

Actas del IV Congreso Español de Videojuegos (CEV'25)
Granada, España

Raúl Lara-Cabrera, Antonio M. Mora

18–19 de septiembre de 2025

Publicado por:

Sociedad Española para las Ciencias del Videojuego

<https://secivi.org>

Créditos:

L^AT_EX editor: Raúl Lara-Cabrera

usando el paquete de L^AT_EX's 'confproc', version 0.8

This work is licensed under a [Creative Commons “Attribution-NonCommercial-ShareAlike 4.0 International”](#) license.



Prólogo, Presidente del CEV'25

La cuarta edición del Congreso Español de Videojuegos (CEV), celebrada en Granada los días 18 y 19 de septiembre de 2025 reunió a una gran representación de investigadores e investigadoras españoles/as de las diferentes ramas relacionadas con el estudio y el desarrollo de videojuegos.

Los videojuegos tienen, por naturaleza, un marcado carácter multidisciplinar y, en este contexto, CEV se ha consolidado como un espacio de encuentro para investigadores de muy diversas áreas de conocimiento, entre ellas la Ingeniería, la Arquitectura, las Artes y Humanidades, las Ciencias de la Educación y las Ciencias Sociales y Jurídicas. Todas ellas comparten un interés común por los videojuegos, ya sea como objeto de estudio o como herramienta para alcanzar otros objetivos, tanto en el ámbito del entretenimiento como en el de la gamificación y los juegos serios.

En cuanto al contenido de estas actas, en la presente edición destaca la consolidación de trabajos que abordan la aplicación de tecnologías y enfoques propios de los videojuegos en contextos distintos al lúdico, como el artístico, el cultural, el educativo o incluso el terapéutico. Así como, por supuesto, la aplicación de los incipientes LLMs (Large Language Models) o técnicas tradicionales de Inteligencia Artificial y Aprendizaje Automático (Machine Learning).

De esta forma, CEV ha continuado avanzando en su vocación de servir como punto de encuentro multidisciplinar, complementando a otras conferencias más especializadas en áreas tradicionales.

Antonio Mora, Pablo García, Nuria Medina

Organizadores del CEV

Comité Organizador

- Presidente del congreso: Antonio Mora García
- Presidente del comité de programa: Pablo García Sánchez
- Presidenta de organización local: Nuria Medina Medina
- Organización local: Antonio M. Mora García, Pablo García Sánchez, Nuria Medina Medina

Comité de Programa

- Alejandro Lozano Muñoz, *Universidad de Salamanca*
- Ana María Botella Nicolás, *Universidad de Valencia*
- Antonio Fernández Ares, *Universidad de Granada*
- Antonio Mora, *Universidad de Granada*
- Antonio A. Sánchez-Ruiz, *Universidad Complutense de Madrid*
- Antonio César Moreno, *Universidad Complutense de Madrid*
- Antonio J. Fernández Leiva, *Universidad de Málaga*
- Carlos Blanco Bueno, *Universidad de Cantabria*
- Carlos Marín Lora, *Universitat Jaume I*
- Elena Shliakhovchuk, *Universitat Jaume I*
- Federico Peinado, *Universidad Complutense de Madrid*
- Francisco Ignacio Revuelta Domínguez, *Universidad de Extremadura*
- Francisco J. Gallego-Durán, *Universidad de Alicante*
- Francisco Luis Gutiérrez Vela, *Universidad de Granada*
- Guillermo Paredes, *Universidad de Sevilla*
- Gustavo A. Patow, *Universitat de Girona*

- Irene Camps-Ortueta, *Universidad Complutense de Madrid*
- Iria Cabrera Balbuena, *Escuela Superior Politécnica del Litoral*
- Isabel Barbancho, *Universidad de Málaga*
- Ismael Sagredo Olivenza, *Universidad Internacional de La Rioja*
- Jesús Mayor, *Universidad Politécnica de Madrid*
- Jorge Echeverria, *San Jorge Universidad*
- José Iván San José Vieco, *Universidad Isabel I de Burgos*
- Josefa Díaz Álvarez, *Universidad de Extremadura*
- Josefa Ros Velasco, *Universidad Complutense de Madrid*
- Juan Antonio Caballero-Hernández, *Universidad de Cádiz*
- Julio M. Álvarez-Bautista, *Universidad Complutense de Madrid*
- Manuel Palomo-Duarte, *Universidad de Cádiz*
- Marco A. Gómez-Martín, *Universidad Complutense de Madrid*
- María Del Carmen Gertrudis, *Universidad Rey Juan Carlos*
- María Eugenia Larreina Morales, *Universitat Autònoma de Barcelona*
- Marta Fernández Ruiz, *Universitat Politècnica de Catalunya*
- Miguel Ángel Chover Sellés, *Universitat Jaume I*
- Miguel Ángel Roque López, *Universidad de Castilla-La Mancha*
- Nuria Medina Medina, *Universidad de Granada*
- Pablo García-Sánchez, *Universidad de Granada*
- Patricia Paderewski, *Universidad de Granada*
- Pedro Antonio González Calero, *Universidad Complutense de Madrid*
- Pedro Castillo, *Universidad de Granada*
- Raul Lara-Cabrera, *Universidad Politécnica de Madrid*
- Rubén Baena Perez, *Universidad de Cádiz*

- Salvador Fernández González, *Universidad de Málaga*
- Sergio Gutiérrez Manjón, *Universidad Complutense de Madrid*

PROGRAMA

Día 1

Sesión J1 – Trabajos Cortos

Jana María Pérez Romero

La obsolescencia en la estética videolúdica: vanguardia indie vs. hiperrealismo mainstream 1

María del Pilar Bermúdez, Roberto Martín

Explorando el diseño narrativo en videojuegos educativos: el caso Mucho más que abejas 7

Daniel Castilla-Rodríguez, María del Pilar Bermúdez Martínez

Serious games y estudios culturales: desarrollo participativo y académico de *ñJuana I*: la puerta de la reina 13

Juan Antonio Roa-Domínguez, Juan Antonio Caballero-Hernández, Manuel Palomo-Duarte, Lucía Benítez-Eyzaguirre

Diseño de un videojuego en Twine para la exploración de imaginarios utópicos 19

Iñigo Gómez, Daniel Castilla, Roberto Martín

Integración del Patrimonio Cultural en el Diseño de Videojuegos Narrativos 23

Oriana Laplana Curtido

Playing to Raise Awareness: A Narrative Video Game Against Meningitis 31

Sesión J2 – Trabajos Regulares

Alberto Contreras Puerto, Manuel Palomo-Duarte, Andrés Muñoz Ortega, Juan Antonio Caballero-Hernández

Análisis de Sentimientos y Polarización de Reseñas de Videojuegos a Través de Inteligencia Artificial 35

Joan Arnedo-Moreno, Paula Rodríguez-Rivera

Una revisión general del estado de la investigación sobre el vínculo entre videojuegos y juegos de apuestas 45

Nàdia Montaña-Miranda, María Beatriz Villar-López, Carlos Marín Lora, Micaela Y. Martín, Miguel Chover

De juego móvil a realidad virtual en cinta de correr: adaptación y evaluación de la experiencia de juego 57

Sesión J3 – Trabajos Regulares

David Rivera, Roi Quintas, Miguel Ramírez, Claudia Zarzuela, Raúl Pérez, Carlos León

Large Language Models for Dynamically Reacting to Unexpected Video Game Interactions 69

Pedro León Miranda, Matt Castellanos Silva, Ismael Sagredo-Olivenza

AI Crossing: LLM integration in NPC interaction with the designer's assistance 81

Cristina Del Águila Martín, Nuria Medina Medina, Patricia Paderewski Rodríguez, Alberto Hitos Garcia, Francisco Luis Gutiérrez Vela

Co-design of a serious game with Alexa for older adults 93

Día 2

Sesión V3 – Trabajos Regulares

Roberto Martín, Mario Alaguero, Iñigo Gómez

Semiótica y estética en la expresividad de personajes en videojuegos educativos: el caso de Indagapp 107

Anabel Cambero, Ismael Sagredo-Olivenza, Lara Sánchez-Coterón

El simbolismo grotesco surrealista en el diseño de videojuegos experimentales 113

Pablo Sánchez-Martín, Alejandro Barrachina-Argudo, Alejandro Romero-Hernández, Ismael Sagredo-Olivenza

Motion capture for gesture detection in virtual environments 127

Sesión V4 – Trabajos Regulares

Sergio J. Viudes-Carbonell, Francisco J. Gallego-Durán, Faraón Llorens-Largo

From Clinic to Smartphone: Clinical Efficacy of a Tilt-Controlled Mobile Serious Game for Overcoming Convergence Insufficiency 139

Pedro Juárez-González, María José Cano-Iglesias, Antonio Joaquín Franco-Mariscal

Elemental Home, un videojuego serio para educar en sostenibilidad ambiental 151

Rocío Barragán, Antonio Mora, Pablo García-Sánchez

Optimización de un agente autónomo para combates de Pokémon 163

Índice de autores

175

La obsolescencia en la estética videolúdica: vanguardia *indie* vs. hiperrealismo *mainstream*

Jana María Pérez Romero¹

¹ Programa de Doctorado en Historia y Artes, Especialidad en Cultura Artística de la Universidad de Granada, España.

Resumen

La obsolescencia estética en el medio videolúdico, analizada en función de sus características teórico-artísticas, así como de su sinergia con la narrativa y del efecto que produce en el jugador, no ha sido suficientemente explorada. Así pues, esta comunicación, derivada de una tesis doctoral en curso, propone estudiar el concepto desde una perspectiva hermenéutica y cualitativa, apoyándose en los planteamientos de Walter Benjamin, Jean Baudrillard y Wassily Kandinsky, quienes anticiparon fenómenos tan actuales como la pérdida del aura en la obra de arte, el agotamiento de la experiencia estética por sobreestímulo o la disolución de la espiritualidad artística a causa de la imitación vacía de la naturaleza. Este marco teórico se aplica al análisis de una muestra seleccionada de videojuegos desarrollados entre 2005 y la actualidad, cuyos resultados iniciales apuntan a que los títulos con una estética alternativa, como la empleada por la industria independiente, presentan una mayor longevidad sensorial frente a aquellos hiperrealistas cuyo diseño está íntimamente ligado a la evolución tecnológica del medio. Esta diferencia parece responder al modo en que la pequeña industria integra elementos estéticos y narrativos con mayor ingenio, mientras que la gran industria centra sus esfuerzos en liderar la carrera técnica, relegando en ocasiones la coherencia narrativa, el refinamiento mecánico o la estabilidad de sus productos. En consecuencia, aunque los videojuegos comerciales alcancen una calidad gráfica superlativa, tienden a mostrar un acusado declive en su rejugabilidad, frente a aquellos con una estética alternativa, capaces de generar vínculos afectivos duraderos que favorecen su continuidad en el ecosistema videolúdico.

