Unlocking AI Literacy: The SMaILE-App Gaming Experience.

Santiago Franco
Dept. of Computer Science
Royal Holloway University of London
santiago.francoaixela@rhul.ac.uk

Maria Giulia Ballatore
Dept. of Mathematical Sciences
'G.L. Lagrange'
Politecnico di Torino
maria.ballatore@polito.it

Sara Bernardini
Dept. of Computer Science
Royal Holloway University of London
sara.bernardini@rhul.ac.uk

Giacomo Como
Dept. of Mathematical Sciences
'G.L. Lagrange'
Politecnico di Torino
giacomo.como@polito.it

Luca Damonte
Economic and Finance
LUISS
ldamonte@luiss.it

Alexandra Neacsu
Dept. of Computer Science
Royal Holloway University of London
alexandra.neacsu.2019@live.rhul.ac.uk

Martina Vanelli
ICTEAM Institute
Université catholique de Louvain
martina.vanelli@uclouvain.be

***Abstract*—AI's promise to become embedded in every aspect of modern life is quickly becoming a reality. However, a significant gap emerges in educational resources tailored for adolescents. This void often cultivates misconceptions about AI's capabilities and potential risks. To close this gap, we present SMaILE-App, a novel educational cross-platform app designed to foster AI literacy among individuals aged 10 to 18. While inclusive, the app design is purposely biased towards rectifying the gender gap in STEM. Applying a constructionist educational framework, SMaILE-App promotes learning AI fundamentals through interactive gameplay, merging entertainment with personalized knowledge acquisition. SMaILE-App comprises a suite of *minigames* in an engaging narrative, each focusing on a different AI concept, as well as interactive creational and instructive modules. SMaILE-App rests on two key ideas: AI is not a singular, all-encompassing term but a spectrum of methodologies, each with its nuances, and AI is not without its limitations.**

***Index Terms*—Artificial Intelligence, Educational Game.**

I. Introduction

The widespread adoption of AI technologies holds both immense potential and inherent risks. Ensuring AI's beneficial impact hinges on equipping individuals with robust digital competencies [1], [2], including expertise in AI. Investing in education to develop these skills is crucial for individual employability, global competitiveness, and the capacity to navigate future technological challenges effectively.

The SMaILE (Simple Methods of Artificial Intelligence for Learning and Education) research project aims to address the urgent need for AI education amid increasing interest from both the public and private sectors. Targeting Generation Z students aged 10-18 in lower and upper secondary schools, SMaILE strives to revolutionize informal educational approaches. By making AI more accessible and engaging through game elements, we can not only boost interest but also build confidence in these subjects, see www.smaile.it.

Furthermore, the gender gap in STEM (Science, Technology, Engineering, and Mathematics) and AI fields is a critical issue affecting workforce diversity and innovation [3]. Studies suggest that this gap emerges in adolescence, here girls may not receive the same level of encouragement or exposure to STEM subjects as boys do [1]. We address this challenge through the design of the SMaILE-App, a gamified educational application tailored for adolescent girls. For reference, a video demonstrating the app is available at the link https://youtu.be/KOPIWFlXa20?si=p8BLf-i9FNDh-Qrn. This paper offers an in-depth exploration of the app, highlighting its educational goals, pedagogical approach, and foundational design principles. Additionally, we present an analysis of initial user feedback, outlining both the strengths and areas for improvement of this educational tool.

While the game is inclusive, special focus has been given to attracting female players. Existing research on gender and gaming suggests that male players typically enjoy elements of mastery, competition, and spatial puzzles, whereas female players are more drawn to games that involve emotional narratives, real-world contexts, and verbal puzzles [1], [4], [5, pp. 103-105]. All those elements are part of SMaILE-App, which incorporates features appealing to both genders through an intuitive interface. This section further delves into the game's elements — story, mechanics, aesthetics, and technology — guided by Shell's elemental tetrad [5].

