



# Learn to Machine Learn *via* Games in the Classroom

Marvin Zammit\*, Iro Voulgari, Antonios Liapis and Georgios N. Yannakakis

*Institute of Digital Games, University of Malta, Msida, Malta*

Artificial Intelligence (AI) and Machine Learning (ML) algorithms are increasingly being adopted to create and filter online digital content viewed by audiences from diverse demographics. From an early age, children grow into habitual use of online services but are usually unaware of how such algorithms operate, or even of their presence. Design decisions and biases inherent in the ML algorithms or in the datasets they are trained on shape the everyday digital lives of present and future generations. It is therefore important to disseminate a general understanding of AI and ML, and the ethical concerns associated with their use. As a response, the digital game *ArtBot* was designed and developed to teach fundamental principles about AI and ML, and to promote critical thinking about their functionality and shortcomings in everyday digital life. The game is intended as a learning tool in primary and secondary school classrooms. To assess the effectiveness of the *ArtBot* game as a learning experience we collected data from over 2,000 players across different platforms focusing on the degree of usage, interface efficiency, learners' performance and user experience. The quantitative usage data collected within the game was complemented by over 160 survey responses from teachers and students during early pilots of *ArtBot*. The evaluation analysis performed in this paper gauges the usability and usefulness of the game, and identifies areas of the game design which need improvement.

## OPEN ACCESS

### Edited by:

Iza Marfisi-Schottman,  
EA4023 Laboratoire d'Informatique de  
l'Université du Mans (LIUM), France

### Reviewed by:

Samia Bachir,  
Le Mans Université, France  
Michael Kickmeier-Rust,  
University of Teacher Education St.  
Gallen, Switzerland

### \*Correspondence:

Marvin Zammit  
marvin.zammit@um.edu.mt

### Specialty section:

This article was submitted to  
Digital Learning Innovations,  
a section of the journal  
Frontiers in Education

Received: 05 April 2022

Accepted: 20 May 2022

Published: 17 June 2022

### Citation:

Zammit M, Voulgari I, Liapis A and  
Yannakakis GN (2022) Learn to  
Machine Learn *via* Games in the  
Classroom. *Front. Educ.* 7:913530.  
doi: 10.3389/feduc.2022.913530

**Keywords:** machine learning, artificial intelligence, serious games, educational games, game analytics, digital literacy, supervised learning, reinforcement learning

## 1. INTRODUCTION

Through the use of the world-wide web and access to applications such as social media, games, Google search, YouTube, and the Internet of Things, digital technologies are mediating children's learning, entertainment and social interactions, and are shaping their everyday lives (Rahwan et al., 2019). The online environment has become ubiquitous, and artificial intelligence (AI) and machine learning (ML) algorithms are employed across most digital platforms to track user preferences, suggest content, and even generate novel material for a wide variety of purposes (Yannakakis and Togelius, 2018). This pervasiveness can enhance the user experience but is fraught with controversy, as user data is often stored and possibly shared with third parties without clarity to the user. This may lead to breach of personal information, or more nefarious acts such as steering of public opinion or distortion of facts. Despite the implications of such digital technologies, children's conceptions of technology, its potential and implications may be vague, inaccurate or distorted (Druga et al., 2017; Mertala, 2019). It is therefore crucial—especially for the younger generations—to understand how AI is used, how it works, and what its pitfalls are. Critical thinking skills, such as making inferences, decision-making, and problem solving (Lai, 2011) have to be promoted. This is necessary for children to understand the implications and context of the online content they access, to recognize the use of AI and how it encroaches into their daily life and their social environment.

The work presented in this paper is situated in this context; we present ArtBot, a game that aims to provide young learners with the background knowledge of basic concepts behind AI and ML, and to show how basic algorithms can be used to solve different problems. In addition, we aim to highlight the challenges emerging from AI, thereby exposing students to the capabilities and limitations of these algorithms. Extensive research over the past few decades has shown the effectiveness and potential of digital games as learning tools; games may support motivation, engagement, and active participation of learners, enhance attention, involvement, and understanding of abstract and complex concepts (Hainey et al., 2016; Tsai and Tsai, 2020). Our game, ArtBot, builds upon this potential of games and aims to introduce young learners to the two fundamental processes of ML: supervised learning (SL) and reinforcement learning (RL). ArtBot was developed to be a teaching tool in classrooms to raise awareness and critical thinking about AI processes. The game was made available primarily through web browsers<sup>1</sup>, but also for download on Windows operating systems, and as a mobile app on Android devices through the Google Play store. ArtBot comprises of two mini-games, one containing a level dedicated to SL, and the other consisting of ten levels dedicated to RL. The game narrative has the players tasked with retrieving art objects which have gone missing. The player is assisted by an AI helper (ArtBot) which they must train first to distinguish between statues and paintings, then to navigate rooms to collect statues scattered within them, avoid hazards and reach the exit. The game offers players customization of the avatar depicting ArtBot from a set of predefined models and color schemes.

ArtBot has been deployed on all platforms concurrently in April 2021, and included the collection of anonymous usage metrics. In our past publications (Voulgari et al., 2021; Zammit et al., 2021), we have focused on the educational design of ArtBot, how the requirements which emerged through focus groups with stakeholders were translated and adapted into a practical game design, and its implementation, deployment and initial reception. This paper, instead, explores the longitudinal usage of the ArtBot game in real-world settings over a period of almost 1 year.

Since its deployment to the general public, the game has been played by over 2,000 unique users across all platforms. This has supplied us with a substantial body of interaction data to analyse the game objectively, and draw conclusions about its design and interface based on user behavior. In this paper we review this data and try to obtain practical insights that expose the strengths and weaknesses of the game, and to evaluate whether the players' interaction with the game followed patterns intended by the design process.

## 2. GAMES FOR AI LITERACY

Using games for teaching AI is not an entirely new concept. Initiatives such as those of Clarke and Noriega (2003) and Hartness (2004), which involved a war simulation game and

<sup>1</sup>Initially launched at <http://learnml.eu/games.php> and later moved to a dedicated website at <http://art-bot.net/> on 22 March 2022.

*Robocode*, respectively, introduced games for teaching AI algorithms to undergraduate computer science students. What has shifted, though, over the past few years is the framing of AI education through games and the age of the target group. Building upon the potential of games to support systems thinking, computational thinking, and understanding of complex concepts and processes (Clark et al., 2009; Voulgari, 2020), platforms, games, and applications to support AI literacy have been developed for learners as young as 4 years old, addressing the technical, societal, and ethical aspects of AI (Giannakos et al., 2020; Zammit et al., 2021).

Games, such as the commercial game *While True: Learn()*<sup>2</sup> and *ViPER* (Parker and Becker, 2014) are appropriate for younger students and aim to scaffold the players through understanding ML concepts such as optimization, loops, and model accuracy. In *While True: Learn()* players assume the role of a computer programmer who tries to develop a model for communicating with their cat and complete tasks using visual programming, while in *ViPER* players train a robot, through coding, to navigate its way on one of Jupiter's moons. Another commercial game, *Human Resource Machine*<sup>3</sup>, which introduces the players to concepts such as automation and optimization by programming the employees in an office environment, has been adapted for students and teachers in educational contexts such as participation in the *Hour of Code*.<sup>4</sup>

Available platforms and tools allow learners to build their own ML models (engaging them in exploratory and constructivist learning practices) and to situate the models in the context of their own interests or real-world problems. *Google Teachable Machine* (GTM), for instance, has been used by 12–13 years old students, introducing them to ML concepts through the design of their own ML models, relevant to their interests or real-world problems (Toivonen et al., 2020; Vartiainen et al., 2021). Results have shown that GTM is appropriate for students with little or no programming experience. Situated in an appropriate learning context, activities through GTM allowed the students to exhibit design thinking, reason inductively about the quality of the datasets and the accuracy of their models, and show empathy to other people's needs in order to develop appropriate ML applications. Zimmermann-Niefield et al. (2019), using a tool they developed (*AlpacaML*), also engaged high school students in the design of their own ML models based on their own interest (i.e., athletic activities). The students experimented with the design of their models and reflected upon the characteristics of a good model and how the models work.