Palabras clave

Estética, videojuego, arte, indie, mainstream, obsolescencia

1. Introducción

Aunque el medio videolúdico atrae cada vez a más investigadores de los ámbitos más diversos, autores como Graeme Kirkpatrick, Frank Lantz, John Sharp, Mateo Terrasa Torres, Jesper Juul, Sonia Fizek, Ximena Hidalgo Vásquez, Ruth García, Clara Fernández-Vara, José Enrique Romero Ocaña o Laura Baigorri han abierto el camino hacia una lectura simbólica del videojuego, que permite abordar también su dimensión estética. Pese a ello, la teoría artística aplicada al medio videolúdico sigue estando en minoría frente a estudios centrados en la tecnología, la pedagogía, el mercado o la experiencia del usuario, entre otros. Por ello, esta

¹ Comunicación realizada para el IV Congreso español de videojuegos.

✉ janacyram@correo.ugr.es / janampromero@gmail.com

ORCID [0009-0006-8173-4939](https://orcid.org/0009-0006-8173-4939) (J.M. Pérez Romero)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

comunicación se propone reforzar esa línea a través de un análisis crítico centrado en sus valores estéticos y en su dimensión cultural, desde el ámbito de la historia del arte.

De este modo, la presente investigación tiene como objetivo principal ahondar en el concepto de obsolescencia estética, derivado del análisis de las etapas vitales del videojuego —preproducción, lanzamiento y primera experiencia de juego, cristalización experiencial, desapego por desgaste y redescubrimiento— desarrolladas con mayor profundidad en la tesis doctoral en curso. Dicho lo cual, se busca indagar en la diversidad estética de los videojuegos creados por la industria independiente —en adelante, *indie*, entendida como aquella que opera con personal y medios limitados— frente a la homogeneidad hiperrealista de la gran industria —en adelante, *mainstream*, caracterizada por disponer de grandes equipos y abundancia de recursos—; y por otro, valorar cómo estas diferencias estéticas influyen en la experiencia del jugador a lo largo de dichas fases.

2. Metodología

Partiendo de un enfoque hermenéutico y cualitativo, esta comunicación analiza una selección puntual de videojuegos lanzados entre 2005 y la actualidad. Dada la brevedad del texto, se comparan especialmente *Planet Crafter* (Miju Games, 2022) y *Starfield* (Bethesda, 2023), por ofrecer un contraste estético claro. También se han considerado, por su relevancia, otros títulos como *RimWorld* (Ludeon Studios, 2018), *Anno 1800* (Ubisoft, 2019), *Hyper Light Drifter* (Heart Machine, 2016) y *Diablo IV* (Blizzard Entertainment, 2024), elegidos según criterios de género, mecánicas, estética y procedencia industrial.

El análisis contempla variables estéticas como colorimetría, composición, perspectiva, estilo gráfico —incluidos HUD e interfaz— y su adecuación iconológica al género narrativo. La obsolescencia estética se relaciona con la cantidad de jugadores activos, usando fuentes como SteamCharts, Steamdb y Play Counter. Así, títulos técnicamente sofisticados, pero con escasa comunidad activa se perciben como envejecidos más rápidamente.

El marco teórico recurre a Walter Benjamin, Jean Baudrillard y Wassily Kandinsky, cuyas ideas permiten abordar la relación entre reproducción, mimesis y valor estético. Benjamin advierte que la reproducción técnica disuelve el aura al uniformar lo único [2]; Baudrillard define la simulación como desaparición de lo real en favor del modelo [1]; y Kandinsky defiende la capacidad del arte para transmitir el valor interno del entorno más allá de la imitación [3]. Estas referencias no se aplican a un título concreto, sino que fundamentan la metodología propuesta para valorar la vigencia estética y su impacto en la experiencia del jugador.

3. La pérdida del aura

No puede analizarse el videojuego como objeto artístico sin atender a los teóricos que alertaron sobre la pérdida del aura con el avance tecnológico. Baudrillard sostiene que la hiperrealidad en títulos bélicos como *Call of Duty* genera un «cortocircuito de la realidad», al eliminar la distancia crítica entre lo real y lo racional [1].

A partir de esta premisa, el estudio identifica dos grandes tendencias estéticas: el hiperrealismo de la industria *mainstream* y la experimentación visual de la independiente. En el primer caso, la estética está supeditada a la tecnología, lo que genera variaciones de color, luz y perspectiva entre entregas que desembocan en una competencia interna y cierta inconsistencia gráfica (ver Figura 1). Según Kandinsky, esta repetición pierde valor cuando cambia la atmósfera que le daba sentido, anestesiando al jugador con estímulos previsibles [3].



Figura 1: Detalle de las texturas de agua, estructuras y del cielo nuboso. Izquierda, *The Witcher: Enhanced Edition* (GOG.com). Derecha, *The Witcher 3: Wild Hunt* (Steam).

En la escena independiente, las limitaciones técnicas suelen alargar el desarrollo, pero permiten un control más detallado y una propuesta estética más coherente. Ello da lugar a estilos únicos que expresan ideas complejas, dotando al proyecto de una identidad reconocible [3]. Al centrarse en el mensaje antes que en la apariencia, estos títulos suelen mantenerse vigentes al margen de las modas, fidelizando a su comunidad y reduciendo su obsolescencia estética.

En el segundo caso, la industria independiente compensa su infrafinanciación y escasez de personal con un proceso creativo más largo, pero también más cuidado, que permite un control de calidad minucioso. Esto da lugar a una gran variedad de estilos que expresan conceptos abstractos complejos, dotando al proyecto de una identidad propia [3]. A diferencia del hiperrealismo comercial —salvo excepciones—, esta estética no depende de las modas, sino del mensaje, lo que favorece una comunidad fiel y una menor obsolescencia estética.

En esta línea, destaca el fenómeno del *demake*, que reinterpreta videojuegos actuales con gráficos propios de generaciones anteriores. Lejos de ser un simple ejercicio nostálgico, prioriza la jugabilidad frente a la carga gráfica y confirma la hipótesis sobre el desgaste del hiperrealismo. Retoma así la idea de Kandinsky sobre el valor interno del arte frente a la mimesis, en sintonía con el planteamiento del arte conceptual.

4. Humo y espejos

Una tendencia observada en esta investigación se relaciona con la cantidad de jugadores que mantienen activos ciertos videojuegos, donde la principal diferencia radica en la estética, pese a compartir género y mecánicas similares. El ejemplo comparativo elegido es *Starfield* (Bethesda, 2023) y *Planet Crafter* (Miju Games, 2022). *Starfield*, un título de la industria *mainstream* con más de 500 personas en su equipo y una inversión superior a 200 millones de dólares [4], lanzó una campaña promocional masiva. *Planet Crafter*, creado por un pequeño

estudio de dos personas —ahora seis—, optó por técnicas procedurales y una estética realista, pero menos mimética que *Starfield* (ver Figura 2).

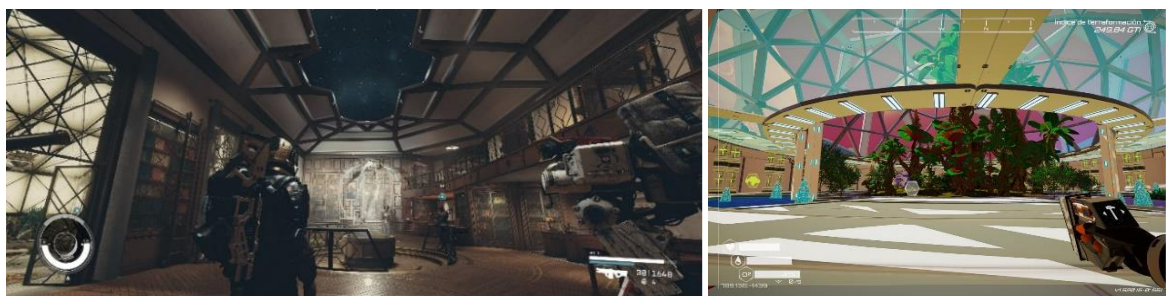


Figura 2: Captura de pantalla de distintos interiores. Izquierda imagen de *Starfield* (Steam, usuario TheReayning). Derecha, imagen de *Planet Crafter* (Steam, propia).

Por otro lado, *Starfield* emplea técnicas cinematográficas: planos generales para bases, americanos y primeros para personajes, conjuntos para misiones y cenitales para naves [5]. Las texturas simulan reflejos, calibración cromática y diseño gráfico en el HUD [6], con resolución *ultrawide* y franjas negras, adaptando la colorimetría al entorno. *Planet Crafter*, en cambio, evita la mímesis filmica: opta por colores planos y formas simples que previenen la cinetosis [7], reduciendo el malestar propio de la vista en primera persona.

Según Steamcharts [8], *Starfield* alcanzó un pico de 330.597 jugadores, frente a los 1.280 de *Planet Crafter*, constantes desde el año anterior. Sin embargo, su rápida caída igualó su actividad, y desde entonces *Planet Crafter* lo supera de forma sostenida, con entre 3.000 y 6.000 jugadores en las últimas 48 horas. Las valoraciones negativas hacia Bethesda contrastan con las positivas hacia Miju, lo que sugiere que el hiperrealismo técnico puede enmascarar debilidades estructurales que acaban por aflorar. No se afirma que los AAA sean inferiores ni carentes de mérito técnico o estético —muchos ofrecen experiencias visuales impactantes—, pero los datos sugieren que la estética *indie*, más contenida y menos dependiente de la hiperrealidad, deja una huella más duradera en la experiencia del jugador, favoreciendo comunidades activas y satisfechas a largo plazo.

5. Conclusiones

La presente investigación confirma la hipótesis según la cual la estética videolúdica se vuelve obsoleta cuando carece de un componente interno que trascienda lo tecnológico, estableciendo un vínculo emocional que impulse al jugador a revisitar el juego o incluirlo en su rutina. Usar el arte solo como vehículo de beneficio económico mediante estéticas hiperrealistas que explotan el avance técnico desaprovecha el potencial expresivo del videojuego independiente, que dota a la obra de significado, identidad y perdurabilidad.

Lejos de ser una crítica al videojuego AAA, este ejercicio comparativo busca destacar cómo ciertas estrategias visuales alternativas generan experiencias más duraderas y comunidades más estables. Esta propuesta no ofrece una conclusión cerrada, sino que esboza una hipótesis interpretativa abierta a futuras investigaciones, donde podrían incorporarse más casos y metodologías.