II. Related Work

There are only a few tools combining educational AI tools and smart devices for a K-12 (5-18 years old) audience. Guerreiro-Santalla et al. [6] describes an interesting albeit smaller more targeted project involving fifty students focused on image recognition. However, this tool targets older students (14-17) with prior programming experience, knowledge of specific software tools, and a strong maths foundation. Naya-Varela et al. [7] integrate smartphones with the Robobo educational robot [8] to craft an AI educational tool for senior K-12 students. Unlike our approach, based on a constructionist gamification, theirs requires either a robot or a computer for simulation. However, both approaches aspire to offer open-source resources for AI educational tool development. Out of the nine minigames in SMaILE-APP, the MeatBot, and Artbot

CRediT author statement: All of the authors collectively engaged in this chapter's conceptualization and methodology phases. MGB, LD, SF, AN, and MV were responsible for the validation, investigation, and resource allocation. Additionally, they were the primary contributors to the original draft. SB and GC oversaw the writing process, and they reviewed and edited the manuscript. LD, SF, and AN contributed to the software development. Furthermore, GC was responsible for the project's administration and funding management. All of the authors discussed the text and revised the final manuscript.

[9] games inspired the School and Supermarket minigames. To our knowledge, this is the first generalist AI-based learning tool for smart devices that does not necessitate programming skills. We aim to acquaint players with AI paradigms using only a smart device, without any prerequisite skills.

III. Educational Theoretical Approach

SMaILE is a three-year project that started in 2021. The first two years were devoted to developing and testing the app, while the third year focused on implementation and evaluation. SMaILE aims to demystify AI, transforming it from an enigmatic technology understood only by experts to a powerful tool that empowers young individuals to unleash their creativity and build practical projects. The project builds on two key ideas: AI is not a singular, all-encompassing term but a spectrum of methodologies, each with its nuances, and AI is not without its limitations.

SMaILE-App is a gaming application explicitly designed for secondary school students, aiming to immerse them in interactive AI activities. The primary objective is to enhance their creativity and self-confidence in understanding and utilizing AI, both within the app environment and in real-world scenarios. Such applications, exemplified by SMaILE-App, play a crucial role in reinforcing fundamental skills such as mathematics, spatial reasoning, and computational thinking [10], thereby equipping the younger generation with essential competencies for the future [11].

Under the cave of Dewey's pragmatism and Piaget's cognitive development theory, SMaILE-App employs a constructionist approach to pedagogy [12], [13]. This approach involves presenting students with simplified real-world problems that could benefit from AI solutions and guiding them through the process of constructing these solutions incrementally. By doing so, the application not only makes learning more engaging but also underscores the interdisciplinary nature of AI, emphasizing the need for a diverse skill set. This approach resonates with Dewey's mantra: "Give the pupils something to do, not something to learn;" and then "learning naturally results" [2, p. 191].

IV. Game Design

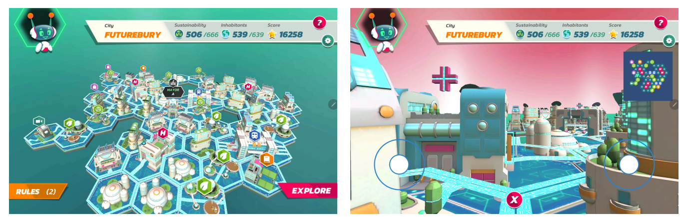
SMaILE-App is a cross-platform app, available on Android and iOS, designed as a single-player city-building simulation. The game aims to engage 10- to 18-year-old students in creating and managing a sustainable, smart city.

A. Story

The game starts with the player assuming the role of mayor for a yet-to-be-built futuristic city. Tasked with crafting a sustainable and livable urban master plan, players build, manage, and maintain various city services and activities. To succeed, they engage in games and challenges designed to impart fundamental AI concepts. Each building type is associated with a distinct activity or *minigame*. Victory in these minigames yields an overall score that combines two rewards: a boost in population and an improved sustainability score. These achievements unlock new building categories for city expansion and present increasingly complex maintenance challenges through a continuous stream of new games.

B. Mechanics

SMaILE-App uses a constructionist approach, therefore the users first play with a specific concept (experience stage), apply it (application stage), and only at the end is free to link it to the theory behind (theory stage). From a pedagogical point of view, each building type corresponds to a specific learning outcome. In particular, each part of the game is based on a fundamental concept of AI. In the TownHall building, players gain access to educational videos after playing the games for the first time. Typically, SMaILE-App offers one theory-focused video and another that provides practical strategies.