Approaches such as those of Turchi et al. (2019) and Microsoft's *Minecraft. Hour of Code: AI for Good*<sup>5</sup> also situate AI and ML into real-world problems. Turchi et al. (2019) used a combination of online and board game play to introduce AI concepts to students and professionals involving the protection of wildlife, while *Minecraft. Hour of Code: AI for Good* scaffolds

<sup>2</sup><https://luden.io/wtl/>

<sup>3</sup><https://tomorrowcorporation.com/humanresourcemachine>

<sup>4</sup><https://tomorrowcorporation.com/human-resource-machine-hour-of-code-edition>

<sup>5</sup><https://education.minecraft.net/hour-of-code>

players to programme a robot to predict forest fires. Through such approaches learners may not only be introduced to concepts of AI and ML but also understand the role, potential, and impact of AI and ML applications in authentic and meaningful contexts.

Existing platforms, games, and applications seem to either provide an open-ended environment for learners to experiment, design and develop their own models applying concepts of AI and ML, or scaffold learners through a linear sequence of puzzles, to become familiar with AI and ML functions, processes, and algorithms (Voulgari et al., 2021). While the latter approach may facilitate novice students to understand basic principles of AI and ML algorithms, the open-ended approach allows students to engage in problem-solving tasks, reflect on their actions, assess and re-examine their progress and construct their knowledge by assuming a more active role, in line with constructivist and constructionist approaches (Kafai and Burke, 2015). In our game design, we tried to combine elements from both approaches; by guiding the players through linear tasks we aimed to introduce learners to core concepts of AI and ML, and through the more open-ended tasks, we provided space for experimentation, problem-solving, and reflection.

### 3. LEARNML AND ARTBOT

LearnML (or Learn to Machine Learn)<sup>6</sup> is a European-funded project aiming to develop digital literacy and awareness of AI usage in the digital landscape to learners who are exposed to these technologies from an early age. Its goal is to develop a toolkit for teachers and students which can be used primarily in a classroom environment, but also in non-formal learning settings (e.g., at home). The project involved the development of a number of educational games which teach different aspects of ML. These are supported by teaching materials that supplement the experience through classroom discussions to encourage reflection and critical thinking. A number of workshops and events for teachers and students were also organized in order to disseminate the work done and also gauge feedback directly from stakeholders (Voulgari et al., 2021). The impact on students, teachers, and their needs was always a priority during the development of ArtBot, which was part of the LearnML project and shared its broader goals. ArtBot was designed in collaboration with educators, with requirements collected through participatory design workshops held in three countries (Greece, Malta, and Norway) and included participation of all stakeholder categories: e.g., teachers, students, and AI researchers (Zammit et al., 2021).

The goals of ArtBot were distilled from the stakeholder needs and consisted of the following (Voulgari et al., 2021):

- To introduce the process of supervised learning, including terminology and concepts of training and testing datasets, classification, labeling, image recognition, decision trees, and prediction accuracy, and outline their role and behavior in an AI system.

- To introduce reinforcement learning and related concepts, such as rewards and penalties, learning rate, exploration, exploitation, and pathfinding.
- To show that the design decisions behind the implementation of an AI system have a considerable bearing on its behavior and outputs, thereby highlighting the fact that human bias may seep through the workings of the algorithms.
- To provoke reflection and discussions on the impact of AI systems in everyday life situations e.g., facial recognition, self-driving vehicles, etc.

With these objectives in mind, two mini-games were developed within ArtBot, focusing on supervised learning (SL) and reinforcement learning (RL), respectively. Specifically, regarding SL, concepts such as training set, testing set, data labeling, classification, and decision trees were introduced. For RL, we focused on introducing concepts such as rewards, learning rate, and exploration. The premise of the game was that a number of statues have gone missing, and the players are tasked with their retrieval. To assist in their quest, the players are given an autonomous helper (ArtBot). We tried to set the story in a meaningful narrative background since narrative seems to motivate the students, support understanding of abstract concepts and the construction of mental models, and re-frame the activities and challenges of the game into an authentic context (Glaser et al., 2009). ArtBot must first be trained to distinguish between statues and paintings, before it can be sent to retrieve the missing statues. Therefore, the SL mini-game is played first, and once completed the RL mini-game becomes available; in general, players need to complete the previous mini-game or level in order to proceed to the next one. Each mini-game is described below, while more details on the design, interfaces, and algorithms included in ArtBot are provided by Zammit et al. (2021).

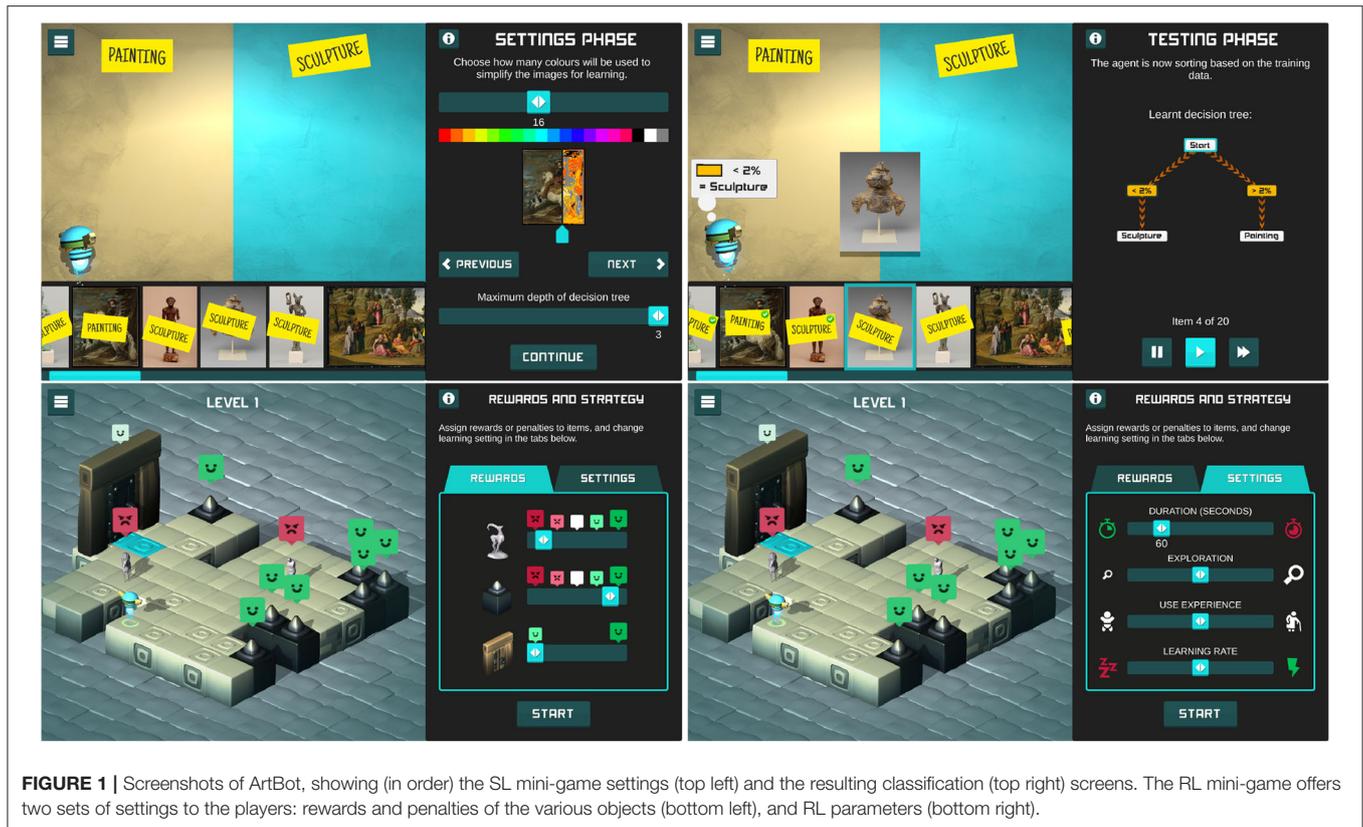
#### 3.1. Description of the Supervised Learning Mini-Game

Supervised learning (SL) is an umbrella term for ML algorithms which are used when a considerable amount of data pertaining to the problem is available (Zhou and Liu, 2021). SL is commonly applied to classification problems. An existing set of labeled data is processed through the algorithm, which tries to find some complex function that accommodates the majority of points. The available data is usually split into a training set, which is used to teach the model, and a testing set, which is used to verify the accuracy of the trained model on unseen data. Some inherent problems with SL are that data labeling is a laborious process, and that the data itself may be biased, or even incorrect, leading to this bias being learned by the model itself.

The SL mini-game tasks players to label a number of images of paintings and statues; after the labeling process is complete, ArtBot uses a decision tree based on the supplied labels to classify hitherto unseen images. All images were obtained from the Open Access Artworks<sup>7</sup> collection of the Metropolitan Museum of Art in New York, USA, and are photographs of real paintings and statues.

