of ownership as they master activities and earn points and badges. The student develops more confidence and independence through the key components of relatedness, competence, and autonomy. The theory states that these three key components are also human psychological needs to be able to collaborate with others, compete with others, and make individual choices. *The author concluded by recommending that instructors take advantage of existing platforms and applications that add gamification to computer science courses.*

Murillo-Zamorano et al. [12] created a novel **8-Pointed Higher Education Gamification Star framework** to be utilized in higher education games (see Figure 3). This framework highlighted the eight game elements that these researchers felt should be included in a successful gamification experience. These elements were points, badges, levels, leaderboards, challenges, storytelling, empowerment, and social influence. Using this framework, the researchers also developed the Scale-HEGx measurement system to measure the presence of gamification elements in a game and used it with the creation of a game, The ECOn Star Battles. This game included both in-class and out-of-class activities over 15 weeks, performed in teams of three to four students. Teams worked through five levels increasing in difficulty, competing against other teams while earning points and badges. The teams were ranked on a leaderboard and the winning team was declared the champion. The levels included challenges that were story-based. Teams could create their team names, attire, and be involved in the creation of the quizzes. The students were given a questionnaire divided into 11 blocks where students rated their levels of the 8-pointed elements, plus engagement, knowledge, and satisfaction. Results were analyzed such that relationships could be found between the game and students' knowledge, engagement, and satisfaction. Murillo-Zamorano et al. found that gamification directly influenced knowledge and engagement, but indirectly influenced satisfaction through that knowledge and engagement. *The researchers admitted that the small sample size (N = 90) was a limitation, as well as a limited group of nationalities, cultures, gender, and personality. The researchers also acknowledged the combination of gamification and artificial intelligence as a way to further enhance students' experiences.*

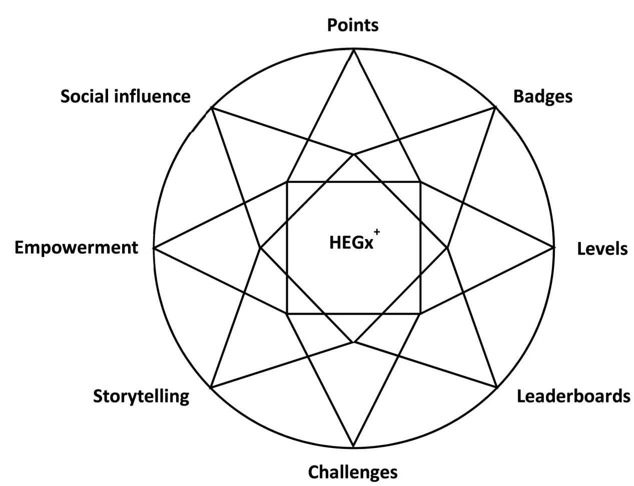

Fig. 3. 8-Pointed Higher Education Gamification Star [12]

Sadiku et al. [11] proposed that **gamification with competition** was the most effective form of motivation because of its impact

Table 2. Gamification in Education

Author	Title	Key Findings
Alsadoon et al.	The Impact of Gamification...	Gamification and self-determination theory
Murillo-Zamorano et al.	Gamification in higher education...	Student knowledge, engagement, and satisfaction
Sadiku et al.	Gamification in Computer Science	Gamification and three components of student engagement
Lampropoulos et al.	Integrating Augmented Reality...	Student-centered approach
Weitl-Harms et al.	Assessing User Experiences with ZORQ...	Student satisfaction and sentiment

on thinking skills and learning. The authors defined engagement as having three components: behavioral, affective, and cognitive. Behavioral engagement was defined as relating to school activities and positive conduct without any disruptive behavior. Affective engagement was defined as the "willingness to do the work". Cognitive engagement was defined as the student's willingness to achieve a "deep understanding and expertise". The authors claimed that creating a game to help students achieve these three types of engagement would impact student motivation. Benefits of gamification in computer science were enumerated as helping the students with problem-solving skills, fostering a collaborative environment, and improving student self-efficacy. The authors listed challenges such as the fine line between a game that engages and a game that bribes. *The authors proposed the question of whether or not a game is teaching retainable knowledge or creating a system of rewards and a "means of control".*