(a) Aerial View (b) Immersive View

Fig. 1: Cityscapes

The app features "Moodboard" graphics with a futuristic aesthetic. It employs an *aerial* first-person perspective for city-improving tasks, Fig. 1a, and switches to an *immersive* view for city exploration, Fig. 1b. The sound effects serve a dual purpose: enhancing entertainment during single-player games, and signaling guidance from the robot assistant. The graphic style is deliberately soft and inviting, aligning with our objective to create a more female-friendly gaming experience. The player is at the center of the screen with a close-up city view. Scores for population, sustainability, and overall are in the top-right corner, while the SMaILE-App assistant is ready to offer advice on the left.

C. Technology

The game is designed to be compatible with a wide range of modern devices, including smartphones and tablets. It primarily consists of three canvases: the 3D Aerial cityscape, the 3D exploratory immersive cityscape and the 2D minigame specific frameworks, e.g. Fig. 2. To execute its AI functionalities, the app interfaces with a Linux server where the AI algorithms are run. This approach ensures that the performance of the AI algorithms is not constrained by the user's device hardware, thereby providing a consistent user experience.

V. AI Concepts and Implementation Details

SMaILE-App employs a client-server architectural model. The client side, functioning as the player's graphical user interface (GUI), resides locally on the user's device. Whenever the game demands AI-related processing, the client forwards the relevant parameters to the server side, which then computes and returns the outcomes. As an illustration, if a user inputs a series of Constraint Satisfaction Problem (CSP) rules [14] for the city construction minigame, the server looks for a solution using the built-in CSP model and the user-provided rules. The results are conveyed to the client upon finding a viable solution for the CSP problem. The client then contextualizes these outcomes within the minigame's framework. The client-sever approach is motivated by the following reasons. First, it ensures a uniform user experience irrespective of a device's computational capabilities. Secondly, a majority of the AI tools employed in our game development are not compatible with prevalent mobile operating systems. Finally, it obviates the need to produce and manage multiple code variants tailored to distinct hardware configurations.

SMaILE-App has three distinct stages, each expanding the variety of building types. To advance to the next stage, players must meet a specific score threshold by completing the minigames associated with each building type. Each minigame features a maximum score cap to encourage players to try the new games, ensuring that they can't simply excel in one game to progress to subsequent stages. Unlocking a new minigame grants access to two instructional videos: the first delves into the pertinent AI theory, while the second offers game-specific strategies.

A. CityBuilding: CSP

The *CityBuilding* minigame operates on hexagonal grids with sizes ranging from 6×6 to 10×10. Players gradually define a sequence of CSP rules, with each rule setting the distance between specific buildings and houses. The available buildings include House, Hospital, Park, Supermarket, Station, School, and Cinema, e.g. Fig. 2a. Scoring is calculated according to the proximity of buildings to houses. Each specific building has an associated reward for being near a house. While the primary gameplay objective is maximizing one's score, the pedagogical aim is to foster an intuitive grasp of CSP models by allowing players to adjust a given set of rules.

The general rule template is: "X building-type must have at least N building(s) of type Y at distance $* D$". X and Y represent building categories. N is the number of buildings of type Y, $*$ is an inequality operator from the set $\{=, <, >, \leq, \geq\}$, and D can be either 1, 2, or 3 tile-units. However, distances exceeding 2 units between any service-house pair yield no score. In Fig. 2a, we have: *House must have at least 4 parks at a distance of less than 3*. The video dedicated to CSP theory elucidates the essential elements of CSP, specifically focusing on variables, domains, and constraints formulated as inequalities. To illustrate these concepts, we employ the classical map-coloring example, wherein no two neighboring regions are permitted to share the same color.

B. Hospital: Search and Heuristics

The *Hospital* game is a re-imagining of the well-known Sokoban puzzle game [15], adapted to a hexagonal grid. The game board size varies up to a 9×9 hexagonal grid, featuring an agent (nurse), designated goal cells, and as many beds. The objective is to push all the beds to goal cells in as few moves as possible. Fig. 2b shows one scenario of the twelve increasingly challenging scenarios where the number of beds and overall difficulty incrementally escalate.