<sup>6</sup><http://learnml.eu/>

<sup>7</sup><https://www.metmuseum.org>



**FIGURE 1** | Screenshots of ArtBot, showing (in order) the SL mini-game settings (top left) and the resulting classification (top right) screens. The RL mini-game offers two sets of settings to the players: rewards and penalties of the various objects (bottom left), and RL parameters (bottom right).

Players are allowed to assign incorrect labels, experiment, and see how this affects ArtBot’s learned classification skills. The classification was simplified to a left swipe (to label an image as painting) or right swipe (to label it as statue) in order to speed up the process. However, this is an inherently repetitive task and can rapidly get boring. To mitigate this, we limited the classification task to 20 images, and also provided an auto-sort button which assigns the ground truth label to each image automatically.

In order to decrease the computational resources required to process and group the dataset images, the SL algorithm splits the pixel spectrum into a smaller number of colors which is controlled by the players. A decision tree is then trained on the training set that has been labeled by the player using the C4.5 algorithm (Quinlan, 1992), and used to classify the image as a painting or statue depending on the pixel count in each of the resulting color bins. The player is also given control on the maximum allowed depth of the decision tree. The interface for this mini-game is shown in **Figure 1**. Note that when showing the player the results of the SL mini-game, the training accuracy is shown as well as a testing accuracy on a set of 20 images that are unseen by both the players and the game’s designers. This training and testing accuracy introduces learners to the concept of training and test sets, and illustrates how supervised learning can be used to predict patterns in unseen data but can suffer from overfitting to patterns in the training data.

### 3.2. Description of the Reinforcement Learning Mini-Game

Reinforcement Learning (RL) is applied when the problem at hand requires a policy or a behavior which will generate the expected solution (Sutton and Barto, 2018). In such cases, there is generally either no predefined dataset available, or the problem is incongruous with a structured one. The learning process starts with a random policy (i.e., taking actions at random) while rewards or penalties are awarded for each action. The algorithm updates its knowledge about the problem according to these rewards, over a repeated number of trials—or episodes. The process tries to maximize the rewards obtained by changing its policy and observe the resulting performance. Some of the issues with this approach are the lengthy “trial and error” approach, as well as the challenge of balancing out how much weight the algorithm should put toward its acquired knowledge vs. the exploration of yet unknown actions.

In the RL mini-game, the player oversees how the ArtBot can learn to navigate through 10 different levels, avoiding hazards (spikes), collecting statues, and finding the exit. Players can set rewards and penalties to be used in the learning algorithm for each type of game object. In addition they are given a number of controls over RL parameters, such as the learning rate, the balance between exploration of new areas and utilization of already discovered information, and the total time allowed for training. When players choose rewards and parameters, the game will use a basic Q-learning algorithm (Watkins and Dayan, 1992) to negotiate a path across the level. Initially ArtBot will start

taking random steps in the environment and record the resulting rewards. As more rewards are encountered across the different episodes, the agent learns an optimal path that maximizes the total reward.

The players can progress to the next level if the exit is found and at least one statue is collected. They can however change the settings and retry each level at will. The user interface for this mini-game and the corresponding settings available to the players are shown in **Figure 1**.

## 4. DEPLOYMENT AND DATA COLLECTION

ArtBot was developed in the Unity game engine<sup>8</sup> due to the engine's capabilities to deploy to multiple platforms. We made use of Unity's integrated analytics service to collect in-game events anonymously across all platforms in order to better understand the performance of the game in terms of adoption, use, and player experience. ArtBot was deployed on the 8th April 2021 on the LearnML website<sup>9</sup>; however, it was shown to some focus groups even earlier as discussed in Section 4.2. The game can be played directly inside web browsers supporting WebGL technology, downloaded for Microsoft Windows operating systems, or downloaded to Android devices from the Google Play Store.<sup>10</sup> The game has been localized in English, Greek, and Norwegian languages, to facilitate a more widespread adoption in the countries of the LearnML project partners.

### 4.1. Collection of Game Analytics

The data presented in this document covers a period from the launch date (8 April 2021) up to 28 February 2022, although data collection is still ongoing.

**Figure 2** shows the total number of distinct users as well as the monthly active users. Player uniqueness can be ascertained for the Windows and Android platforms, but web browser sessions are less identifiable (e.g., anonymous browsing modes in the web browser client will make different sessions by the same user appear as different users). However, we assume that this measurement error is not significant enough, especially since there is no apparent benefit for the users to obfuscate their activities in the game. **Table 1** shows the distribution of users across the platforms over which the game was deployed. It is interesting to note that the majority of the players chose to play the game through the browser, which indicates that facilitating immediate launch of the game (as opposed to the download and installation required on the other platforms) is important to the end-user. This is particularly relevant as the game is (also) intended for classroom use, and special privileges are usually required on public computers for installing new software; instead, the browser experience is available to all. It is also worth mentioning that most players that launched the app also initiated a game, and played at least the first SL mini-game.

<sup>8</sup><https://unity.com>

<sup>9</sup><http://learnml.eu/games.php>

<sup>10</sup><https://play.google.com/store/apps/details?id=com.InstituteofDigitalGames.ArtBot>

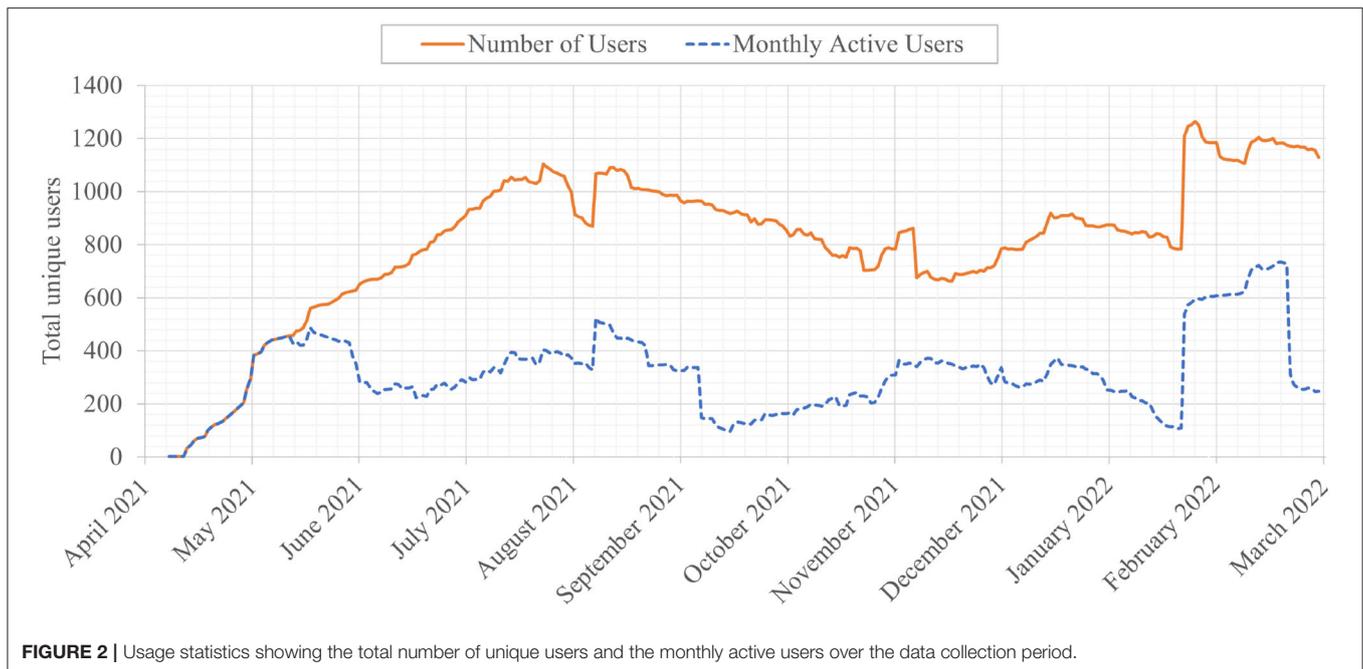
The language change within the game was also monitored, and it was noted that 148 users (i.e., in 6.7% of game starts) switched to Greek, but no user has so far selected Norwegian. The game was launched in English and Greek, then updated with Norwegian localization on 2 September 2021, which has some bearing on this resulting lack of adoption.