6. Referencias

- [1] J. Baudrillard, Cultura y simulacro. Editorial Kairós, 2024.
- [2] W. Benjamin, La obra de arte en la época de su reproductibilidad técnica. Ediciones Godot, 2024.
- [3] W. Kandinsky, De la espiritualidad en el arte. Barral Editores, SA, Barcelona, 1982.
- [4] A. Ibarra, Revelan cual fue el presupuesto de Bethesda para el desarrollo de Starfield, 2023. URL: <https://www.egames.news/xbox/Revelan-cual-fue-el-presupuesto-de-Bethesda-para-el-desarrollo-de-Starfield-20230827-0003.html>
- [5] D. Bordwell, and K. Thompson, El arte cinematográfico: una introducción, volumen 68, Paidós comunicación, 1995.
- [6] L. Manovich, The Language of New Media. 2002.
- [7] C. Sánchez Blanco, R. Yáñez González, C. R. Gordon y A. Batuecas Caletrio, Cinetosis o “Motion Sickness”, Sociedad Otorrinolaringológica de Castilla y León, Cantabria y La Rioja, Vol. 5, 2014.
- [8] Steamcharts, estudio comparativo de jugadores entre Starfield y Planet Crafter, histórico, URL: <https://steamcharts.com/cmp/1716740,1284190#All> y en las últimas 48 horas, URL: <https://steamcharts.com/cmp/1716740,1284190#48h>
- [9] Steamdb: web de información sobre cantidad de jugadores en Steam, <https://steamdb.info/>
- [10] Player Counter: web de información sobre cantidad de jugadores en servidores online, <https://playercounter.com/>

A. Recursos online (imágenes)

Figura 1.

The Witcher: Enhanced Edition: <https://acortar.link/lkxf3f>

(<https://steamcommunity.com/app/20900/screenshots/>)

The Witcher 3: Wild Hunt: <https://acortar.link/27MtSn>

(<https://steamcommunity.com/app/292030/screenshots/>)

Figura 2.

Starfield – <https://steamcommunity.com/app/1716740/screenshots/#scrollTop=15204>

Zoom a la imagen: <https://acortar.link/LYCy0j>

Planet Crafter – Zoom a la imagen: <https://acortar.link/pfwR6z>

Explorando el diseño narrativo en videojuegos educativos: el caso Mucho más que abejas.

María del Pilar Bermúdez¹, Roberto Martín²

¹ Universidad de Burgos, Facultad de Humanidades y Comunicación, P.º de los Comendadores, S/N, 09001 Burgos, España. **Email:** mpbermúdez@ubu.es

² Universidad de Burgos, Facultad de Humanidades y Comunicación, P.º de los Comendadores, S/N, 09001 Burgos, España. **Email:** Roberto.martin@ubu.es

Resumen

Este indaga sobre el diseño narrativo en videojuegos educativos mediante el caso Mucho más que abejas, un juego orientado a concienciar al alumnado de primaria sobre la crisis ambiental que afecta a las abejas. El diseño integra mecánicas de juego, narrativa interactiva y semiótica visual para generar una experiencia lúdica y formativa. A través de una combinación de narrativa incrustada y emergente, el juego permite al jugador aprender de forma activa y emocional. Los resultados destacan el potencial de los *serious games* como herramientas didácticas para abordar contenidos complejos desde una perspectiva visual, simbólica e interactiva.

Palabras clave

Narrativa, semiótica, serious games, STEAM

1. Introducción

Este estudio se centra en la exploración de la influencia de las mecánicas de juego y la semiótica en la ludonarrativa en los videojuegos orientados al ámbito didáctico escolar. Se plantea un caso de estudio basado en el videojuego “Mucho más que abejas” que forma parte del proyecto de investigación y transferencia “Prototipos Curriculares Integrados e Inclusivos para propuestas STEAM” financiado por la Agencia Estatal de Investigación.

2. Marco Teórico

El aprendizaje basado en juegos se ha consolidado progresivamente como una disciplina en expansión, caracterizada tanto por la adaptación de enfoques pedagógicos tradicionales al lenguaje y dinámica de los videojuegos [1], como por el diseño de experiencias lúdicas específicamente concebidas con fines didácticos. Esta doble vertiente ha permitido explorar nuevas formas de enseñanza que integran la interacción, la simulación y la narrativa como elementos potenciadores del proceso de aprendizaje. Los avances en la investigación han ido trasladándose al ámbito educativo con un impacto creciente, y la literatura especializada ha documentado ampliamente que los *serious games* pueden mejorar significativamente la motivación, el compromiso y la retención de contenidos [2]. En paralelo, el desarrollo de estos juegos ha dejado de basarse únicamente en la intuición de diseñadores o docentes, dando paso a propuestas metodológicas más estructuradas. Así, diversos autores han planteado marcos y modelos de diseño que buscan integrar desde el inicio principios sólidos tanto de la pedagogía como del diseño de juegos, con el fin de asegurar una mayor coherencia entre objetivos educativos, mecánicas lúdicas y estrategias de evaluación [3]. Asimismo, se han propuesto marcos de análisis para *Serious Games* que destacan la necesidad de coherencia entre mecánicas de juego, objetivos pedagógicos y evaluación [4], lo que ayuda a que el planteamiento de los juegos didácticos sea cada vez más eficaz siempre que los diseñadores sigan estas premisas.

Los videojuegos educativos se consolidan como un formato pedagógico innovador, capaz de motivar al alumnado mediante el aprendizaje interactivo y de contrarrestar el desgaste propio de los modelos educativos tradicionales. No obstante, su diseño enfrenta un doble desafío: garantizar la transmisión efectiva del conocimiento sin sacrificar la dimensión lúdica. Ante la ausencia de estándares en su desarrollo, han surgido múltiples modelos donde frecuentemente predominan los aspectos recreativos sobre los educativos. Existe un desequilibrio persistente entre entretenimiento y aprendizaje, y subrayan la necesidad de centrar el diseño en tres ejes fundamentales: el contenido pedagógico, los objetivos educativos y el proceso de aprendizaje [5].

Aunque la literatura científica ha señalado desde hace tiempo la importancia tanto del juego como de la narración como componentes clave en los procesos cognitivos y educativos, los estudios actuales tienden a analizarlos por separado, estableciendo jerarquías que alternan la primacía de uno u otro. Ante esta fragmentación, se plantea la necesidad de desarrollar métodos y modelos educativos que aprovechen la capacidad innata de la mente humana para integrar la información dentro de un contexto unificado y coherente [6].

3. Metodología

Este estudio se centra en el diseño y desarrollo del videojuego educativo *Mucho más que abejas*, concebido como una herramienta didáctica orientada a sensibilizar al alumnado de educación primaria sobre los factores que afectan negativamente al hábitat y a la dispersión de las abejas. El principal objetivo del proyecto fue crear una experiencia interactiva que, además de transmitir contenidos educativos de forma rigurosa, resultara atractiva y comprensible para el público infantil. Para abordar este reto, se optó por un enfoque de diseño fundamentado en la literatura académica sobre *serious games*, al que se integraron elementos propios de la narrativa interactiva y principios de la semiótica visual, especialmente en lo referente al uso intencionado de la forma y el color como recursos comunicativos y a los principios de la narrativa audiovisual. Desde esta perspectiva, la comunicación audiovisual se revela como una disciplina clave para enriquecer el diseño de videojuegos educativos, al aportar estrategias expresivas y simbólicas que potencian la comprensión y el impacto emocional del mensaje. El desarrollo del juego fue llevado a cabo de manera colaborativa por un equipo multidisciplinar formado por pedagogos, especialistas en comunicación y diseñadores de juegos, quienes trabajaron siguiendo las fases metodológicas recogidas en los principales modelos de diseño revisados en el estado de la cuestión.

4. Resultados

“Mucho más que abejas” es un juego del género *pasillo infinito*, en el que el jugador controla un enjambre de abejas que atraviesa distintos escenarios divididos en casillas, cada una representando territorios con distintos niveles de intervención humana. A través de esta mecánica simple pero eficaz, el juego transmite de forma continua y visual el impacto ambiental, facilitando que los jugadores interioricen conceptos ecológicos clave mientras se enfrentan a retos interactivos.

Parte de la narrativa del juego no se expresa de forma lineal o textual, sino mediante un lenguaje visual simbólico basado en la semiótica del color y la forma. Los paisajes y objetos han sido diseñados bajo una estética *low-poly*, que reduce el nivel de detalle visual para convertir cada elemento en un signo fácilmente reconocible. Esta simplificación permite representar elementos naturales y artificiales con claridad, reforzando el vínculo entre lo que el jugador ve y el significado que se le quiere transmitir. Así, la forma se convierte en un canal de información que guía la experiencia del jugador sin necesidad de instrucciones explícitas.

El color, por su parte, actúa como un código narrativo fundamental: los tonos verdes representan zonas naturales, los marrones indican degradación ambiental y los grises o azules simbolizan la pérdida total de vida. Cada casilla del juego comunica una situación ambiental concreta a través de estos códigos cromáticos, permitiendo que el jugador infiera el mensaje sin necesidad de texto. Este diseño busca aprovechar la capacidad intuitiva del ser humano para organizar y comprender la información visual como un todo coherente, convirtiendo la partida en una experiencia estética y pedagógica simultánea.

La ludonarrativa en *Mucho más que abejas* se construye a partir de la interacción complementaria entre dos formas principales de relato: la narrativa incrustada (embedded narrative) y la narrativa emergente (emergent narrative). Esta distinción, ampliamente abordada en los estudios de narrativa en videojuegos, permite comprender cómo el juego combina elementos predefinidos con experiencias personalizadas para comunicar un mensaje ecológico de forma progresiva.

La narrativa incrustada en “Mucho más que abejas” se materializa a través de una serie de cinemáticas breves que se activan al inicio de cada nivel. En ellas, se introduce un nuevo factor de riesgo para las abejas —como pesticidas, herbicidas, avispa gigantes, automóviles, polución o deforestación—, construyendo una secuencia dramática que estructura la experiencia del jugador. Estas cinemáticas, aunque breves, cumplen una función fundamental: establecen el marco narrativo del nivel y preparan emocional y cognitivamente al jugador para el tipo de amenaza que encontrará en el entorno. A medida que se avanza en el juego, la acumulación de factores de riesgo intensifica la tensión narrativa y refuerza el mensaje central del juego: la vulnerabilidad del ecosistema frente a la acción humana. Esta narrativa guiada permite una comprensión ordenada de los contenidos ecológicos y ayuda a reforzar el carácter progresivo del aprendizaje.

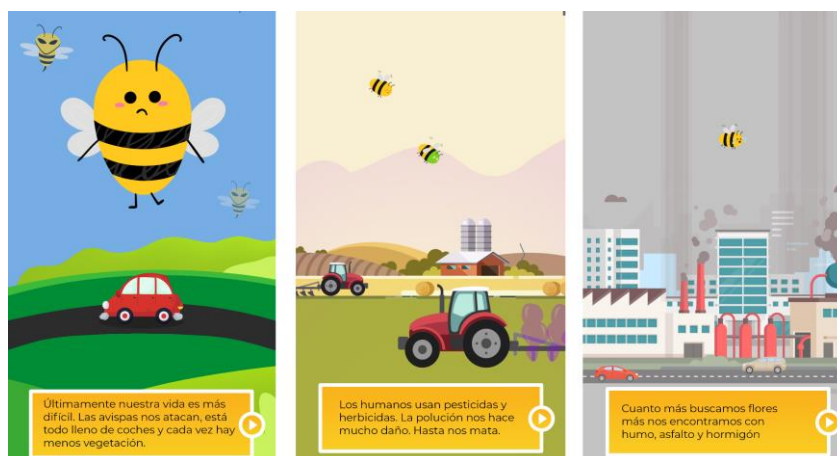


Figura 1. Ejemplos de viñetas de las cinemáticas iniciales de cada nivel.