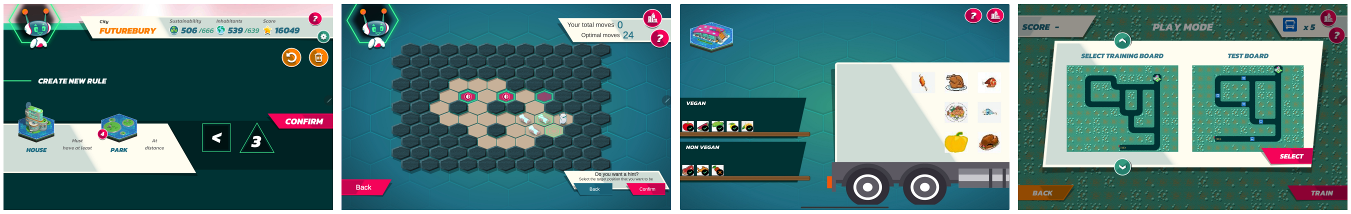
The educational objective of this game is to teach that heuristics serve as a means to simplify complex problems, even though solutions based on heuristics are not necessarily optimal or even valid. The game incorporates an AI *hint* system, which shows a sequence of moves to get a bed to a user-selected goal position. However, these simplified solutions can block a solution for the remaining beds.

C. Park: Combinatorial Game Theory

The park building type contains three minigames: *Park-Selection*, *Planting*, and *Recycling*. In **ParkSelection**, the player's challenge is to strike a balance: selecting options that optimize citizen satisfaction while also minimizing the environmental impact. The game tasks players with balancing city sustainability and resident happiness while building a park. Players rely on citizen feedback, scaled from 1 (not interested) to 10 (very interested), and environmental impact data to make informed decisions. Higher interest scores lead to greater environmental impact. To aid decision-making, the game features a decision tree with varying levels of nodes, from the root at level 0 to leaves at the lowest level. These nodes represent construction options like fields or benches, categorized into groups like *gardens* or *playgrounds*. ParkSelection utilizes the Minimax decision algorithm, as described in [14]. The Minimax algorithm operates on decision trees and is particularly useful in two-player, zero-sum games. It aims to optimize the choices for the maximizing player by considering the best possible counter-moves by the minimizing player. The educational objective of this minigame is to introduce players to decision trees and the broader topic of adversarial search in decision-making problems.

Planting is an enhanced version of *Tic-Tac-Toe* on an 11×11 grid, aiming to connect four tiles instead of three. Players can face off against $Easy$, $Medium$, or $Hard$ opponents, who are trained using an adapted version of the AlphaZero [16] algorithm. Available before the Station game, Planting serves to familiarize the player with the game, acquainting users with varying AI strategies. For further insights, refer to the *Station* section. The game features a single educational video that highlights the limitations of the Minimax algorithm in large search spaces and introduces the idea of progressively improving a virtual player with iterative training.

Recycling acquaints players to the *Nim* game's optimal strategy [17] via a waste-collection challenge. Players take turns removing waste from various piles, vying to claim the last item. The game allows difficulty customization by letting players choose the AI opponent's skill level and board complexity. In terms of AI behavior, the 'Easy' opponent consistently chooses losing moves, 'Medium' alternates between winning and losing moves randomly, while 'Hard' opts for winning moves when available, otherwise picking a single random item. There is a deterministic winning strategy as long as the digital sum of elements is not zero before a move.



(a) CityBuilding Rule Selection (b) Hospital Level Example (c) Supermarket Level Example (d) School Level Example

Fig. 2: minigames Screenshots

An educational video elucidates this by introducing binary arithmetic and XOR operations, equipping players to beat expert-level AI. A supplementary video further clarifies the decision-tree concept by mapping it onto a Nim game tree. An in-game hint button displays the Nim number.