### 4.2. Collection of User Evaluations via Surveys

Beyond anonymous data collected from log files, an online survey was also used to examine players' attitudes toward the game. Two surveys were designed: one for students (130 participants) and one for educators (35 participants). Seventy-four percent of the educators identified as female and 26% as male, while for the students 51% identified as male, 45% as female, and 4% preferred not to answer. Mean (*M*) age of the students was 14.3 years old, with a standard deviation (*SD*) of 5. The surveys included open questions regarding the positive and negative aspects of the game, and closed questions (5-item Likert scale) on whether they enjoyed the game and its learning potential. Gaming frequency varied among students and educators, reporting from 0 to 30 h of game playing over the past week for students (*M* = 7.6, *SD* = 10.6) and from 0 to 10 h for educators (*M* = 1.6, *SD* = 2.6). The educators came from a wide range of fields such as physics, language, mathematics, and information technology.

ArtBot was disseminated through the following events and avenues, where the game was demonstrated to participants, the participants were asked to play the game for a few minutes, and then asked to complete an evaluation survey:

- Distributed to secondary education teachers and students of a private school in Athens, Greece, in March 2021 (before the official launch).
- Demonstrated during online workshops with primary and secondary education teachers, mainly in computer science but also teachers from science education, linguistics and arts, in the framework of the Athens Science Festival in March 2021.
- Demonstrated during an online seminar mainly for secondary education teachers, organized by the 3rd Secondary Education Office in Attica, Greece, in May 2021.
- Demonstrated as part of an online teacher training event, in June 2021, organized in Malta. Participants were primary and secondary education teachers from a wide range of fields such as computer science, mathematics, economics, biology, Maltese, and ethics/religion, as well as other stakeholders (e.g., heads of school, school inspectors, researchers) from state, private, and church schools.
- Showcased and tested during a 3-day teacher training event for primary and secondary education teachers in October 2021.
- Showcased during a LearnML Info Day event in October 2021, mainly addressing educators and researchers.
- Shown as part of a keynote speech at the 3rd International Conference on Digital Culture & AudioVisual Challenges, addressing researchers and lecturers from a wide range of academic fields, in May 2021.



**FIGURE 2** | Usage statistics showing the total number of unique users and the monthly active users over the data collection period.

**TABLE 1** | Unique players sorted by platform.

	Total users	Browser	Android	Windows
Unique user visits	2,484	2,265	154	65
Unique user game starts	2,222	2,034	131	57

## 5. DATA ANALYSIS

Since its first launch in April 2021, the ArtBot game has been played by a total of 2,222 unique users. The users' interaction data with the game in general, and its two constituent mini-games around supervised learning (SL mini-game) and reinforcement learning (RL mini-game) are analyzed below, while the feedback of users (students and educators) to surveys solicited during dedicated events (see Section 4.2) is analyzed in Section 5.4.

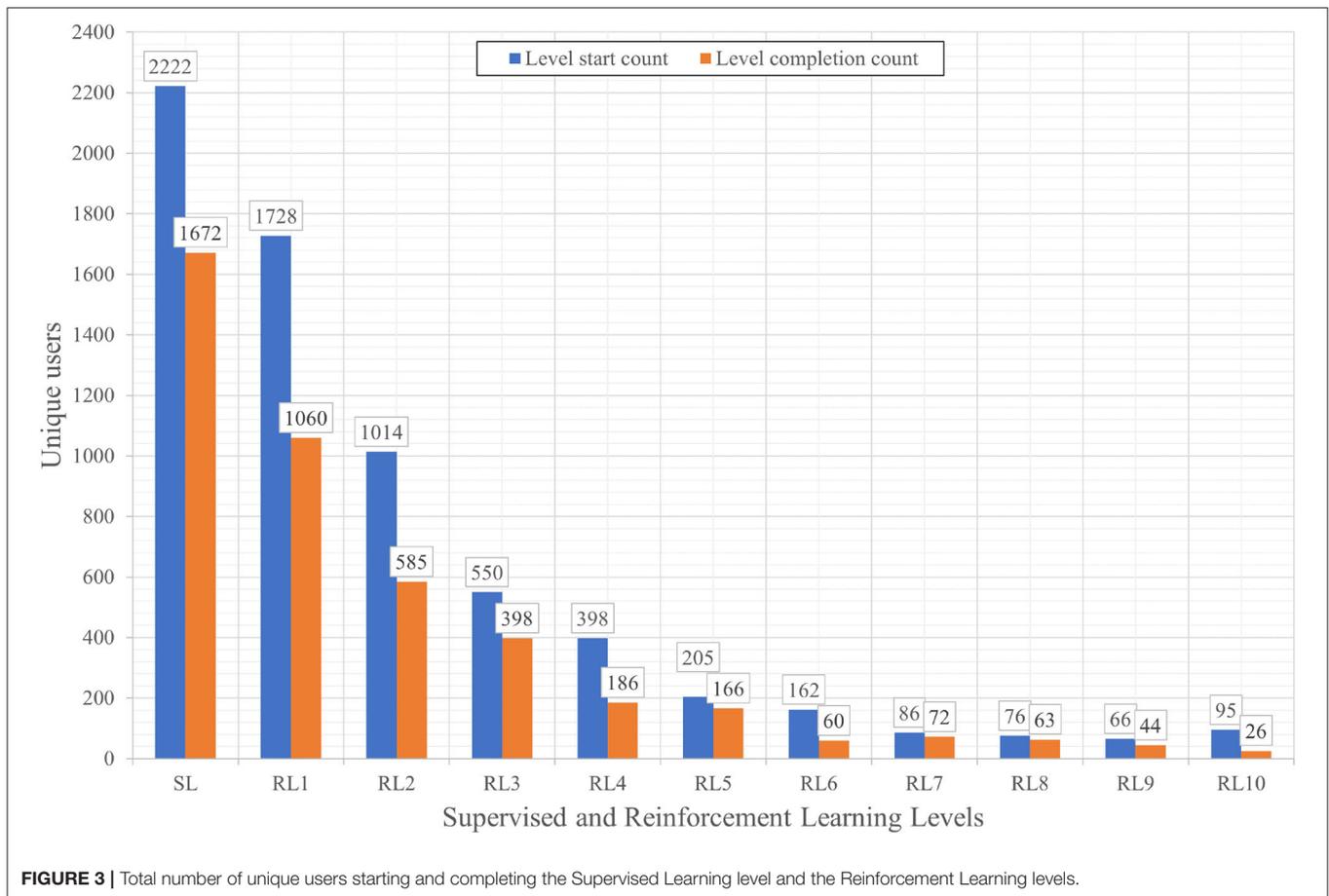
### 5.1. General Use

As a first indication of the engagement of players with the game, we explore how many unique users interacted with different portions of the game. Since the game progresses sequentially from the menu page to the SL mini-game and then to the RL mini-game (the latter consisting of 10 levels), we observe how many users visited each portion in **Figure 3**. The SL game and the first RL level had a large number of players and a relatively high completion rate (82 and 74%, respectively), but the number of players moving on to later levels of the RL game decreases drastically after that. The completion rate remained high, indicating that persevering players were still engaged with the game. However, the high drop in the amount of players might indicate that the different levels of RL mini-game did not offer enough novelty to secure player retention.

Another practical metric for user engagement is the duration of playtime in each level. This proved to be a challenging measure to evaluate for the online playable version of the game. We noticed a number of users with extremely long times spent in each level (e.g., over an hour), which indicates that the game was most probably left running unattended in the browser while the user switched to another activity. Since the browser version was the most popular platform being used (see **Table 1**), this practice introduced a number of outlying data which skewed the statistics. The overall regular usage, however, was frequent enough to mitigate this, and we manually removed the outliers in terms of duration for this analysis.

When considering each level per mini-game, the mean time spent by the player in each of the levels was 4.2 min, although there was considerable variation. The SL mini-game and each level of the RL mini-game had a similar duration (3.8 min on average for SL, 4.4 for each RL level).