Paralelamente, la narrativa emergente se desarrolla durante la propia interacción del jugador con el entorno, a través de las mecánicas del juego. A diferencia de la narrativa incrustada, aquí no se trata de recibir una historia, sino de construirla mediante la acción. En este caso, el jugador controla un enjambre de abejas que debe desplazarse por territorios compuestos por casillas que representan distintas formas de intervención humana en el paisaje. Cada casilla es, en sí misma, un signo visual: los colores —verdes para la vegetación, marrones para la degradación, grises y azules para la ausencia de vida— y las formas —simplificadas en estilo low-poly— permiten al jugador decodificar rápidamente la situación ecológica de cada escenario sin necesidad de texto.



Figura 2. Ejemplos de ambientes en cada nivel.

Esta estética, cuidadosamente diseñada, convierte el entorno del juego en una narración visual que evoluciona en función de las decisiones del jugador y del nivel alcanzado.

La clave de esta narrativa emergente reside en que el jugador vive el deterioro ambiental a través del recorrido del enjambre. No se trata solo de una representación simbólica, sino de una experiencia interactiva donde los obstáculos, enemigos y recursos presentes en el entorno reflejan directamente las amenazas introducidas en las cinemáticas previas. Por ejemplo, tras la cinemática que presenta a las avispas gigantes como amenaza, estas aparecen efectivamente en el escenario, modificando la dinámica del juego e intensificando el desafío. De este modo, cada nuevo nivel no solo avanza en dificultad, sino también en densidad narrativa, reforzando el vínculo entre historia y mecánica.

5. Conclusiones

La interacción entre estas dos formas de narrativa propone un diseño orientado a favorecer una experiencia ludonarrativa cohesiva, en la que el aprendizaje no se limita a la recepción pasiva de información, sino que se construye en tiempo real mediante la acción, la exploración y la toma de decisiones. Desde una perspectiva educativa, este diseño potencia una comprensión significativa: el jugador no solo conoce las causas de la desaparición de las abejas, sino que las experimenta simbólicamente. La progresión narrativa, alineada con la progresión mecánica, convierte el videojuego en una herramienta didáctica que apela tanto a la dimensión racional como a la emocional del aprendizaje.

En conclusión, Mucho más que abejas logra articular una ludonarrativa efectiva que combina estructura y libertad, exposición y exploración, guion y juego. Esta estrategia narrativa refuerza el carácter didáctico del videojuego, al permitir que el conocimiento se construya desde múltiples dimensiones —visual, simbólica, interactiva y emocional—, y evidencia el potencial de los serious games para abordar temáticas complejas como la crisis ecológica mediante el lenguaje propio del medio lúdico.

References

- [1] Starks, K. (2023). Guidelines for Ludonarrative Conversion: A Case Study for a Children's Book Converted to a Game. *International Journal of Games and Social Impact*.
- [2] Brandl, L.C., & Schrader, A. (2024). Serious Games in Higher Education in the Transforming Process to Education 4.0—Systematized Review. *Education Sciences*.
- [3] Linehan, C., Kirman, B., Lawson, S., & Chan, G. (2011). Practical, appropriate, empirically-validated guidelines for designing educational games. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1979–1988.

- [4] Mitgutsch, K., & Alvarado, N. (2012). Purposeful by design? A serious game design assessment framework. *FDG '12: Proceedings of the International Conference on the Foundations of Digital Games*, 121–128.
- [5] A. Merino-Cajaraville, S. Reyes-de-Cózar, M. Salguero-Pazos, P. Navazo-Ostúa (2024) TOWARDS THE STANDARDIZATION OF SERIOUS GAMES: PEDAGOGICAL GUIDELINES IN THE DESIGN AND DEVELOPMENT OF EDUCATIONAL VIDEOGAMES, *EDULEARN24 Proceedings*, pp. 4556-4562.
- [6] Antoniazzi, A. (2013). Storytelling and Videogames. An interdisciplinary approach to interactive multimedia models for children. *Studi sulla Formazione*, 16, 67-76.

Serious games y estudios culturales: desarrollo participativo y académico de «Juana I: la puerta de la reina»

Daniel Castilla-Rodríguez¹, M. Pilar Bermúdez-Martínez²

¹Universidad de Burgos, Centro de Innovación y Tecnología en Videojuegos y Comunicación Audiovisual (ITACA), Burgos, España

²Universidad de Burgos, Centro de Innovación y Tecnología en Videojuegos y Comunicación Audiovisual (ITACA), Burgos, España

Resumen

El presente trabajo expone el desarrollo de un videojuego histórico desde el ámbito académico como herramienta experimental en los estudios culturales. Bajo el título «Juana I: la puerta de la reina», el proyecto propone una aproximación innovadora al diseño y análisis de videojuegos como medio de divulgación histórica y reflexión crítica desde una perspectiva de género. A diferencia de estudios que se centran en productos ya comercializados, esta iniciativa se fundamenta en el desarrollo integral de un videojuego en contexto universitario, lo que permite analizar el proceso creativo y sus implicaciones investigativas. El juego se centra en la figura de Juana I de Castilla, empleando una narrativa secundaria protagonizada por una doncella ficticia que sirve como hilo conductor para explorar su entorno histórico. Se combinan metodologías de investigación cualitativas y participativas con enfoques interdisciplinares que integran historia, comunicación audiovisual y diseño interactivo. El proyecto busca romper estereotipos asociados al papel de la mujer en la historia, proponiendo un diseño inclusivo tanto en lo narrativo como en lo estético. El resultado es una experiencia jugable que funciona como laboratorio académico, promoviendo nuevas formas de representación histórica en los videojuegos.

Palabras clave

Serious game, videojuegos históricos, estudios culturales, Juana I de Castilla, estudios de género, juegos educativos, diseño de videojuegos

1. Introducción y contexto académico

El papel de los videojuegos en la construcción de imaginarios históricos ha sido objeto de creciente interés en los estudios culturales y en el ámbito de la investigación interdisciplinar en humanidades digitales. Tradicionalmente, los estudios académicos han centrado su análisis en productos ya publicados, examinando su impacto social, su valor educativo o su representación histórica. Esta aproximación, aunque válida, presenta limitaciones al estar desligada del proceso creativo del videojuego, lo que impide observar cómo se construyen las decisiones narrativas, visuales y mecánicas desde el diseño inicial hasta su ejecución final [1].

El presente trabajo plantea una propuesta alternativa: el desarrollo y análisis conjunto de un videojuego desde el entorno académico. «Juana I: la puerta de la reina» (figure 1) es un proyecto de innovación docente enmarcado en un entorno universitario multidisciplinar que integra historia, comunicación audiovisual, diseño interactivo y estudios culturales. Su objetivo principal es doble: por un lado, experimentar con el diseño de videojuegos como herramienta de investigación sobre procesos históricos y representación de género; por otro, generar un producto jugable que funcione como laboratorio para el estudio del propio proceso de creación [2].

En concreto, el juego se articula en torno a la figura histórica de Juana I de Castilla, un personaje tradicionalmente malinterpretado tanto en la historiografía popular como en las representaciones audiovisuales. A través de la narrativa de una doncella ficticia que forma parte de su séquito, se exploran aspectos políticos, personales y simbólicos de la reina, desde una perspectiva que combina rigor histórico

✉ dcr1010@alu.ubu.es (D. Castilla-Rodríguez); mpbermudez@ubu.es (M. P. Bermúdez-Martínez)

ORCID 0009-0003-2789-565X (D. Castilla-Rodríguez); 0009-0008-7120-1906 (M. P. Bermúdez-Martínez)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



Figure 1: Logo de la portada del videojuego: Juana I La puerta de la reina.

con accesibilidad interactiva. Este enfoque permite establecer un punto de equilibrio entre el respeto por la complejidad del pasado y la necesidad de crear una experiencia interactiva significativa [3][4].

Frente a la opacidad de los procesos de desarrollo en la industria convencional del videojuego, este proyecto propone un marco de trabajo transparente y colaborativo, donde las decisiones narrativas, gráficas y mecánicas son documentadas y analizadas a lo largo de su evolución. Esta estrategia responde a la necesidad creciente de generar casos de estudio que permitan experimentar con nuevas formas de modelización de la historia en videojuegos [5][6].

El contexto académico en el que se inscribe la producción permite romper con los condicionantes comerciales del medio, priorizando objetivos de investigación, experimentación y crítica cultural. En este sentido, el proyecto no solo genera un producto divulgativo, sino que también propone un nuevo modelo metodológico para el diseño de videojuegos históricos desde la universidad.

2. Fundamentación teórica y estado de la cuestión

El videojuego histórico ha emergido como un objeto de análisis legítimo en los estudios culturales y las humanidades digitales. A lo largo de las últimas dos décadas, se ha consolidado un cuerpo teórico que reconoce el potencial del medio para transmitir y reinterpretar narrativas del pasado [5][1]. Más allá de su función como entretenimiento, los videojuegos son hoy considerados espacios de producción simbólica que participan activamente en la construcción de la memoria colectiva [7].

Los estudios sobre videojuegos históricos (*historical game studies*) como disciplina ha permitido cuestionar las representaciones históricas dominantes en los videojuegos comerciales, muchas veces simplificadas, espectacularizadas o tergiversadas [5]. En particular, se ha señalado cómo ciertos títulos reducen episodios complejos a marcos narrativos lineales o centrados exclusivamente en el conflicto bélico, ignorando dimensiones políticas, sociales o de género. Esta crítica cobra especial relevancia cuando se abordan figuras femeninas históricas, cuya representación suele estar mediatizada por estereotipos o ausencias [8].

En este sentido, el personaje de Juana I de Castilla representa un caso paradigmático. Su figura ha sido frecuentemente interpretada desde una mirada patologizante o romántica, asociada a la locura, el amor y la reclusión, invisibilizando su papel político activo en la historia de la monarquía hispánica. Esta construcción ha sido reforzada por narrativas literarias y audiovisuales, y apenas ha sido abordada desde el medio interactivo con un enfoque crítico o desde una perspectiva feminista.

La literatura especializada ha destacado la necesidad de incorporar el análisis de género al diseño de videojuegos históricos, atendiendo tanto a los contenidos como a las formas de interacción, estructura narrativa y diseño de personajes [8][9]. Asimismo, se señala la importancia de cuestionar el modelo dominante de «verosimilitud histórica» que muchas veces se traduce en una aparente neutralidad estética que oculta sesgos ideológicos [6].



Figure 2: Concept art del videojuego mostrando a Aldonza, personaje controlado por el jugador.