D. Supermarket: Neural Networks and Image Recognition

The *Supermarket* game teaches image classification using Convolutional Neural Networks (CNNs). Users first train an AI to identify vegan and non-vegan items using 30 images, see Fig. 2c. Then, they fine-tune model parameters by setting the number of colors (Colors $= \{2, 4, 8, 16\}$), image resolution (Size $= \{10 \times 10, 20 \times 20, 50 \times 50\}\}$), and neural network layers (Layers $= \{1, 2, 3\}$). The neural network is optimized for a small set of images so that training lasts no longer than 20 seconds. To align with the game's thematic elements, we use cartoonish images, which proved to be effective given the constraints on the training set size and the sub-20-second training time requirement. The theoretical video introduces players to fundamental concepts in Artificial Neural Networks (ANNs). The video elucidates essential elements like image layers, input methods, filter application, and labeling techniques, as well as the distinctions between training and test data sets. It also underscores the significance of salient features and issues a cautionary note on overfitting.

E. Station: Adversarial Games and AlphaZero

In the *Station* game, the player learns to fine-tune the AlphaZero algorithm [16] for creating virtual opponents for the Planting minigame. Players have the flexibility to adjust three key parameters for AlphaZero's training: *rollouts* (choices: $5, 25, 50$), *cuts* (choices: $1, 5, 10$), and *winning threshold* (choices: $0.4, 0.6, 0.8$). Subsequently, they can select other virtual opponents to compete against. These opponents are characterized by displayed values for the same three parameters, giving players insights into their virtual competitors.

In the theoretical video for the *Station* game, players receive a historical overview of AlphaZero, contextualizing it as an evolution of the Minimax algorithms employed in Deep Blue during the 1990s. The Minimax algorithm is previously introduced in the ParkSelection game, while CNNs are discussed in the Supermarket game. The video clarifies the role of CNNs in AlphaZero as heuristics for evaluating the quality of a game state through iterative training. The basics of Monte-Carlo search as a heuristic search algorithm are also covered; this concept is linked to heuristic search, which is initially introduced in the Hospital game. Players learn that Monte-Carlo is a parameterized stochastic search algorithm and gain insights into how *rollouts* function. The core intuition behind AlphaZero is then revealed: it leverages CNNs to refine known strategies and uses Monte-Carlo search for efficient exploration of new moves. In the game-specific video, players are guided through the trade-offs of selecting too low or too high parameter values, connecting these choices to the *exploration vs exploitation* concept.

F. School: Reinforcement Learning

School introduces players to Reinforcement Learning (RL) principles, focusing on the concept of *state* as a collection of features. Modeled on a school bus scenario, the game challenges an autonomous agent to optimize its route for picking up students. Eight incremental levels unlock sequentially, each featuring distinct training and testing maps, e.g. Fig. 2d. The agent navigates a grid comprised of five cell types (road, sidewalk, bus stop, school, and visited) and can move north, east, south, or west, except over sidewalks or outside the map. Its state is defined by the types of adjacent cells and its orientation. We employ Q-learning [18], with a reward function promoting efficient routing to bus stops and schools while penalizing inefficiencies. To ensure accessibility, the video illustrates the agent's learning process, focusing on the RL concepts of exploration and exploitation rather than detailing the Q-function.

G. Cinema: Natural Language Processing

The *Cinema* module serves as a demonstration of AI in an advisory role. Users input a script focused on environmental awareness, and the AI performs two key tasks: it generates a word cloud from the script and matches it against a library of hundreds of educational films created by K-12 students. The system then ranks the five most similar existing movies, which users can watch using the app's built-in video player.

VI. Initial Results and Observations

The research described herein has been duly approved by the Ethical Review Commission at Politecnico di Torino.

A. Setup

The app underwent a three-phase evaluation. Initially, we convened 5 gender-balanced focus groups of 16 children aged 10-13 in the school partner. Participants assessed CityBuilding, ParkSelection, Planting, and Recycling, in terms of mechanics, difficulty, and playability. To focus on gameplay and narrative, we utilized paper versions of the games while separately running the Python CSP engine for partial automation, as in the app. Overall, the story intrigued the participants leading to their engagement in the gameplay; they found CityBuilding both challenging and enjoyable, and the consensus opinion was that it was “fun, like a puzzle”. In the second phase, two gender-balanced focus groups with 20 new students aged 11-12 were conducted in the school partner to assess the app’s interface and usability. The first group only tested three games, and the other group tested the remaining ones. We adjusted the game tutorials and narratives based on the feedback to enhance their accessibility. We also tested the app with a group of 40 students, all aged 10-18, in two separate workshops at public events. In the third phase, the fully completed app was tested by involving 900 middle school students in Piedmont in a Randomized Control Trial (RCT) study.