At every portion of the game, information screens were made available for additional explanation of how to use the game controls as well as further clarification about the underlying ML process. This information changed at every phase of the mini-game, updating the information with more relevant instructions and facts related to that specific part. Since the information was verbose, the display of these screens was entirely optional, and was displayed at the request of the player. Very few players opened these informational screens; the data indicates that most players who opened the information screen only did so during the first phase of each mini-game. This could be an indication that it was not clear to them that the content of that screen changed during subsequent phases. That said, the help options in the RL mini-game were requested more than two times as often as the respective help options for the SL mini-game. This indicates that



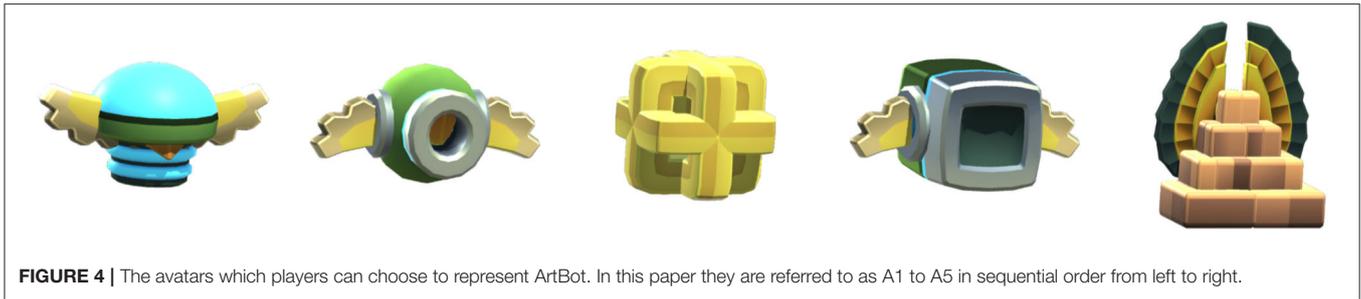
the many parameters that could be tweaked in the RL sessions required more explanation than the few and intuitive options for the SL sessions.

At the beginning of the game the players are given the option to customize their avatar, selecting between five different avatars for ArtBot and three different color schemes per avatar. During the requirements collection stage of the game development process, it was determined that there exists a widespread misconception that AI is used mostly in robots. Consequently, the avatars were intentionally created without any anthropomorphic or highly technological connotations which may misconstrue them as being robots (Zammit et al., 2021). The avatar choices are shown in **Figure 4**. For ease of reference throughout this text the avatars are referred to as A1 to A5, each having three possible color schemes. The default avatar (A1 with the first color scheme) was the most commonly selected (36% of users). However, customization did occur and all possible avatar and color variations were chosen by at least 16 users. The most popular avatars (across all color schemes) were in order: A1 (54% of users), A2 (19%), A4 (11%), A3 (9%), A5 (7%). The more abstract avatars (A3, A5) were less commonly picked, while more anthropomorphic avatars (with bilateral symmetry and a distinction between front and back) were preferred despite our efforts to avoid these connotations.

## 5.2. The Supervised Learning Mini-Game

Due to our concern about the repetitiveness of the manual labeling process in the SL mini-game, we monitored the usage rate of the auto-sort function. Auto-sort was used to label the images automatically in 65% of the games played. Following the completion of an SL mini-game session, 25% of players repeated the labeling process, and 32% of players opted to change the parameters of the algorithm to see how the learned classification changed. These findings indicate that although some players were interested enough in the process to test different settings, the manual labeling activity was either not enjoyable, or does not yield enough of a different outcome for the experience to be considered worth the player's while.

Another important metric is *when* players were choosing to retry the level by changing the settings or relabeling the images. We noted each player's retries and what the resulting accuracy of the testing set was before and after the retry action: 32% of retries resulted in a greater accuracy, 28% regressed to a worse accuracy and the remaining 40% showed no change in accuracy. This suggests that the players were trying to improve their accuracy score, but the high rate of unvarying accuracy also implies that the settings we offer in the game might not allow enough room for improvement.



**FIGURE 4** | The avatars which players can choose to represent ArtBot. In this paper they are referred to as A1 to A5 in sequential order from left to right.

The accuracy of labeling by the players compared to the ground truth was usually above 90%, even when ignoring the high number of auto-sorted runs; almost 99% of users that did manual labeling had 90% or above labeling accuracy on the training set. A few users labeled the images (paintings vs. statues) with lower accuracy, among which 10 unique users classified all paintings as statues and vice versa. Since the classification task itself is very easy for a human being, we deduce that this mislabeling was intentional and exploratory, which was among our initial objectives. Mislabeling the dataset would be a good entry point for discussion about the accuracy of trained models and the role of AI developers and data quality. However, players did not explore such disruptive labeling strategies which indicates the need to add prompts for the players to try to mislabel the data and reflect upon the results. Moreover, since the vast majority of users classified all images correctly (73% of players who did manual labeling), this would indicate that the task is trivial, and why auto-sorting was used to avoid it.

When considering the accuracy of the AI at the end of the supervised learning process, data shows that accuracy on the training dataset was very high, as expected ( $\geq 90\%$  accuracy in 99% of sessions). The accuracy on a hitherto unseen testing dataset however, was much lower on average (67% across all sessions) and varied much more wildly. In 93% of the games played, the test accuracy was between 50% and 90%, and only went higher than that in 4% of sessions. Although it is normal for accuracy to drop during inference on a testing dataset, the reason why it is so pronounced here is probably the small amount (20) of labeled images that are used to train the algorithm. We opted for this small number to strike a good balance between accuracy and the tedium of the manual labeling task if extended to larger training datasets. This shortcoming, however, could act as a trigger for discussion with the learners on the factors affecting the quality of the trained model. A relevant note has actually been added at the information panel regarding the relation of the size of the training set and the accuracy of a model.

An analysis of the algorithmic settings used shows that the full range was used both for the color bin settings (from 4 to 32 in steps of 8) as well as for the maximum depth of the decision tree (1 to 3). The mean number of colors used as inputs to the SL algorithm was 15.0 across all players, with a standard deviation of 9.6. The high variance indicates that the color bin settings was being changed across replays. The tree depth was set to its maximum allowed value of 3 by 87% of the players, while 7.6% of players used a tree depth of 2 and 5.1% used a tree depth of 1.

Overall, the large variance of algorithmic settings, accompanied by the variance in testing accuracy, points to some exploratory behavior from players in order to improve the behavior of the SL algorithm.

### 5.3. The Reinforcement Learning Mini-Game

The ArtBot mini-game focusing on reinforcement learning (RL) is richer in content, comprising of ten distinct levels when compared to the single SL mini-game. It also has a larger set of RL parameters for the player to explore, and its visuals are more congruent with those of commercial games. It was thus foreseeable that the engagement time for this part would be higher, as discussed in Section 5.1.

This mini-game includes a training time for the AI agent (between 30 s and 3 min) which can be chosen by the user. The mean training time set by players varied between 70 and 84 s across the different levels, with later levels (RL7 to RL10) falling closer to the upper end of this range. Since the game only allows training times to be in increments of 30 s, most players chose shorter training times (60 or 90 s).

Since rewards drive the AI behavior in RL problems, we observe what rewards and penalties the players assign to different game objects across the ten levels. Players may assign a reward of values between 1 and 5 to exits, and between -5 and 5 to hazards and collectibles. We monitored the mean and standard deviation of the assigned rewards across the different levels, which yielded useful insight. The standard deviation was consistently high in earlier levels of the RL mini-game (RL1 to RL5), indicating that players were assigning different rewards and penalties to explore how the learning of the agent is affected. In later levels (RL6 to RL10) the variance drops, indicating that those players who made it that far had developed some intuition for the more optimal assignments of these values. The mean reward for reaching the exit was consistently between 3.5 and 3.9 across levels. Players were often not assigning the maximum reward for the exit, in order to allow the agent to explore and obtain rewards from collectibles while learning. The mean reward for reaching collectibles was around 3.5 for the initial levels, then increased to around 4.3 at later levels. This again confirms more frequent variations in the settings in the earlier levels than in later levels, and that collectibles were prioritized over exits in order to promote collection of more statues in each level. The mean rewards for reaching hazards followed a reverse pattern, with negative rewards of  $-3$  at earlier levels, up to  $-4.2$  later in

the game. It also indicates an understanding by the players of the relation between the rewards or penalties assigned and the behavior of the agent.

We noted a different behavior with respect to the algorithm parameters. The variation within each RL parameter across sessions did not change much from level to level. This is interesting, as we were expecting that players would eventually find the best parameter setup and keep it consistent at later levels. The mean learning rate was consistently high throughout ( $>0.9$ ), whereas the exploration rate had the highest variance from level to level, ranging between 76 and 89%. It is reasonable to set a high learning rate when training time is limited. Since the players were trying to collect more treasures before reaching the exit, it also makes sense for the exploration rate to be an intuitive parameter to vary. This indicates that players understood the underlying principles of reinforcement learning, and how human-determined settings have the potential of varying the outcome of an AI algorithm.