Desde una perspectiva metodológica, la creación de videojuegos en contextos académicos contribuye a superar algunas limitaciones de los estudios tradicionales sobre el medio. Tal como señala Rocheleau[2], el desarrollo de modelos tridimensionales con base histórica no solo permite nuevas formas de representación, sino también nuevas formas de validación de hipótesis historiográficas. En este marco, el proyecto «Juana I: la puerta de la reina» no se limita a la producción de un objeto final, sino que documenta y analiza cada fase del diseño como proceso investigativo (figure 2).

Además, experiencias recientes como la creación de videojuegos con base literaria —por ejemplo, la adaptación interactiva de «Don Quijote de la Mancha»[10]— demuestran el potencial del medio para articular narrativas complejas a través de mecánicas jugables que interpelan directamente al jugador desde lo cognitivo y lo emocional.

3. Metodología y proceso de desarrollo

El diseño y desarrollo de «Juana I: la puerta de la reina» se enmarca en un proyecto de innovación docente que emplea el videojuego como herramienta de experimentación en investigación cultural. La metodología utilizada es de carácter cualitativo, participativo e interdisciplinar, articulando distintos niveles de intervención académica con fases de producción iterativa. Este enfoque permite generar un laboratorio vivo de creación donde las decisiones de diseño, guionización, arte y mecánicas no solo responden a criterios técnicos o estéticos, sino también investigativos.

Uno de los principales desafíos del desarrollo de videojuegos en entornos académicos es su discontinuidad. La incorporación de estudiantes en fases específicas del proceso —ligadas a asignaturas, proyectos fin de grado o prácticas extracurriculares— exige que el diseño se mantenga abierto y flexible. Para ello, se adopta un modelo modular que facilita el trabajo autónomo y la integración progresiva de aportaciones dentro de un marco conceptual común, lo que evita que las contribuciones individuales se disocien del conjunto narrativo y funcional del juego.

El proceso de creación parte de una documentación historiográfica rigurosa sobre la vida de Juana I de Castilla, con especial atención a su reclusión en el Palacio de Tordesillas, etapa clave en su vida (figure 3). En lugar de adoptar una estructura narrativa centrada en la figura real, el juego opta por una protagonista ficticia: una doncella de su séquito. Esta decisión responde a la necesidad de evitar interpretaciones especulativas sobre pensamientos o emociones de la reina, permitiendo al jugador explorar el contexto histórico desde una posición de interpretación y no de suplantación [5][1].

Desde el punto de vista técnico, se emplea el formato de aventura gráfica en tercera persona, que favorece el diseño de mecánicas de exploración e interacción con elementos del entorno, reforzando la inmersión en el contexto histórico. Las tareas de la protagonista consisten en cumplir pequeñas misiones o encargos que revelan aspectos del espacio, la época y el personaje de Juana I a través del diálogo, la observación y la toma de decisiones.



Figure 3: Vista desde Blender del modelado de una habitación junto a Aldonza perteneciente al palacio.

Durante el desarrollo, se llevan a cabo sesiones colaborativas con expertos de diversas áreas (historia, narrativa, arte, animación, diseño de niveles, etc.) a través de entrevistas, grupos focales y tutorías dirigidas. Esta dinámica permite integrar múltiples perspectivas académicas en las decisiones de diseño, reforzando el carácter interdisciplinar del proyecto [9] [4].

La iteración y documentación sistemática son componentes clave de esta metodología. Cada fase del desarrollo —desde el diseño conceptual hasta la implementación de mecánicas— es registrada, comentada y analizada desde una óptica crítica. Este material constituye una base para futuras publicaciones, así como para la validación del propio modelo de producción académica de videojuegos.

4. Conclusiones y perspectivas futuras

El proyecto «Juana I: la puerta de la reina» representa una propuesta innovadora en la intersección entre estudios culturales, historia y diseño de videojuegos. Su principal aportación reside en situar el desarrollo de un videojuego no como producto final, sino como proceso investigativo y pedagógico en sí mismo. Al estar concebido en un entorno académico, el juego permite explorar formas alternativas de representar el pasado, desde una perspectiva crítica, feminista y metodológicamente abierta.

Lejos de la lógica de mercado que caracteriza a la industria convencional del videojuego, este enfoque académico fomenta un espacio de experimentación donde la narrativa, la estética y la mecánica están al servicio de objetivos investigativos. En este sentido, se ha evidenciado el potencial del videojuego como herramienta crítica y reflexiva sobre la historia, no solo como vehículo de transmisión de contenidos, sino también como artefacto discursivo que puede cuestionar representaciones hegemónicas del pasado [5][6].

Además, la elección de una figura histórica como Juana I —marcada por múltiples capas de invisibilización, patologización y ficcionalización— aporta un marco fértil para aplicar una mirada crítica desde la perspectiva de género. La decisión de no jugar *como* Juana I, sino desde un punto de vista periférico y contextual, evita caer en especulaciones ficcionales, a la vez que sitúa al jugador en una posición de descubrimiento progresivo, tanto del entorno como de la propia figura histórica [8].

Desde el punto de vista metodológico, el proyecto demuestra la viabilidad del desarrollo modular y colaborativo de videojuegos en entornos universitarios (figure 4). Este modelo permite integrar la investigación con la formación práctica, y abre la puerta a nuevas formas de producción de conocimiento en humanidades digitales. La experiencia adquirida abre la posibilidad de futuras líneas de investigación centradas en la validación del impacto cognitivo y emocional del videojuego, el análisis comparativo de diferentes estructuras narrativas, o la expansión del modelo a otras figuras históricas poco representadas.

Centrados en la proyección futura, se contempla la elaboración de instrumentos de evaluación que permitan analizar la recepción del videojuego en contextos educativos y museísticos, así como su utilidad como herramienta de transferencia cultural. Del mismo modo, se plantea la publicación de estudios paralelos sobre las decisiones de diseño, las tensiones entre rigor histórico y jugabilidad, y las



Figure 4: Vista del centro ITACA, ubicación dónde se está realizando el videojuego de forma interdisciplinar.

estrategias de representación adoptadas.

En suma, este proyecto refuerza la idea de que los videojuegos pueden —y deben— ocupar un lugar central en las agendas de investigación en ciencias sociales y humanidades, no solo como objeto de análisis, sino también como medio expresivo autónomo, capaz de generar nuevas formas de conocimiento histórico y cultural [2][4].

References

- [1] J. Clyde, H. Hopkins, G. Wilkinson, Beyond the “historical” simulation: Using theories of history to inform scholarly game design, *Loading...* 6 (2012).
- [2] M. Rocheleau, La modélisation 3d comme méthode de recherche en sciences historiques, in: *Actes Du 10e Colloque International d’Artefact*, 2011, pp. 245–265.
- [3] A. Chapman, A. Foka, J. Westin, Introduction: what is historical game studies?, *Rethinking History* 21 (2016) 358–371. doi:10.1080/13642529.2016.1256638.
- [4] V. M. Carvalho, Videogames as tools for social science history, *Historian* 79 (2017) 794–819. Crossref.
- [5] A. Chapman, *Digital games as history: How videogames represent the past and offer access to historical practice*, Routledge, 2016. Crossref.
- [6] A. Venegas Ramos, La memoria indolora del videojuego de historia, *Barataria: Revista Castellano-Manchega de Ciencias Sociales* (2020) 24–35. doi:10.20932/barataria.v0i29.545.
- [7] A. Venegas Ramos, El videojuego como forma de memoria estética, *Pasado y Memoria. Revista de Historia Contemporánea* 20 (2020) 277–301. doi:10.14198/PASADO2020.20.12.
- [8] F. Heritage, Magical women: Representations of female characters in the witcher video game series, *Discourse, Context & Media* 49 (2022) 100627. doi:10.1016/j.dcm.2022.100627.
- [9] T. J. Copplestone, Designing and developing a playful past, in: A. Mol, C. Ariesse-Vandemeulebroucke, K. Boom, J. H., A. Politopoulos (Eds.), *The interactive past*, Sidestone Press, 2017, pp. 85–97.
- [10] Desarrollo de un videojuego que introduce en su narrativa el don quijote de la mancha, *Revista Tecnología en Marcha* (2024). doi:10.18845/tm.v37i6.7266.

Diseño de un videojuego en Twine para la exploración de imaginarios utópicos^{*}

Juan Antonio Roa-Domínguez^{1,*}, Juan Antonio Caballero-Hernández²,
Manuel Palomo-Duarte² and Lucía Benítez-Eyzaguirre¹

¹Universidad de Cádiz, Facultad de Ciencias Sociales y de la Comunicación, Av. de la Universidad, s/n, 11405 Jerez de la Frontera, Cádiz, España

²Universidad de Cádiz, Escuela Superior de Ingeniería, Av. de la Universidad 10, 11519 Puerto Real, Cádiz, España

Abstract

El concepto de utopía, entendido como una sociedad futura que supera lo establecido a través de un ideal, ha sido escasamente explorado en los videojuegos, a diferencia de la distopía. Este trabajo presenta el diseño de una narrativa interactiva en Twine que permite al alumnado transitar distintas visiones utópicas a través de decisiones que se registran y analizan mediante minería de procesos. Se ha desarrollado una arquitectura que automatiza la recolección de datos y permite identificar patrones ideológicos y trayectorias de navegación mediante técnicas de minería de procesos. El experimento, aplicado a 23 estudiantes universitarias, revela una preferencia colectiva por modelos socialistas moderados, con énfasis en la igualdad, la colectivización y la sostenibilidad ecológica. La integración de herramientas narrativas y analíticas demuestra el potencial del videojuego como dispositivo pedagógico y crítico. Se concluye que esta metodología permite estudiar cómo el género influye en la construcción de futuros posibles dentro de entornos digitales interactivos.

Keywords

Videojuegos, Utopías, Process mining, Twine, Perspectiva de género

1. Introducción

El videojuego, como producto de la industria cultural, se ha posicionado al servicio de la reproducción de hegemonías ideológicas [1]. Aún así, también se ha reconocido su capacidad para generar espacios simbólicos de resistencia, especialmente cuando se articula en torno a narrativas utópicas que permiten un gameplay ético [2]. En este marco, herramientas como Twine han abierto nuevas posibilidades para la creación de narrativas alternativas, inclusivas y descentralizadas [3].

Así, estudiar el videojuego desde una perspectiva utópica, centrada en su potencial frente al realismo capitalista, se vuelve una tarea necesaria para desentrañar las complejas relaciones entre videojuegos, capitalismo y poder [4]. De igual importancia se presenta la perspectiva de género en el enfoque de la investigación. En un sector dominado por la cultura brogrammer, la perspectiva de género ofrece una mirada crítica que revela cómo los roles de género influyen en nuestra visión del futuro [5].

Este trabajo presenta la arquitectura de un videojuego narrativo elaborado en el software libre especializado en videojuegos de narrativas no lineales: Twine¹, donde el jugador transita por distintas concepciones de lo utópico según sus decisiones y valores, que serán recogidas y procesadas mediante técnicas de minería de procesos. Todo ello con el fin de comprobar cómo las audiencias –en función de su género– dialogan con el concepto de utopía a través del videojuego como artefacto mediador, siendo la minería de procesos necesaria para detectar patrones de comportamiento colectivos.