In this paper, we analyze the results of the Game Experience Questionnaire *GEQ* [19] that involved 450 students accessing the app during the RCT study. 68 students opted not to complete the questionnaire. We prioritize those design features that achieve the primary objective of engaging the target age and gender groups. We employed the *GUESS-13* questionnaire [20], a revised version of GEQ, to assess user experience and satisfaction levels in SMaILE-App. Consisting of thirteen questions, this validated scale evaluates diverse facets of gameplay such as playability, user-friendliness, learning ease, visual appeal, and player engagement. Utilizing this established methodology, we gain valuable insights into the elements that either enhance or detract from the overall quality of the user experience. When comparing groups, statistical significance is determined via t-test, and we used the APA (American Psychological Association) star significance format.

Finally, by combining the GUESS-13 results and the gender data collected, we can also elucidate our gender-specific objective: to design an informal learning environment that prioritizes fostering greater interest in AI among girls.

B. Groups

To analyze the results, we create four categories: *Most Committed*, *Committed*, *Somewhat Committed* and *Uncommitted* (groups 1, 2, 3 & 4). Membership for each group is calculated using *K-Medoids* [21]. To determine engagement with the SMaILE-App as a whole, where each minigame has very different playtimes per session, we created a scoring system that gives each game equal importance. Equation 1 shows the normalized score, where t stands for the playtime in each game, $\mathcal{G}$ stands for the set of minigames and $\mathcal{P}$ for all the players in each game.

$$\text{Score}(player) = \sum_{\forall g \in \mathcal{G}} \frac{t(g, player)}{\sum_{\forall p \in \mathcal{P}} t(g, p)} \tag{1}$$

We consider playtime to represent effective learning time, e.g. excluding menu navigation. Table I displays the scores and total playtime for each group and overall. The table shows there is a strong correlation between the average time spent in the app as a whole and the normalized score in equation 1.

TABLE I: Summary Game Metrics

Group	PlayTime	Score			No.of
	Mean	Mean	Max	Min	Players
1	2h55m	0.078	0.320	0.049	47
2	1h36m	0.036	0.049	0.026	62
3	57m	0.017	0.025	0.010	83
4	19m	0.004	0.010	0.00001	196
All	58m	0.021	0.320	0.00001	388

C. Guess Questionnaire

Table II shows the average score and standard deviation of the GUESS-13 questionnaire and the gender distribution (F/M/O/NA); gender was self-assessed by the participants. Given the limited responses in the *O* and *NA* categories, our gender analysis concentrates on *M* and *F*. Groups 1 and 2 contain a higher percentage of females, validating the app’s success in female engagement. GUESS-13 scores span from -2 (“strongly disagree”) to $+2$ (“strongly agree”), i.e. Likert scale. More committed players generally give higher values. Interestingly, there is a drop in items I1-I13 from groups 1 to 2. However, when running a t-test there is no statistically significant difference when comparing the two groups. The column titled $\%$ represents the average perception of the SMaILE-App, and it is calculated as a percentage of the maximum possible GUESS-13 score. Higher commitment correlates with higher overall satisfaction. SMaILE-App was well received; appreciation grew as commitment increased.

D. Engagement

Table III focuses on GUESS-13 questions with statistical significance between committed players (group A, combining groups 1, 2, & 3) and uncommitted players (group B, also known as 4). Group A participants played for at least close to an hour on average, while group B engaged for less than 20 minutes, also on average. Note that SMaILE-App usage was completely voluntary. Players who engaged for at least 20 minutes found the app more satisfactory, the differences in I1&I7, with 3 stars significance, are particularly striking.

E. Gender

Table IV highlights questions with statistical significance when comparing male (100) and female (87) responses among frequent players. SMaILE-App’s design aims to engage female students, and the results strongly suggest that females found the game more visually appealing, engaging, well-designed, and well-balanced than males did. While both genders agree on the need for enhanced visuals, females distinctly prefer SMaILE-App’s style (I4, I5, I8, I9) and are more inclined to both like it (I1) and recommend it (I13).