We also logged the results of the final runs after training is complete, as they offer additional insights to the individual level design in addition to AI behavior (and players' AI tuning priorities). The exit was found most of the time (above 80% of the time in most levels), but two levels were noteworthy. After training, in level RL4 the agent could reach the exit 44% of the time, indicating that the agent found it difficult to reach the exit across all players' attempts. This level contains the largest number of hazards from all levels, and a long distance between the statues and the exit, which indicates that the level design was indeed more taxing. For level RL9, in contrast, the exit was discovered by the trained agents in all cases across players. This may be in part because players who kept playing the game for nine levels so far were only those dedicated and knowledgeable enough of the RL parameters to achieve such performance; in RL10 for instance the trained agent reached the exit 90% of the time, which is also a high completion rate.

**Figure 5** showcases the differences in layout between these two levels, which seemingly had a strong impact on the behavior of the agent. For RL4, there are spikes in the direct path from ArtBot's starting position to the exit, leading to the low completion rate. For RL9, there is a clear path from ArtBot's starting position to one statue and then the exit, but the other statue is behind a number of spikes and far from the path to the exit, leading to the low collection ratio.

It follows that finding the exit in each level is relatively easy with one exception. On the other hand, the (optional) task of finding all collectibles in the level before reaching the exit is more challenging. Indeed, the ratio of collectibles reached in the best trained agent varied significantly between levels. The most difficult levels in which to accomplish this task were RL10, RL6, RL9 with collection ratios of 10, 16, and 21%, respectively. Interestingly, while RL9 and RL10 were played only by a few (presumably expert) players who managed high completion rates, they had some of the lowest collection ratios; this may point to fatigue from the part of the players regarding the optional task of finding all collectibles.

The game allows players to stop training, which is useful if the learning process does not appear to be productive. The data

shows that players tended to use this function mostly in the first four levels, then the rate gradually decreased; two exceptions were RL6 and RL10, which also had a large number of stops. This is congruent to the previous finding that these two levels were the most difficult in terms of collecting statues, indicating that the users were noticing that only one statue was being collected and stopped the training to revise the settings. In a similar fashion, levels RL6, RL4, RL10 were the ones in which the players most frequently changed the settings and retried the learning after the training was allowed to complete. These results keep underlining the difficulty in reaching the exit in RL4, and the collectibles of RL6 and RL10.

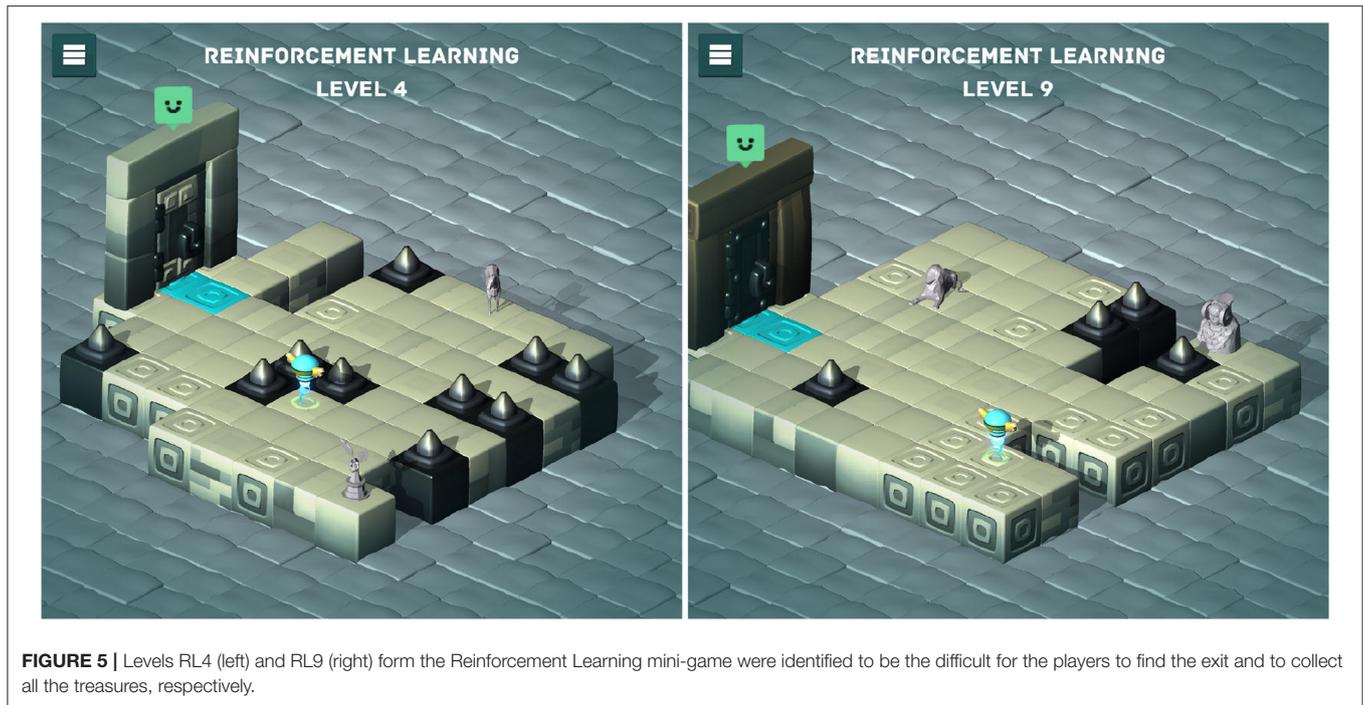
Analogously to what was done for the SL mini-game, we also noted the events leading players to retry the level with different settings. The occurrences of an improved, identical or worse performance upon a level retry were counted. Since the level completion is contingent on the exit being found, an improved collectible count was only considered in this analysis when the exit was also reached before and after the repetition. A newly found exit resulted in 9% of retries, while it was lost after 6% of retries. It is expected that in most cases this value would be unchanged since players tended to repeat the level to get more collectibles. For the latter, 17% of the retries resulted in more collectibles and 10% of retries resulted in fewer. It is still noteworthy that 73% of the time the number of collectibles remained unchanged. This suggests that either the settings offered to the player might not be providing enough agency to change this result, or the lack of such affordance is due to the individual design of each level.

## 5.4. Feedback by Students and Educators

In this section, we analyse the feedback reported by both students and educators to user surveys after these end-users had a chance to play the game in dedicated dissemination events (see Section 4.2). In **Tables 2, 3**, we report results as the mean and standard deviation from survey results in 5-item Likert scales collected from 130 students and 35 teachers, respectively. The students had not been previously exposed to any AI and ML concepts in their classes. Most of the teachers were teaching language or were primary education teachers. Only 10 of the 35 teachers were specialized in fields such as informatics, robotics, technology, and coding. The teachers and students who responded to the survey had no prior involvement with the game; they had not used the game before or taught concepts related to AI and ML.

It seems that the reception of the game was generally positive by students and educators. Both groups found it relatively easy to understand how to play the game, and reported that they would play it again. Students were slightly more reserved regarding whether they would recommend ArtBot to their friends, while many teachers would recommend it to colleagues for use in the classroom. The responses on how fun the game was were fairly positive from all participants.

The learning aspects of the game were also positively received; all respondents reported that it helped them understand more about AI and ML. Teachers were also very positive regarding whether they believed the game would help students understand the concept of AI. The attitudes toward the implementation of



**FIGURE 5 |** Levels RL4 (left) and RL9 (right) from the Reinforcement Learning mini-game were identified to be the difficult for the players to find the exit and to collect all the treasures, respectively.

**TABLE 2 |** Likert scale questions asked to the students in the survey. Scores ranged from 1 (strongly disagree) to 5 (strongly agree).

Question to students	Mean	SD	Pos. (%)	Neg. (%)
It was easy to understand how to play the game ArtBot.	4.2	0.8	88	4
It was easy to play the game.	4.2	0.9	79	3
I would play the ArtBot game again.	3.9	1.1	66	9
I thought the ArtBot game was boring.	2.2	1.1	11	64
I would recommend to my friends to play the game ArtBot.	3.5	1.1	58	18
I could learn how to play the game easily.	4.3	0.9	83	5
I thought the ArtBot game was fun.	4.0	1.1	76	9
The ArtBot game helped me understand how Artificial Intelligence works.	4.0	1.0	75	5
In the game, it was easy to understand what machine learning is.	4.0	1.0	75	5
I would like to use the ArtBot game at school.	4.1	1.1	69	9
I would recommend to my teacher to use the ArtBot game in the classroom so as to learn more about Artificial intelligence.	3.7	1.4	68	22

The mean, standard deviation (SD), and the percentage of positive (>3) and negative (<3) replies of the collected values are reported.

the game in the school environment were again positive; students and teachers reported that they would like to use the game in the classroom and, for the teachers, that they intended to use the game in their teaching.