IV Congreso Español de Videojuegos: Granada, 18 y 19 de septiembre de 2025

*Corresponding author.

✉ juan.roadominguez@alum.uca.es (J. A. Roa-Domínguez); juanantonio.caballero@uca.es (J. A. Caballero-Hernández); manuel.palomo@uca.es (M. Palomo-Duarte); lucia.benitez@uca.es (L. Benítez-Eyzaguirre)

🌐 <https://www.linkedin.com/in/juan-antonio-roa-374554172/> (J. A. Roa-Domínguez);

<https://produccioncientifica.uca.es/investigadores/205633/detalle> (J. A. Caballero-Hernández);

<https://indess.uca.es/manuel-palomo-duarte> (M. Palomo-Duarte); <https://luciabenitez.es/> (L. Benítez-Eyzaguirre)

🆔 0000-0002-5453-6180 (J. A. Roa-Domínguez); 0000-0001-9012-8961 (J. A. Caballero-Hernández); 000-0001-6002-0319

(M. Palomo-Duarte); 0000-0002-8879-7979 (L. Benítez-Eyzaguirre)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹<https://twinery.org/>

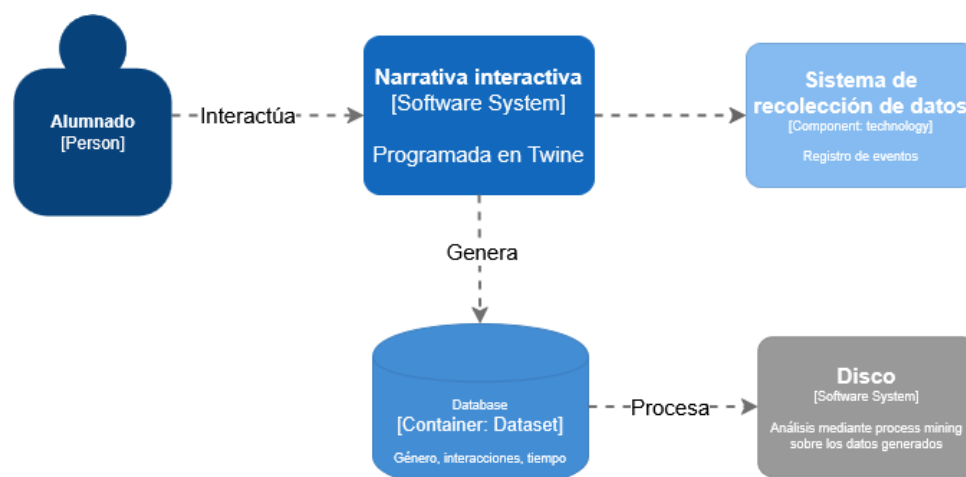


Figure 1: Arquitectura del videojuego según modelo C4.

2. Trabajos relacionados

Nuestra investigación integra dos enfoques habitualmente separados: el análisis del comportamiento en entornos interactivos mediante process mining y la creación de narrativas críticas con Twine. La minería de procesos (process mining en inglés) pretende el descubrimiento, monitorización y mejora de procesos a través del registro de eventos [6]. Así, se ha aplicado tanto para evaluar competencias académicas mediante serious games [7] como para ajustar dinámicamente la dificultad del juego según las capacidades de la persona usuaria [8].

Esta propuesta une ambas perspectivas, combinando análisis automatizado y narrativa utópica interactiva.

Este estudio aporta al estado del arte la combinación de narrativas utópicas en Twine con análisis automatizado mediante minería de procesos. Esta integración permite interpretar las decisiones del alumnado tanto como expresión crítica, como una perspectiva computacional, uniendo ambos campos.

3. Diseño del videojuego y conjunto de datos

En este apartado se presenta el diseño narrativo del videojuego y la arquitectura desarrollada para registrar y analizar las interacciones. Asimismo, se describe el conjunto de datos recogido durante la experiencia de juego.

El videojuego ha sido desarrollado en Twine por su carácter libre, accesible y orientado a narrativas no lineales, y propone cinco rutas principales inspiradas en diferentes concepciones utópicas (platónica, renacentista, ilustrada, socialista y contemporánea) según [9]. A lo largo del recorrido, la persona jugadora puede transitar entre estas rutas en momentos clave, respondiendo a preguntas interactivas que permiten evaluar su afinidad ideológica con los distintos modelos. La interfaz combina narrativa, visuales y decisiones múltiples para fomentar una participación activa, generando así una experiencia crítica.

Para ello, se diseñó una arquitectura de software capaz de registrar, almacenar y analizar automáticamente las decisiones, tiempos y trayectorias del alumnado durante el juego. Estos datos, estructurados en una base, se procesan mediante minería de procesos con la herramienta DISCO², lo que permite reconstruir patrones de navegación y comprender cómo el alumnado interpreta y reproduce modelos utópicos en el entorno interactivo.

En total, se recogieron las interacciones de juego ³ a 23 alumnas de la Universidad de Cádiz, en los grados de Publicidad y Relaciones Públicas y Marketing e Investigación de Mercados. Los datos

²<https://fluxicon.com/disco/>

³Grafo completo disponible en: 10.6084/m9.figshare.29381252

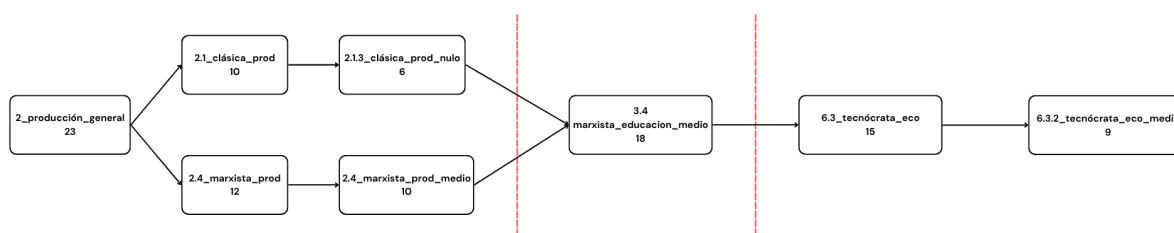


Figure 2: Extracto flujo de interacción.

recogidos comprenden: (I) género de la persona, (II) interacciones por las distintas rutas de juego, (III) tiempo en cada ruta, (IV) tiempo total de la experiencia de juego. Los datos fueron recogidos en formato JSON a través de un script (ver figura 1).

4. Resultados

En la figura 2 se presenta un extracto del recorrido realizado en el videojuego por un grupo de 23 chicas, de entre 19 y 25 años, seleccionadas aleatoriamente entre las alumnas de las asignaturas de Marketing y Cuentas (Grado en Publicidad y Relaciones Públicas) y Herramientas Informáticas para la Publicación en Web (Grado en Marketing e Investigación de Mercados). El análisis del grafo generado revela un perfil colectivo claramente orientado hacia valores socialistas moderados, con fuerte énfasis en la igualdad social, la educación pública y una economía colectivizada. Esta tendencia se traduce en una alta frecuencia de paso por nodos como “3.4.2 marxista-educación-media”, que registran los mayores grosores de flujo, indicando que la mayoría de las participantes optaron por estas alternativas. En al menos cuatro de los siete niveles del recorrido, los caminos más transitados corresponden a ramas socialistas en sus versiones intermedias, lo que denota una preferencia por estructuras colectivas pero equilibradas, alejadas tanto del capitalismo como de posturas autoritarias. En la dimensión ecológica (nivel 6), destaca la elección combinada en “6.3.2 tecnóeco-eco-medio”, lo que apunta a una voluntad de integrar desarrollo tecnológico con sostenibilidad ecológica, articulando así una visión de futuro tecnificada y ambientalmente consciente. Asimismo, se evidencia una ausencia casi total de opciones extremas —tanto en marcos clásicos como socialistas— y una escasa representación de rutas liberales o neoliberales, prácticamente marginales en el grafo, lo cual refuerza la hipótesis de una utopía moderada, pragmática y plural, sostenida en una preferencia ideológica clara y común entre las participantes.

5. Amenazas a la validez

La muestra, compuesta por 23 estudiantes universitarias, es reducida y homogénea, lo que limita la generalización de los resultados. Además, las decisiones tomadas en el entorno del juego pueden no reflejar de forma directa convicciones ideológicas estables, afectando la validez del artefacto. De igual forma, el carácter simulado del videojuego implica una baja validez de concordancia con entornos no simulados, ya que las elecciones en contextos lúdicos no siempre se trasladan al comportamiento real.

6. Conclusiones y trabajo futuro

Este estudio ha demostrado que es posible integrar herramientas de creación narrativa como Twine con metodologías de análisis automatizado como la minería de procesos, generando así un entorno híbrido que permite tanto la exploración ideológica del alumnado como la recolección sistemática de datos interpretativos. El análisis de las decisiones tomadas revela una tendencia hacia visiones utópicas propias de la tradición marxista más moderadas, con especial énfasis en la equidad social, la educación pública y una economía colectivizada y tecnificada. Además, el enfoque del juego ha propiciado una

aproximación crítica que visibiliza la relación entre género y utopía, en un contexto aún dominado por la cultura tecnológica masculina.

Como trabajo futuro, se propone ampliar el número y diversidad de participantes para validar la generalización de los patrones encontrados. También se plantea la posibilidad de desarrollar nuevos módulos narrativos que exploren utopías desde otras perspectivas interseccionales, fortaleciendo el potencial inclusivo y experimental de esta herramienta.

Agradecimientos

Esta publicación es parte del proyecto PID2023-149674OB-I00, financiado por MICI-U/AEI/10.13039/501100011033 y por ERDF, EU.

El software usado en este estudio cuenta con gráficos realizados por Raúl Moreno Gómez bajo licencia CC-by. La autoría agradece su aportación.

References

- [1] J. Dean, Communicative capitalism: Circulation and the foreclosure of politics, *Cultural politics* 1 (2005) 51–74.
- [2] M. Sicart, Mundos y sistemas: entendiendo el diseño de la gameplay ética, *Comunicación: revista Internacional de Comunicación Audiovisual, Publicidad y Estudios Culturales*, 1 (7), 45-61. (2009).
- [3] A. Harvey, Twine's revolution: Democratization, depoliticization, and the queering of game design, *GAME: The Italian Journal of Game Studies* (2014) 95–107. URL: http://www.gamejournal.it/3_harvey/.
- [4] M. Fisher, Realismo capitalista ¿no hay alternativa?, *Caja Negra* (2023).
- [5] L. Benítez-Eyzaguirre, La cultura programador y el sexismo digital, in: J. Alberich, F. Sierra (Eds.), *Cultura digital, nuevas mediaciones sociales e identidades culturales*, Comunicación Social Ediciones y Publicaciones, 2021, pp. 71–96.
- [6] W. M. Van der Aalst, M. H. Schonenberg, M. Song, Time prediction based on process mining, *Information systems* 36 (2011) 450–475. doi:10.1016/j.is.2010.09.001.
- [7] J. A. Caballero Hernández, M. Palomo Duarte, J. M. Dodero Beardo, D. Gašević, Supporting skill assessment in learning experiences based on serious games through process mining techniques, *International Journal of Interactive Multimedia and Artificial Intelligence* 8 (2023) 146–159. doi:10.9781/ijimai.2023.05.002.
- [8] D. S. Lora Ariza, Data mining techniques in video games, Master's thesis, Universidad Complutense de Madrid, 2015.
- [9] R. Levitas, *The concept of utopia*, volume 3, Peter Lang, 2010.