VII. Conclusion

This paper presented an in-depth analysis of SMaILE-App, a city-building simulation app designed for educational engagement focusing on female inclusivity. Our findings, validated through GUESS-13 surveys and focus groups, indicate that

TABLE II: **GUESS-13 Questionnaire**

G	F/M/O/NA	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	%
1	24/21/0/2	0.51 (1.16)	0.45 (1.14)	0.60 (1.23)	−0.28 (1.10)	0.34 (1.11)	−0.19 (1.31)	0.51 (1.14)	0.49 (1.04)	0.47 (1.08)	0.28 (1.08)	0.19 (0.95)	0.40 (0.95)	0.23 (1.22)	58
2	32/29/1/0	0.42 (0.97)	0.29 (1.08)	0.37 (0.94)	−0.39 (1.06)	0.16 (1.04)	0.08 (1.00)	0.16 (1.03)	0.48 (1.00)	0.66 (0.94)	0.65 (0.87)	0.21 (0.91)	0.5 (0.9)	0.27 (1.20)	57
3	31/50/0/2	0.42 (0.94)	0.17 (1.10)	0.43 (1.06)	−0.49 (1.15)	0.14 (1.03)	0.05 (1.09)	0.19 (1.13)	0.42 (1.04)	0.42 (1.04)	0.48 (0.97)	0.28 (0.98)	0.41 (0.96)	0.04 (1.27)	56
4	92/101/1/2	−0.04 (1.13)	0.11 (1.18)	0.32 (1.03)	−0.47 (1.07)	−0.11 (1.06)	0.02 (1.17)	−0.14 (1.15)	0.23 (1.05)	0.51 (1.09)	0.36 (1.07)	0.08 (1.03)	0.23 (0.99)	−0.24 (1.27)	52
All	179/201/2/6	0.2 (1.09)	0.19 (1.14)	0.38 (1.05)	−0.44 (1.09)	0.04 (1.07)	0.01 (1.14)	0.06 (1.15)	0.35 (1.04)	0.51 (1.06)	0.42 (1.03)	0.16 (0.99)	0.33 (0.97)	−0.04 (1.27)	54

G: Group. F/M/O/NA: Female/Male/Others/NotAnswered. I1:I like to play. I2:Easy to learn. I3:Easy to use. I4:Visually appealing. I5:Engaging. I6:Hard. I7:Fun to play.I8:Well Balanced. I9:Well Designed. I10:Well Organized. I11: Reactive. I12:Stable. I13:Recommend. %: App's average perception

TABLE III: **Comparison By Frequency of Play.**

G	I1***	I5**	I7***	I8*	I12*	I13**
A	0.44 (1.00)	0.20 (1.05)	0.26 (1.10)	0.46 (1.02)	0.44 (0.94)	0.16 (1.24)
B	−0.04 (1.13)	−0.11 (1.06)	−0.14 (1.15)	0.23 (1.05)	0.23 (0.99)	−0.24 (1.27)

A: Players in groups 1 &2 & 3. B: Players in group 4.
I1:I like to play. I5:Engaging. I7:Fun to play.
I8:Well Balanced. I12:Stable. I13:Recommend.

TABLE IV: **Comparison by Gender.**

G	I1*	I4**	I5**	I8**	I9**	I13*
F 87	0.63 (0.93)	−0.13 (1.08)	0.45 (0.92)	0.72 (0.89)	0.75 (0.91)	0.41 (1.17)
M 100	0.34 (1.01)	−0.62 (1.10)	0.04 (1.10)	0.26 (1.10)	0.33 (1.09)	0.02 (1.24)

Only groups 1 & 2 & 3 being considered.
I1:I like to play.I4:Visually appealing. I5:Engaging.
I8:Well Balanced. I9:Well Designed. I13:Recommend.

the app has been successful in engaging players aged 10-18, particularly females. The app's integration of AI concepts not only adds a layer of complexity but also serves pedagogical purposes, making it a unique addition to the educational gaming landscape. However, there are areas for improvement, particularly in graphics quality. Future work includes expanding the app's educational capabilities and using more sophisticated statistical analysis methods. The positive reception from committed players and statistical significance in gender preferences provide a promising avenue for the app's development and potential impact in educational settings.