Lampropoulos et al. [3] looked at the use of gamification and **augmented reality in the computer science classroom**. This type of environment combines physical and virtual objects for student interaction. The authors researched the idea of combining augmented reality with games that focus on educational aspects. The research was performed on mobile devices using the Unity game engine and the Vuforia Engine SDK for augmented reality enhancements. The students involved in the study (N = 117) were from the International Hellenic University, 87% male, and were also involved in the initial design of the games. This "student-centered approach" was novel, showing that students should be involved in the development of such educational games. Following engagement with the games, students filled out paper-based questionnaires. Within the results, Lampropoulos et al. examined the emotions felt by the students, including joy, surprise, anticipation, and trust, all of which were reported to lead to better learning outcomes and motivation (see Figure 4). Students also reported that the games created collaborative learning and active engagement. *The authors admitted to the limitation that the participants were already familiar with mobile applications which could have biased the usability and learnability results. This suggested doing further research on a broader group of participants.*

Weitl-Harms et al. [16] studied the use of **dynamic gamification in a computer science setting**, where students participated in the application as well as its initial design, implementation, and ongoing customizations. The researchers studied the ZORQ gamification framework where students program and design autonomous ships in a 2D world while also learning underlying data science concepts in the code. The analyzed results were related to the students' satisfaction and sentiments in the experiment that ran over five years. In

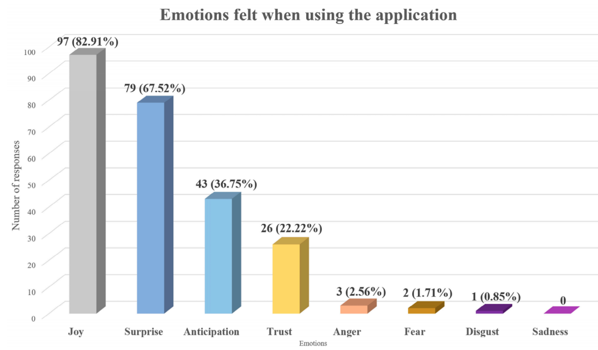

Fig. 4. Emotions Felt [3]

one survey, students selected from a list of 55 words to describe the experience, resulting in top selections of fun, stimulating, valuable, exciting, motivating, and customizable (see Figure 5). The work was novel in its study of gamification in computer science while looking at student engagement within the actual creation of the game, not merely the playing of the game. *Weitl-Harms et al. noted future work as adding ZORQ to more courses and introducing ZORQ earlier in the semester, as well as increasing sample size and adding a control group.*

Fig. 5. ZORQ Word Cloud Analysis [16]

Gamification in computer science education has been shown to increase student motivation and engagement. In this section gamification was explored in the context of measuring non-tangible student motivation results along with novel measurements of confidence, competence, satisfaction, self-efficacy, sentiment, and retention. The studies added various game elements or applications to computer science classes, then assessed students' feelings through post surveys. *Each study had its limitations related to participant size or length of the study, but overall had encouraging conclusions. Results showed that gamification increased student motivation and created positive emotions, especially when the student was both a participant and had ownership in the game development.*

4 LARGE LANGUAGE MODELS AND EDUCATIONAL GAMIFICATION

The intersection of LLMs and educational gamification is a new phenomena. LLMs are being used in educational games in different countries, different grade levels, and for different reasons. In this section, five papers from 2023 were chosen that highlight this state of the art research area (see Table 3). Emerging themes include measuring the emotions of the game players to be able to better personalize the game, allowing the student to be a part of the game creation, and adding a chat-bot functionality for students to interact with the game. These findings show that the state of the art in leveraging LLMs for educational gamification is game personalization and live interaction with AI participants.

Cao [2] performed a study on using gamification and LLMs to help Chinese programming students have an increased sense of belonging and understanding in the classroom. The paper described "AI-enhanced gamification", **AI-driven tutoring systems that create a personalized learning game** for the students. The author sought to research international students' sense of belonging within an environment of AI and gamification. The study began with a questionnaire to assess the challenges faced by a group of Chinese students (N = 57). The study then deployed a prototype of a story-based game using the learning tool of Blackboard. Based on positive feedback from the prototype, the author created a story-based intelligent tutoring system (ITS) using GPT-3. *The author admitted to needing further work with more participants and more detailed studies. The author's work showed promising results with the LLM's ability to answer questions and increase students' understanding and sense of belonging.*