Integración del Patrimonio Cultural en el Diseño de Videojuegos Narrativos

Iñigo Gómez¹, Daniel Castilla², Roberto Martín³

¹ Universidad de Burgos, Facultad de Humanidades y Comunicación, P.º de los Comendadores, S/N, 09001 Burgos, España. **Email:** igz1001@alu.ubu.es

² Universidad de Burgos, Facultad de Humanidades y Comunicación, P.º de los Comendadores, S/N, 09001 Burgos, España. **Email:** dcr1010@alu.ubu.es

³ Universidad de Burgos, Facultad de Humanidades y Comunicación, P.º de los Comendadores, S/N, 09001 Burgos, España. **Email:** roberto.martin@ubu.es

Resumen

El presente estudio analiza el desarrollo de un videojuego narrativo en primera persona ambientado en el pueblo despoblado de Sarnago (Soria, España), concebido como herramienta para la sensibilización y difusión del patrimonio cultural. Partiendo del creciente interés académico por los videojuegos como medios para acercar a los jugadores a valores culturales, esta investigación aborda los retos de representar el patrimonio de manera precisa y significativa dentro del entorno lúdico. El proyecto plantea una propuesta alternativa a los enfoques comerciales, al poner al jugador en un rol activo en la reconstrucción simbólica del patrimonio, mediante mecánicas centradas en la exploración y la interpretación. Aunque se encuentra aún en fase experimental y sin pruebas empíricas con usuarios, los resultados preliminares sugieren el potencial del videojuego como medio para fomentar una relación afectiva y reflexiva con la memoria colectiva.

Palabras clave

Videojuegos culturales, patrimonio cultural, memoria colectiva, diseño narrativo

1. Introducción

El uso de videojuegos como herramienta para difundir el patrimonio cultural ha despertado interés académico debido a su potencial para acercar a los jugadores a distintos aspectos culturales (Camuñas-García et al., 2024). Sin embargo, la inclusión de estos elementos presenta desafíos en cuanto a precisión y transmisión adecuada de valores culturales. Muchos juegos incorporan escenarios y personajes históricos para ambientar sus narrativas, pero estas representaciones suelen priorizar el entretenimiento sobre la enseñanza o preservación del patrimonio (Hanes & Stone, 2019). Así, el patrimonio a menudo funciona solo como un telón de fondo, relegado frente a la jugabilidad (García-Fernández & Medeiros, 2019).

Además, el fenómeno de la reutilización del patrimonio implica que elementos históricos son recontextualizados en nuevos marcos, lo que puede acercar a los jugadores a aspectos culturales desconocidos, pero también puede distorsionar su significado y valor cultural (Guglielmucci & Scaraffuni Ribeiro, 2016; Eyüce & Eyüce, 2010). Por otro lado, los videojuegos pueden generar una relación afectiva y participativa entre el jugador y el patrimonio (Flynn, 2005), aunque esta interacción suele ser limitada, con pocas oportunidades para recrear prácticas culturales, que generalmente quedan en manos de personajes no jugables (Barbara, 2020). Por ello, para evitar la modificación de la memoria colectiva, hay que estudiar detenidamente de qué manera interactúan los jugadores con los elementos patrimoniales.

En este contexto, el presente trabajo explora el desarrollo de Calera, un videojuego narrativo en primera persona ambientado en el pueblo despoblado de Sarnago (Soria, España). El proyecto se concibió como una herramienta de sensibilización patrimonial, con el objetivo de representar los elementos tangibles e intangibles de la cultura local. Esta investigación analiza cómo se

integran las dimensiones del patrimonio en el diseño del juego y qué potencial tiene este tipo de experiencias interactivas para la conservación de la memoria colectiva.

2. Marco Teórico

Para lograr una integración efectiva entre la educación patrimonial y el diseño de videojuegos, era necesario desarrollar un sistema capaz de medir y evaluar cómo se representan las diversas dimensiones del patrimonio cultural dentro de estas experiencias interactivas. Un sistema métrico de este tipo puede servir como guía para los desarrolladores, facilitando no solo la incorporación significativa de elementos patrimoniales en los espacios de juego, sino también asegurando que dichos elementos se presenten de manera coherente, educativa y culturalmente respetuosa.

Con este objetivo, los investigadores Noreena Yi-Chin Liu, Kevin Chi-Jen Miao y Darren Chen Yi Ying (2022), diseñaron una herramienta métrica basada en un enfoque estructurado que permite analizar cómo se representa el patrimonio, tanto tangible como intangible, en los distintos componentes del videojuego. Esta herramienta se construye a partir de las “palabras enlace” propuestas por Harvey (2001): exhibir, experimentar, representar y registrar. Estas palabras ayudan a entender de qué manera actúa cada elemento del patrimonio. Las palabras concuerdan con el propósito de los museos y divide el patrimonio en arte y artefactos, entorno, personas e historia:

1. **Arte y artefactos:** Objetos creados por el ser humano que, dentro del juego, funcionan como elementos visuales o interactivos que conectan al jugador con el contexto cultural e histórico. En los museos, los artefactos son *exhibidos* para captar la atención de los visitantes (Petrelli et al., 2017). De la misma manera, estos elementos pueden ser representados dentro de un videojuego en forma de objetos y escenarios (Liu et al., 2022).
2. **Entorno:** Paisajes y espacios culturales recreados digitalmente. Trasladar la estructura de un lugar real a un entorno de juego crea un escenario inmersivo en el que se puede *experimentar* la actividad cultural local. El usuario puede explorar el espacio de juego e investigar lo que podría haber sucedido y cómo podría haber vivido la gente según la manera en la que interactúa con el entorno.
3. **Personas:** *Representación* de individuos o comunidades a través de sus costumbres, rituales, lenguaje o creencias, abordando especialmente el patrimonio inmaterial.
4. **Historia:** Forma en la que un juego *registra* la historia. La historia del patrimonio cultural puede estar relacionada con la narrativa, el guion, el diálogo o el texto del juego.

Cada una de estas dimensiones puede integrarse en diferentes componentes del videojuego, como:

- **Jugabilidad:** Interacción del jugador con los elementos patrimoniales mediante mecánicas de juego (retos, exploración, uso de herramientas, etc.).
- **Narrativa:** Integración de historias o leyendas relacionadas con el patrimonio cultural para motivar y contextualizar la experiencia.
- **Actuación/Interpretación:** Representación de movimientos, expresiones o comportamientos que encarnan rituales, danzas u otras prácticas culturales.
- **Arte/Diseño:** Estética visual que representa el patrimonio a través de gráficos, diseño de escenarios, personajes e interfaz.
- **Música/Sonido:** Uso de elementos sonoros para ambientar y reforzar la conexión con el patrimonio cultural (música tradicional, efectos ambientales, voces, etc.).

A partir de la combinación entre las dimensiones del patrimonio y los componentes del videojuego, Liu et al. (2022) construyen una tabla métrica, visible en la [Tabla 1](#), que permite visualizar y evaluar de forma clara cómo se integran los distintos elementos patrimoniales en el diseño de los juegos.

Tabla 1

Relación de las dimensiones del patrimonio con cada componente del videojuego.

Fuente: “A Metric Instrument for the Games with Cultural Heritage”.

	Jugabilidad	Narrativa	Actuación	Arte/Diseño	Sonido
Arte y artefactos	El juego permite interactuar con el arte, la arquitectura, etc.	La narrativa está relacionada con la arquitectura o las obras de arte.	La actuación interpreta el arte, la arquitectura, etc.	El diseño del juego muestra la arquitectura o las obras de arte de un lugar.	La música y el sonido del juego se inspiran en la arquitectura, escultura, folclore, etc.
Entorno	El juego permite interactuar con entornos de interés patrimonial.	La narrativa está relacionada con lugares naturales de interés patrimonial.	La actuación tiene lugar en entornos de interés patrimonial	El diseño del juego muestra lugares naturales de interés patrimonial.	La música y el sonido del juego se inspiran en lugares de interés paisajístico, en el clima, etc.
Personas	El juego permite interactuar con la gente de un lugar.	La narrativa del juego está relacionada con la gente de un lugar y como se comportan.	La actuación retrata la gente de un lugar y su comportamiento.	El diseño del juego muestra la gente de un lugar, sus ropas, su manera de actuar, etc.	La música y el sonido del juego se inspiran en la gente de las tribus y aldeas, en su comportamiento, etc.
Historia	El juego permite interactuar con lugares o personajes históricos	La narrativa del juego está relacionada con un hecho, personaje o época historia.	La actuación retrata una época o lugar histórico.	El diseño del juego muestra lugares o personajes históricos.	La música y el sonido del juego se inspiran en lugares o personajes históricos, en costumbres, en épocas históricas, etc.

3. Metodología

La metodología aplicada para el diseño de Calera se basó en la herramienta métrica propuesta por Liu et al. (2022), adaptado para el patrimonio cultural específico de Sarnago. El objetivo principal fue traducir las dimensiones del patrimonio identificadas en el estado de la cuestión en elementos concretos y funcionales dentro del videojuego.

Este proceso comenzó con la elaboración de una tabla que cruzaba las dimensiones del patrimonio cultural con los componentes clave del videojuego: jugabilidad, narrativa, actuación/interpretación, arte/diseño y música/sonido. A partir de las “palabras enlace” de Harvey (2001) — exhibir, experimentar, representar y registrar — se formularon preguntas específicas para cada combinación, con el fin de definir cómo cada elemento patrimonial podía integrarse efectivamente en el diseño.

Por ejemplo, para la dimensión Arte y artefactos, se plantearon interrogantes sobre cómo esos objetos podían ser exhibidos en el entorno visual del juego, cómo el jugador podría interactuar con ellos, qué papel narrativo podrían desempeñar y qué tipo de sonidos podrían asociarse para reforzar su presencia cultural.

Este ejercicio permitió identificar qué elementos patrimoniales resultaban más adecuados para su inclusión, facilitando decisiones informadas para el desarrollo de mecánicas, construcción narrativa, ambientación visual y sonora. Así, la tabla funcionó como una herramienta práctica para guiar el diseño integral de Calera.