ACKNOWLEDGMENTS

This research was supported by the *Fondazione Compagnia di San Paolo*. The authors are grateful to their project partners Convitto Nazionale Umberto I di Torino, AIACE, USR Piemonte, and MelaZeta. They also thank the participating schools and the individuals who contributed, including N. Sapio, D. Battaglia, R. Goyal, and S. Chopra.

REFERENCES

[1] W. Admiraal, J. Huizenga, I. Heemskerk, E. Kuiper, M. Volman, and G. Dam, "Gender-inclusive game-based learning in secondary education," *International Journal of Inclusive Education*, vol. 18, no. 11, pp. 1208–1218, 2014.

[2] J. Dewey, *Democracy and Education: An Introduction To The Philosophy of Education*. New York: The Free Press, 1916.

[3] J. S. Eccles, "Where are all the women? gender differences in participation in physical science and engineering," in *Why aren't more women in science? Top researchers debate the evidence*. American Psychological Association, 2007, pp. 199–210.

[4] J. Robertson, "Making games in the classroom: Benefits and gender concerns," *Computers & Education*, vol. 59, pp. 385–398, 2012.

[5] J. Shell, *The Art of Game Design: A Book of Lenses*. New York: A K Peters/CRC Press, 2019.

[6] S. Guerreiro-Santalla, A. Mallo, T. Baamonde, and F. Bellas, "Smartphone-based game development to introduce k12 students in applied artificial intelligence," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 11, pp. 12 758–12 765, 2022.

[7] M. Naya-Varela, S. Guerreiro-Santalla, T. Baamonde, and F. Bellas, "Robobo smartcity: An autonomous driving model for computational intelligence learning through educational robotics," *IEEE Transactions on Learning Technologies*, 2023.

[8] F. Bellas, M. Naya, G. Varela, L. Llamas, A. Prieto, J. C. Becerra, M. Bautista, A. Faina, and R. Duro, "The robobo project: Bringing educational robotics closer to real-world applications," in *Robotics in Education: Latest Results and Developments*. Springer, 2018, pp. 226–237.

[9] LearnML, "Learnml games," https://learnml.eu/games.php, 2023, accessed: 2023-08-22.

[10] N. Newcombe, "Harnessing spatial thinking to support stem learning," *OECD Education Working Papers*, vol. 161, 2017. [Online]. Available: https://doi.org/10.1787/7d5dcae6-en

[11] R. Williams, H.-W. Park, L. Oh, and C. Breazeal, "Popbots: Designing an artificial intelligence curriculum for early childhood education," in *9th Symposium on Educational Advances in Artificial Intelligence*, 2019.

[12] S. Papert, *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books, Inc., 1980.

[13] M. Resnick, *Lifelong Kindergarten: Cultivating Creativity Through Projects, Passion, Peers, and Play*. Cambridge, Massachusetts: MIT Press, 2017.

[14] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Pearson, 2016.

[15] U. Studio, "Sokoban online," https://www.sokobanonline.com/, 2023, accessed: 2023-08-22.

[16] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel *et al.*, "A general reinforcement learning algorithm that masters chess, shogi, and go through self-play," *Science*, vol. 362, no. 6419, pp. 1140–1144, 2018.

[17] E. R. Berlekamp, J. H. Conway, and R. K. Guy, *Winning Ways for your Mathematical Plays*. AK Peters/CRC Press, 2001.

[18] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, pp. 279–292, 1992.

[19] W. A. IJsselsteijn, Y. A. De Kort, and K. Poels, "The game experience questionnaire," https://research.tue.nl/en/publications/the-game-experience-questionnaire, 2013.

[20] M. H. Phan, J. R. Keebler, and B. S. Chaparro, "The development and validation of the game user experience satisfaction scale (guess)," *Human factors*, vol. 58, no. 8, pp. 1217–1247, 2016.

[21] L. Kaufman and P. Rousseeuw, "Clustering by means of medoids," in *Proceedings of the statistical data analysis based on the L1 norm conference, neuchatel, switzerland*, vol. 31, 1987.