In freeform responses by respondents, one of the most positive aspects cited by students (20 cases) was the learning aspect of the game. Students reported that they enjoyed the combination

of learning content and game, they enjoyed the fact that they could learn new concepts through a game, that they had to think, solve problems, and “use their brain”. In 17 cases they reported that the game was fun and entertaining, in 12 cases that the graphics and colors were among the positive aspects of the game, and in 4 cases, that they found the game creative. In 15 cases, students reported that they enjoyed their active role, the complexity, and the fact that they had control over the training of the AI agent. Unexpectedly, in 8 cases the learners reported playing with friends as a positive aspect of the game. Since the game is not designed for multiple players, we assume that they referred to the context of playing the game in the same location as their friends; this substantiates the role of the social environment as a motivation for play for young learners (Ferguson and Olson, 2013).

Most of the positive elements of the game reported by teachers were relevant to the learning aspect of ArtBot and its potential to teach students AI and ML concepts and processes through a playful environment (13 cases). The interface, the ease of use, the graphics, and the friendly environment were also among the most cited positive aspects of the game by the teachers (11 cases). Teachers reported that the game environment would attract students and especially students who are already interested in games. Other positive aspects described by the teachers was the explanatory information, the avatar selection, the archaeology aspect, and the fact that the students are manipulating an AI agent.

The most cited negative aspect of the game by students was its pacing; in 9 cases the students described the game as slow, monotonous, time consuming, or boring while in 6 cases, they

**TABLE 3** | Likert scale questions asked to the teachers in the survey. Scores ranged from 1 (strongly disagree) to 5 (strongly agree).

Question to teachers	Mean	SD	Pos. (%)	Neg. (%)
It was easy to understand how to play this game.	3.9	0.9	71	3
The game was easy to use.	3.9	0.9	71	3
I would play this game again.	4.0	1.0	74	9
I found the game boring.	2.2	1.2	17	63
I would recommend to my colleagues to use this game in their classroom.	4.1	1.0	77	6
It was easy to learn how to play this game.	3.9	1.0	74	9
I thought the game was fun.	3.8	0.8	74	6
The game helped me understand how Artificial Intelligence works.	4.2	1.0	83	6
It was easy to understand what machine learning is, through this game.	4.2	0.9	86	6
I would like to use this game in the classroom.	4.1	1.0	77	9
I would like to use this game in my teaching in the future.	4.2	0.8	80	3
I believe that this game will help children understand the concept of Artificial intelligence.	4.3	0.7	89	0

The mean, standard deviation (SD), and the percentage of positive (>3) and negative (<3) replies of the collected values are reported.

thought the game was too complex or too difficult to understand and therefore they needed more guidance. In 3 cases the learners reported that they would like to see more levels or more difficult challenges customized to the learners' age. The graphics, the colors, and the avatar were described as negative aspects in 6 cases. The complexity and the difficulty to understand concepts of AI (e.g., the "use experience" parameter) was also highlighted by teachers as a negative aspect (in 7 cases). Teachers reported that more guidelines and tutorials are needed for explaining the process, variables, and concepts to the students. Teachers also listed the quality of the graphics and sound as negative aspects in 3 cases, and in 2 cases teachers suggested that the game should have more variables for the students to manipulate.

The attitudes of teachers and students were generally positive, although there seem to still be some challenges regarding the complexity of the content and the difficulty to address students and teachers of varying levels of AI expertise (see also Zammit et al., 2021).

## 6. DISCUSSION

The analysis of the data collected has consistently shown a number of important findings regarding the learning and player behavior of the users. The browser platform is evidently very convenient for casual players, as this was by far the most popular for this game. This finding matches our intent to make ArtBot as accessible as possible to a wide audience.

The data indicates that a substantial amount of players meaningfully engaged with both mini-games of ArtBot. Players explored the parameters of the algorithms and were interested to see how manipulating these parameters would vary the outcome of machine learning. The game is therefore successful in its objectives to impart information and awareness about the basics of SL and RL algorithms, their related terminology and processes. However, the lack of interaction with the information screens also shows that it is not evident to users when additional details about the game and background algorithms are made available to them. Based also on feedback by students and teachers in dedicated dissemination events, more effort is needed to better engage the players with the background information and learning content.

The manual image labeling in the SL mini game did not appeal to players, and the outcome of supervised learning does not vary enough with changes in the settings to hold players' attention. In addition, while the first few RL levels were frequently played, the subsequent ones did not offer enough variety or novelty to retain player interest. This could hinder our goal to disseminate the game to a wide audience, and to enhance the learning process by active participation through game-based learning. To address this, additional prompts or datasets closer to the learners' interests could better indicate the role and impact of labeling on the training of the model and the behavior of the agent in the SL process. Similarly, the RL levels could be fine-tuned to provide settings and agent behaviors which are varied and obvious to the players. Additionally, interaction data have brought to light specific design issues with individual levels, such as RL4 and RL9, which require further tuning to align their difficulty to the intended difficulty progression in the level order.

The game-based learning aspects of ArtBot in formal education was positively received by both students and teachers. Both end-users mostly agreed that the implementation of ArtBot in the classroom for teaching and learning about AI and ML would motivate and engage the students, and could play a more active role in their learning. We note that the sample may be biased, especially for the educators, since the survey was completed by teachers who chose to participate in the relevant events; these teachers had, most probably, positive attitudes toward new tools and new concepts. That said, we can assume that games—and particularly ArtBot—can be a useful tool for educators planning to introduce AI education and literacy in their classroom.

### 6.1. Limitations

In this paper, we explored how anonymous usage data and surveys targeting educators and students can be used to gauge how a game designed for imparting AI and ML concepts was perceived by a broad audience. As such, our findings are specific to the game ArtBot and certainly not generalisable to all games on AI and ML. However, this paper aims to provide insights on the design and analytics of games aiming to teach AI and ML concepts to young students, and to highlight the potential of games to teach AI and ML.

On that note, ArtBot is intended to be used by teachers and students in a classroom. Since the data collection process is anonymous, no information about the players' ages or their roles

is available. This gives limited context in which to evaluate the interactions with the game. For example, it would have been interesting to understand which age groups were playing the later levels, and which types of end-users (teachers vs. students) were accessing the additional information.

During the design process of ArtBot, we strove to appeal to a broad age group by using simple graphics, familiar controls, and gameplay that has an immediate reward but can be explored deeper in accordance to the curiosity and understanding of the player. Our objective was a game that is understandable enough for primary school students, yet one that can still offer a challenge to secondary level ones. This versatility comes at the cost that ArtBot can not fully address the distinct pedagogical requirements of one specific group over the other. The game therefore trades off a more targeted teaching approach for a wider reach. This was corroborated by responses from educators in our surveys, as teachers of different topics and at different educational institutions and levels gave very different directions toward improvement of the game.

The browser platform, despite the advantages of neither requiring any installation nor a specific device, posed significant problems in the analysis. Users and sessions cannot be clearly distinguished, especially if users close or reopen their browsers, use different browsers, or even multiple tabs. Furthermore, it is very easy for users to switch to different tabs and leave the game running in the background, returning to it and continuing later. This behavior hinders the analysis due to the misleading timestamps corresponding to the same user and session. Moreover, the number of unique users is likely inflated due to the above behavior and/or the use of incognito windows and cookie blockers.

Regarding the attainment of the learning objectives and understanding of the AI and ML concepts addressed in the game, both teachers and students reported a positive impact in dedicated feedback sessions. However, further tests that combine quantitative (e.g., pre- and post-tests) and qualitative (e.g., interviews) analysis are needed to more objectively examine the learning impact of the game and identifying potential misconceptions.

## 6.2. Future Work

The analysis reported in Section 5 clearly outlined areas where the game can be improved. The SL mini-game can be reformatted to give more weight to user settings, by perhaps introducing different parameters of the algorithm which have a more drastic effect on the learning process, and thereby relying less on image labeling. The RL levels require additional features and a gradual increase in difficulty, with perhaps a better reward system for players, such as a points system with a leader board showing best results, or a list of achievements for the player to accomplish. The additional information button can be highlighted whenever it is populated with new information, or it could be shown automatically the first time that new information is available to the players, in the fashion of a game tutorial. In the latter case, the text would be revised and made less verbose.