Martinez et al. [9] developed a prototype of a Study-Buddy platform to help students engage with learning material in a **gamified and LLM-driven environment**. Questions were created using the LLM, then presented to the students in an environment that provided feedback, points, scores, and a leaderboard (see Figure 6). Study-Buddy would also proactively notify students when new content was added or if student's ranking was declining. The authors were able to provide a gamified learning experience and a way for teachers to track student progress at two high schools in Bolivia. *The authors noted future plans of expanding Study-Buddy to other groups of participants as well as adding a way to measure student emotions to help in the personalization of the game.*

Gonzalez-Gonzalez et al. [4] proposed a **personalized educational game** for children that used a "Chatbot Reactive Architecture". In this architecture, the authors addressed four main characteristics of the chatbot: 1) that it be reactive to changing user inputs; 2) that it be customizable depending on the learning objectives; 3) that it be personalizable; and 4) that it adapt dynamically to the

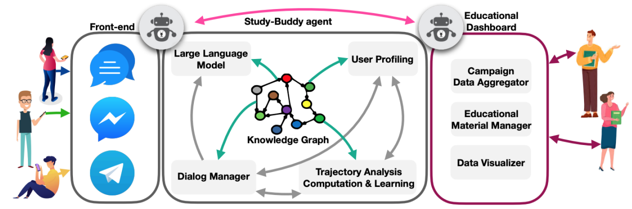

Fig. 6. Study-Buddy Architecture [9]

student's abilities. The researchers focused on the "student-player" model, where the user of the game is both a student and a game player. Keeping this view in the forefront, the authors were able to present an engine that provided educational material and real-time personalization. The chatbot architecture was novel in its use of an "orchestrator" to help maintain coherent conversations with a user depending on the context (see Figure 7). *The authors noted future work to include making the prototype more generic to work with other disciplines, and adding ways to assess the prototype for its effectiveness.*

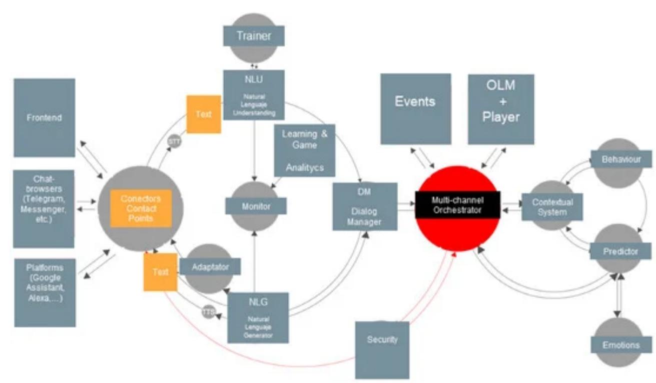

Fig. 7. Chatbot Architecture [4]

Janson et al. [6] reviewed three papers related to gamification and AI. The third paper focused on the use of a **"virtual laboratory" and gamification**, where students worked in the "metaverse" of a lab and were analyzed using a leaderboard. In the study, students were split into two groups: gamified and non-gamified, then asked to work on web services. The students with the leaderboard were found to implement more web services than those without the leaderboard, and specifically more complex services. The researchers found the leaderboard to be a major source of motivation. Taking ideas from the papers read, the authors proposed theories on how to leverage game design alongside LLMs to create positive outcomes and "foster social bonds" between the students and the AI. The authors suggested the importance of AI in the creation of the game, adding dynamic personalization and helping to overcome the novelty effect, the experience of the novelty of a game wearing off over time. *The authors proposed future work in the area of customizing games for education using LLMs.*

Table 3. LLMs and Gamification

Author	Title	Key Findings
Chen Cao	Leveraging Large Language Model and Story-Based Gamification...	AI-enhanced gamification
Martinez et al.	Study-Buddy...	Study-Buddy gamified experience
González-González et al.	Personalized Gamification...	Student-Player model
Janson et al.	... Adaptive and Intelligent Gamification Design	Virtual laboratory
Sudhakaran et al.	MarioGPT...	Dynamic video game levels