Externally to the game, teaching resources and accompanying materials have been developed as classroom aids to enhance

the learning experience with ArtBot<sup>11</sup>. This material can be further developed to address some of the shortcomings identified, namely additional information about the underlying algorithms and their use in different real-world applications. The feedback of both teachers and students reported in Section 5.4 will be an important guide toward improving the educational material to mitigate some of the difficulties in understanding the underlying algorithms and to better connect it to everyday ML uses in students' lives.

## 7. CONCLUSION

ArtBot is part of the toolkit developed through the LearnML project to support experimentation, reflection, and critical thinking about AI and ML to primary and secondary school students. Its goal was to teach the basics of supervised learning and reinforcement learning through a playful and exploratory experience. This paper analyzed how ArtBot has been used by different players since its launch in April of 2021, including the feedback of teachers and students in dedicated playtesting events. Through anonymously collected usage metrics, we identify the usability of the game, its user interface, and design effectiveness. The data revealed that the game was generally well received, having over 2,000 unique users, with the browser version being the most popular platform. This indicates that our efforts to make the game easily accessible were fruitful. The players were largely successful completing the in-game activities; many players explored the various ML parameter setups, but only a few explored additional information to learn more about the topics of ML. The dedicated feedback by students and teachers also indicated a generally positive outlook on the use of ArtBot in the classroom, but also raised concerns regarding the game's pacing and the complexity of some of the concepts introduced. A number of potential areas of improvement were identified, both in broad scope as well as specific design tweaks for each portion of the game. With these findings in hand, the game can be refined to enhance player engagement, and to maximize the benefits of a game-based learning experience in the classroom.

## DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the data has been collected anonymously but pertains to the users who have played the game. Although specific users cannot be identified from the data, we do not wish for the data to be publicly accessible. Requests to access the datasets should be directed to MZ, marvin.zammit@um.edu.mt.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants' legal guardian/next

<sup>11</sup>The LearnML Guidebook is freely available at <http://art-bot.net/teachers/>.

of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

MZ and IV carried out the data collection and analysis reported in the paper. MZ, IV, and AL each contributed to the writing in the various sections of the paper. AL and GY advised on the research direction and the text, and oversaw the implementation, analysis, and authoring process. All authors contributed to the article and approved the submitted version.

## REFERENCES

- Clark, D., Nelson, B., Sengupta, P., and D'Angelo, C. (2009). "Rethinking science learning through digital games and simulations: genres, examples, and evidence," in *Proceedings of the Workshop on Learning Science: Computer Games, Simulations, and Education* (Washington, DC).
- Clarke, D., and Noriega, L. (2003). "Games design for the teaching of artificial intelligence," in *Interactive Convergence: Research in Multimedia* (Prague).
- Druga, S., Williams, R., Breazeal, C., and Resnick, M. (2017). "Hey Google is it ok if I eat you?" Initial explorations in child-agent interaction," in *Proceedings of the Conference on Interaction Design and Children* (New York, NY), 595–600. doi: 10.1145/3078072.3084330
- Ferguson, C. J., and Olson, C. K. (2013). Friends, fun, frustration and fantasy: Child motivations for video game play. *Motivat. Emot.* 37, 154–164. doi: 10.1007/s11031-012-9284-7
- Giannakos, M., Voulgari, I., Papavaslopoulou, S., Papamitsiou, Z., and Yannakakis, G. (2020). "Games for artificial intelligence and machine learning education: review and perspectives," in *Non-Formal and Informal Science Learning in the ICT Era*, ed M. Giannakos (Cham: Springer), 117–133. doi: 10.1007/978-981-15-6747-6\_7
- Glaser, M., Garsoffky, B., and Schwan, S. (2009). Narrative-based learning: possible benefits and problems. *Communications* 34, 429–447. doi: 10.1515/COMM.2009.026
- Hainey, T., Connolly, T. M., Boyle, E. A., Wilson, A., and Razak, A. (2016). A systematic literature review of games-based learning empirical evidence in primary education. *Comput. Educ.* 102, 202–223. doi: 10.1016/j.compedu.2016.09.001
- Hartness, K. (2004). Robocode: using games to teach artificial intelligence. *J. Comput. Sci. Coll.* 19, 287–291. doi: 10.5555/1050231.1050275
- Kafai, Y. B., and Burke, Q. (2015). Constructionist gaming: understanding the benefits of making games for learning. *Educ. Psychol.* 50, 313–334. doi: 10.1080/00461520.2015.1124022
- Lai, E. R. (2011). Critical thinking: a literature review. *Pearsons Res. Rep.* 6, 40–41.
- Mertala, P. (2019). Young children's conceptions of computers, code, and the Internet. *Int. J. Child Comput. Interact.* 19, 56–66. doi: 10.1016/j.ijcci.2018.11.003
- Parker, J., and Becker, K. (2014). "ViPER: Game that teaches machine learning concepts - a postmortem," in *Proceedings of the IEEE Games and Entertainment Media Conference* (Toronto, ON).
- Quinlan, J. (1992). *CA.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann.
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., et al. (2019). Machine behaviour. *Nature* 568, 477–486. doi: 10.1038/s41586-019-1138-y
- Sutton, R. S., and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. Cambridge, MA; London: MIT Press.
- Toivonen, T., Jormanainen, I., Kahila, J., Tedre, M., Valtonen, T., and Vartiainen, H. (2020). "Co-designing machine learning apps in K-12 with primary school children," in *Proceedings of the IEEE International Conference on Advanced Learning Technologies* (Los Alamitos, CA), 308–310. doi: 10.1109/ICALT49669.2020.00099

## FUNDING

This work was supported by the Learn to Machine Learn (LearnML) project, under the Erasmus+ Strategic Partnership program (Project Number: 2019-1-MT01-KA201-051220).

## ACKNOWLEDGMENTS

The authors would like to thank the many anonymous users of ArtBot, as well as all partners of the LearnML project for their feedback throughout the project and their participation in dissemination events for promoting ArtBot to the general public.

- Tsai, Y.-L., and Tsai, C.-C. (2020). A meta-analysis of research on digital game-based science learning. *J. Comput. Assist. Learn.* 36, 280–294. doi: 10.1111/jcal.12430
- Turchi, T., Fogli, D., and Malizia, A. (2019). Fostering computational thinking through collaborative game-based learning. *Multimedia Tools Appl.* 78, 13649–13673. doi: 10.1007/s11042-019-7229-9
- Vartiainen, H., Toivonen, T., Jormanainen, I., Kahila, J., Tedre, M., and Valtonen, T. (2021). Machine learning for middle schoolers: learning through data-driven design. *Int. J. Child Comput. Interact.* 29. doi: 10.1016/j.ijcci.2021.100281
- Voulgari, I. (2020). "Digital games for science learning and scientific literacy," in *Non-Formal and Informal Science Learning in the ICT Era*, ed M. Giannakos (Cham: Springer Nature), 35–49. doi: 10.1007/978-981-15-6747-6\_3
- Voulgari, I., Zammit, M., Stouraitis, E., Liapis, A., and Yannakakis, G. N. (2021). "Learn to machine learn: designing a game based approach for teaching machine learning to primary and secondary education students," in *Proceedings of the ACM Interaction Design and Children Conference* (New York, NY: Association for Computing Machinery), 593–598. doi: 10.1145/3459990.3465176
- Watkins, C. J. C. H., and Dayan, P. (1992). Q-learning. *Mach. Learn.* 8, 279–292. doi: 10.1023/A:1022676722315
- Yannakakis, G. N., and Togelius, J. (2018). *Artificial Intelligence and Games*. (Cham: Springer). doi: 10.1007/978-3-319-63519-4
- Zammit, M., Voulgari, I., Liapis, A., and Yannakakis, G. N. (2021). "The road to AI literacy education: from pedagogical needs to tangible game design," in *Proceedings of the European Conference on Games Based Learning* (Reading: Academic Conferences International Limited), 763–771.
- Zhou, Z., and Liu, S. (2021). *Machine Learning*. Cham: Springer. doi: 10.1007/978-981-15-1967-3
- Zimmermann-Niefield, A., Turner, M., Murphy, B., Kane, S. K., and Shapiro, R. B. (2019). "Youth learning machine learning through building models of athletic moves," in *Proceedings of the ACM International Conference on Interaction Design and Children* (Reading: Academic Conferences International Limited), 121–132. doi: 10.1145/3311927.3323139

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zammit, Voulgari, Liapis and Yannakakis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.