Sudhakaran et al. [14] used a GPT2 model to develop **dynamic game generation**. The model, MarioGPT, was described as a tile-based game level generator for the well known Mario game. The model was able to generate dynamic levels in the same vein as Procedural Content Generation (PCG), while allowing for customized "play-style dynamics" with the use of guided prompts. The authors proposed a text-to-level model (see Figure 8) that generated levels for the Mario game using LLMs and natural language prompts. The authors also proposed a search capability to produce an endless stream of diverse levels. The search capability involved taking a random slice from an archived level, mutating it using MarioGPT sampling, then adding it back into the archive using a "novelty score" (see Figure 9). Sudhakaran et al. developed novel work that allowed for the prediction of player interaction in the levels. The researchers also generated new and diverse playable levels (88% of the time). *The authors noted future work that included bringing human feedback into the loop to truly fine-tune the models.*

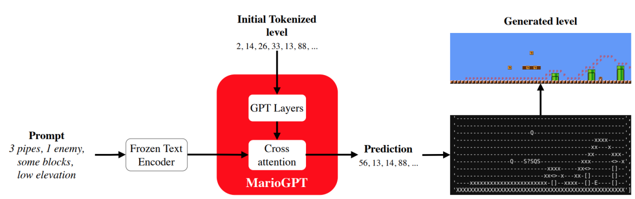

Fig. 8. MarioGPT Text-To-Level [14]

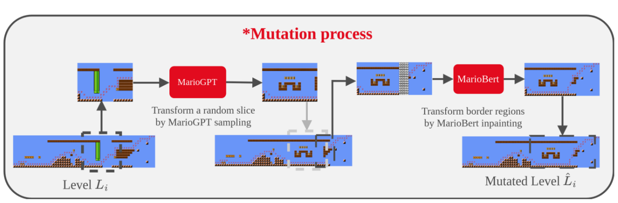

Fig. 9. MarioGPT Level Mutation [14]

LLMs are finding their way into many aspects of computer science including educational computer games. In this section, this combination of LLMs and educational gamification was reviewed, producing varying use cases. *Common themes found were the use of LLMs to increase the personalization of the game, the use of LLMs to act as a dynamic chat system for the game participants, and allowing for student involvement in the game creation.*

5 CONCLUSION

In a time when large language models are becoming a part of daily life, teachers are working to incorporate them into the classroom in an ethical manner. Similarly, research has been done on the effects of adding gamification to the classroom to better motivate students and teach concepts in a visually engaging manner. In this literature survey, 16 papers were studied related to the intersection of these two pedagogies, searching for the state of the art of leveraging LLMs for educational gamification and what limitations currently exist.

First, LLMs in education were examined, finding that LLMs are being used in computer science classes to help students with coding questions, with debugging, and with receiving feedback in a format that mimics that of a teacher. The biggest limitation was in the LLM itself. The tool requires very specific prompts in order to answer in the most accurate and unbiased manner. There is still much manual intervention involved to be able to trust the LLM to act in place of a teacher.

Second, gamification in education was examined, finding that gamification has been a technique used in computer science classrooms for years. Measuring how well a game increases student motivation and engagement has been challenging, as well as determining which game elements have the biggest impact. One common theme was that of adding competition and a leaderboard to the game, but also being aware of the fine line between a game teaching concepts and a game being simply a system of rewards. The way people enjoy games is very personal, both in a basic like or dislike of games in general, but also of how the look and feel of the game affects each individual. With that, another common theme was that of adding personalization to a game so it fits the needs of any student.

Lastly, LLMs and educational gamification were examined, finding that there is an intersection between LLMs and educational games, although that intersection is small to date. The LLM has been used to generate the questions within the game, to generate the actual game levels, and to act as an interactive game chatbot. In each study the common theme was that the work was very new and in need of more research as well as more participants.

The intersection of LLMs and educational gamification has the potential to significantly enhance student engagement and learning outcomes. However, research has shown that LLMs are still very new and cannot be completely trusted, while gamification is still a difficult area to measure with regard to its benefits. Though there are limitations, there are also exciting gaps that can lead to using LLMs to create and personalize educational games that could revolutionize the industry. In conclusion, putting together the research of this study shows huge potential for LLMs, teachers, and students to work together to create custom educational games that suit any student's needs, pushing motivation and engagement to its highest levels.

6 ACKNOWLEDGMENTS

The author thanks Dr. Bo Wu for advice on the direction of this research, tieing together the novel use of large language models with her interest in gamification of computer science materials. She also thanks Dr. Christine Liebe for advice on how to research a topic of interest, how to read a paper, and how to write a literature review.

REFERENCES

[1] Hamadah Alsadoon. 2023. The Impact of Gamification on Student Motivation and Engagement: An Empirical Study. *Dirasat: Educational Sciences* 50, 2 (2023), 386–396. https://doi.org/10.35516/edu.v50i2.255

[2] Chen Cao. 2023. Leveraging Large Language Model and Story-Based Gamification in Intelligent Tutoring System to Scaffold Introductory Programming Courses: A Design-Based Research Study. arXiv:2302.12834 [cs.HC]

[3] K. Diamantaras G. Evangelidis G. Lampropoulos, E. Keramopoulos. 2023. Integrating Augmented Reality, Gamification, and Serious Games in Computer Science Education. *Education sciences* 13 (2023). Issue 6. https://doi.org/doi: 10.3390/educsci13060618

[4] Carina S. González-González, Vanesa Muñoz-Cruz, Pedro Antonio Toledo-Delgado, and Eduardo Nacimiento-García. 2023. Personalized Gamification for Learning: A Reactive Chatbot Architecture Proposal. *Sensors* 23, 1 (2023). https://doi.org/10.3390/s23010545

[5] Yann Hicke, Abhishek Masand, Wentao Guo, and Tushaar Gangavarapu. 2023. Assessing the efficacy of large language models in generating accurate teacher responses. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*. Association for Computational Linguistics, Toronto, Canada, 745–755. https://doi.org/10.18653/v1/2023.bea-1.60

[6] Andreas Janson, Manuel Schmidt-Kraepelin, Sofia Schobel, and Ali Sunyaev. 2023. Special Issue Editorial: Adaptive and Intelligent Gamification Design. *AIS Transactions on Human-Computer Interaction* 15, 2 (2023), 136–145.

[7] Sungmin Kang, Bei Chen, Shin Yoo, and Jian-Guang Lou. 2023. Explainable Automated Debugging via Large Language Model-driven Scientific Debugging. arXiv:2304.02195 [cs.SE]

[8] Stephen MacNeil, Andrew Tran, Arto Hellas, Joanne Kim, Sami Sarsa, Paul Denny, Seth Bernstein, and Juho Leinonen. 2022. Experiences from Using Code Explanations Generated by Large Language Models in a Web Software Development E-Book. arXiv:2211.02265 [cs.SE]

[9] Fernanda Martinez, Diego Collarana, Davide Calvaresi, Martin Arispe, Carla Florida, and Jean-Paul Calbimonte. [n. d.]. Study-Buddy: A Knowledge Graph-Powered Learning Companion for School Students. ([n. d.]).

[10] Jordan Matelsky, Felipe Parodi, Tony Liu, Richard Lange, and Konrad Kording. 2023. A large language model-assisted education tool to provide feedback on open-ended responses.

[11] Janet O. Sadiku Matthew N. O. Sadiku, Uwakwe C. Chukwu. 2023. Gamification in Computer Science. *Central Asian Journal of Mathematical Theory and Computer Sciences* 04 (2023). Issue 5.

[12] Luis R. Murillo-Zamorano, José Ángel López-Sánchez, María José López-Rey, and Carmen Bueno-Muñoz. 2023. Gamification in higher education: The ECOn+ star battles. *Computers Education* 194 (2023), 104699. https://doi.org/10.1016/j.compedu.2022.104699

[13] Tung Phung, José Cambronero, Sumit Gulwani, Tobias Kohn, Rupak Majumdar, Adish Singla, and Gustavo Soares. 2023. Generating High-Precision Feedback for Programming Syntax Errors using Large Language Models. arXiv:2302.04662 [cs.PL]

[14] Shyam Sudhakaran, Miguel González-Duque, Claire Glanois, Matthias Freiberger, Elias Najarro, and Sebastian Risi. 2023. MarioGPT: Open-Ended Text2Level Generation through Large Language Models. arXiv:2302.05981 [cs.AI]

[15] Justin Vasselli, Christopher Vasselli, Adam Nohejl, and Taro Watanabe. 2023. NAISTeacher: A Prompt and Rerank Approach to Generating Teacher Utterances in Educational Dialogues. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*. Association for Computational Linguistics, Toronto, Canada, 772–784. https://doi.org/10.18653/v1/2023.bea-1.63

[16] Spanier Adam Rokusek Matthew Hastings John Weitl-Harms, Sherri. 2023. Assessing User Experiences with ZORQ: A Gamification Framework for Computer Science Education. (2023).