Debido a la extensión de la tabla completa, a continuación, en la [Tabla 2](#), se presenta un ejemplo focalizado en la dimensión de Arte y artefactos ilustrando cómo se aplicó esta metodología para vincular el patrimonio tangible con los distintos componentes del videojuego.

Tabla 2

Tabla que se utilizó antes de comenzar a desarrollar Calera con el objetivo de detectar de qué forma se podía trasladar el patrimonio de Sarnago al videojuego.

Fuente: Elaboración propia.

OBJETIVOS		PREGUNTAS	REPRESENTACIÓN
Arte y artefactos: Evaluar cómo el juego del patrimonio cultural puede exhibir el arte y los artefactos.	Jugabilidad	¿Cómo interactúa el juego con el arte y los objetos?	El jugador recoge materiales como piedra caliza, madera o herramientas tradicionales para reconstruir el horno de cal. Esta interacción se da al explorar espacios concretos del pueblo.
	Narrativa	¿Cómo se utiliza la narrativa del juego para relacionarlo con el arte y los objetos del patrimonio?	Aunque <i>Calera</i> no incluye una narrativa textual extensa, la disposición y el estado de los objetos (por ejemplo, un horno en ruinas o casas abandonadas) cuentan visualmente la historia del abandono del pueblo y su tradición artesanal.
	Actuación	¿Cómo interpreta el juego el arte y los objetos patrimoniales?	El jugador actúa simbólicamente como el último habitante que, al reunir materiales y reconstruir el horno, revive y reinterpreta un antiguo oficio local que formaba parte de la identidad del lugar.
	Arte/Diseño	¿Cómo muestra el juego el arte y los artefactos?	El diseño visual <i>low poly</i> recrea elementos patrimoniales como el horno de cal, el lavadero, la iglesia o las casas con formas geométricas simplificadas que conservan los rasgos clave del entorno original. Se prioriza la evocación estética sobre la fidelidad realista.
	Música/Sonido	¿Cómo se inspira la música y el sonido del juego en el arte y los objetos?	Cada espacio asociado a un objeto patrimonial activa un recuerdo sonoro: el horno incluye sonidos de combustión y trabajo; el lavadero, agua y voces femeninas; la iglesia, repiques de campana y cantos litúrgicos. Estos audios ambientan e interpretan el uso original de cada objeto.

4. Resultados

El desarrollo del videojuego Calera se encuentra en una versión casi final, con una narrativa y un conjunto de mecánicas diseñadas para explorar y transmitir el patrimonio cultural del pueblo de Sarnago. El juego propone una experiencia narrativa en primera persona centrada en la recuperación simbólica de un espacio rural despoblado, donde el jugador asume el papel del último habitante cuyo propósito es reconstruir el horno de cal, símbolo de identidad local.



Figura 1: Horno de cal reconstruido Fuente: Elaboración propia.

Durante la partida, el jugador debe recolectar materiales dispersos por el pueblo para reparar el horno. Esta mecánica de exploración actúa como hilo conductor para recorrer espacios claves del entorno, donde cada lugar activa recuerdos sonoros que representan escenas de la vida cotidiana y prácticas culturales del pasado. Por ejemplo, al acercarse al lavadero, se escuchan conversaciones entre mujeres y el fluir del agua; en la iglesia, las campanas y los cánticos litúrgicos representan un día de misa; y en la plaza, los sonidos vecinales reconstruyen el ambiente social de la comunidad. Estos elementos no solo aportan una capa de inmersión sensorial, sino que funcionan como vehículos para la transmisión del patrimonio inmaterial.



Figura 2: Material que el jugador debe recolectar para reparar el horno Fuente: Elaboración propia.

A nivel técnico y visual, el entorno del pueblo ha sido modelado en 3D con una estética low poly, que apuesta por formas simplificadas y estilizadas para evocar el paisaje rural sin recurrir al realismo. La narrativa no se desarrolla mediante textos extensos ni diálogos, sino que se transmite de forma ambiental a través de sonidos, objetos y la disposición del entorno. La jugabilidad se basa en mecánicas no competitivas, diseñadas para favorecer una experiencia pausada, contemplativa y centrada en la exploración sensorial del patrimonio.



Figura 3: Imágen del entorno en el que se desarrolla la historia del juego Fuente: Elaboración propia.

El diseño general prioriza la coherencia entre experiencia lúdica y contenido cultural, evitando relegar el patrimonio a un simple decorado. En lugar de ello, cada elemento del juego (visual, sonoro, narrativo y mecánico) contribuye a reforzar la conexión afectiva del jugador con la historia del lugar.

5. Conclusiones

El desarrollo del videojuego *Calera* ha servido como experimento para explorar cómo integrar el patrimonio cultural en el diseño de videojuegos más allá de su uso decorativo habitual. Aunque se ha empleado una herramienta métrica (Liu, Miao y Ying, 2022) para estructurar esta integración, el proyecto aún está en fase experimental y carece de pruebas con usuarios. La evaluación futura, mediante estudios empíricos y publicación del juego, será clave para medir si la experiencia realmente promueve una conexión significativa con la memoria colectiva o si los elementos patrimoniales siguen siendo percibidos como secundarios.

6. Referencias

Barbara, J. (2020). CLASSIFICATION OF GAMEPLAY INTERACTION IN DIGITAL CULTURAL HERITAGE

Camuñas-García, D., Cáceres-Reche, M. P., Cambil-Hernández, M. d. l. E., & Lorenzo-Martín, M. E. (2024). Digital Game-Based Heritage Education: Analyzing the potential of heritage-based video games. *Education Sciences*, 14(4), 396. <https://doi.org/10.3390/educsci14040396>

Eyüce, Ö., & Eyüce, A. (2010). Design education for adaptive reuse. *Archnet-IJAR: International Journal of Architectural Research*, 4(2/3), 419–428.

Flynn, B. (2005). Gameplay as Interpretative Cultural Heritage. In *Journal Title IADIS International Journal on WWW/Internet*.

Guglielmucci, A., & Scaraffuni Ribeiro, L. (2016). Site of memory and site of forgetting: The repurposing of the Punta Carretas prison. *Latin American Perspectives*, 43(5), 131–144. <https://doi.org/10.1177/0094582X15570882> ri.conicet.gov.ar+8

García-Fernández, J., & Medeiros, L. (2019). Cultural heritage and communication through simulation videogames—A validation of Minecraft. *Heritage*, 2(3), 2262–2274. <https://doi.org/10.3390/heritage2030138>

Hanes, L., Stone, R. A model of heritage content to support the design and analysis of video games for history education. *J. Comput. Educ.* 6, 587–612 (2019). <https://doi.org/10.1007/s40692-018-0120-2>

Harvey, M. L., Loomis, R. J., Bell, P. A., & Marino, M. (1998). The influence of museum exhibit design on immersion and psychological flow. *Environment and Behavior*, 30(5), 601–627

Liu, N. Y. C., Miao, K. C. J., & Ying, D. C. Y. (2022). A Metric Instrument for the Games with Cultural Heritage. *International Journal of Serious Games*, 9(4), 89–136

Petrelli, D., Marshall, M. T., O'Brien, S., McEntaggart, P., & Gwilt, I. (2017). Tangible data souvenirs as a bridge between a physical museum visit and online digital experience. *Personal and Ubiquitous Computing*, 21(2), 281–295

Playing to Raise Awareness: A Narrative Video Game Against Meningitis

Laplana Curtido Oriana¹

¹ T¹ Spanish Meningitis Association (AEM), Volunteer Team, Madrid, Spain

Abstract

This paper presents the development of a serious video game designed to raise awareness about meningitis, particularly among adolescents, a high-risk and often underinformed group. The game is free, brief, and accessible via both mobile devices and web browsers. It uses emotional, interactive storytelling as its core strategy to communicate critical health information while maintaining player engagement. By placing players in a high-stakes narrative situation, the game fosters empathy and emotional immersion, aiming to create a lasting impact beyond traditional informational approaches. Key design decisions, including narrative structure, voice acting, and gender-neutral character design, are examined, along with the challenges of balancing education and entertainment. The collaboration with a health NGO ensured scientific accuracy and outreach support. The project is currently in the data collection phase to evaluate its effectiveness as a tool for health awareness.

The game is freely accessible and playable on mobile and desktop browsers at [24h: Hope's Machine Redirect Link](#).

Keywords

serious game, meningitis, interactive narrative, emotional storytelling, health, awareness

1. Introduction

Meningitis is a serious and often little-known disease that can cause permanent damage or death if not treated promptly. Its impact is especially concerning among adolescents and young adults, a high-risk group that remains poorly informed. According to the Spanish Association Against Meningitis (AEM), adolescents are one of the populations most vulnerable to meningococcal infection due to low vaccination rates and increased exposure in social environments such as schools, parties, or shared housing situations [1].

This work arises from a direct personal experience with the disease, which motivated the creation of a serious game aimed at helping to inform and raise awareness among those who need it most. By addressing a real risk in an accessible, emotional, and engaging way, the project seeks to transform awareness into preventive action.

2. Objective and Approach

The objective was to design an experience that not only informed but did so in an emotional, accessible, and engaging way for a young audience. As a serious game, it had to balance two goals: teaching rigorously and maintaining the player's interest. The approach was to create an immersive and brief narrative that did not require complex skills, ensuring that anyone, including casual players, could complete it and take away a clear message.

Research shows that immersive narratives combined with meaningful player agency significantly enhance engagement and information retention among adolescents, making this approach particularly effective for health awareness games (Vandersmissen et al., 2021)[2].

We also drew on current public health recommendations, AEM emphasizes that vaccination in adolescents and young adults is essential to prevent the second incidence peak, where unvaccinated groups living in close-contact environments remain vulnerable.

The central design philosophy was to make players feel the urgency and consequences of meningitis, not just learn facts about it. Emotional immersion was treated as a primary vehicle for awareness.

3. Game Description

The game, 24h: The Hope Machine (original title in Spanish: La máquina de la esperanza), is designed as a point-and-click graphic novel focused on exploration and interaction with objects and characters within the environment. The narrative dynamically adapts based on the order of actions and exploration choices made throughout the game. This approach encourages players to discover information about meningitis naturally through subtle interactions.

The player assumes the role of a parent whose child, Luis, has been urgently admitted to the hospital. After diagnostic tests, doctors confirm a severe case of meningitis. In desperation, the protagonist discovers a time machine in the hospital basement and embarks on an impossible mission: traveling to the past to vaccinate their son. Through this journey, the player learns about symptoms, consequences, and prevention of meningitis.

A clear and impactful ending delivers a strong message once the player discovers the machine does not work: time machines do not exist, meningitis is not a game — vaccinate yourself.

3.1. First-Person Narrative and Player Identification

The story is presented in the first person, deliberately avoiding any mention of the player's gender, allowing players to fully identify with the main character regardless of their gender. Playtesting revealed that players naturally project their own parental role (mother or father) without explicit prompts, which enhances immersion and engagement.

3.2. Branching Paths and Player Agency

Multiple narrative paths were developed based on player actions, including the order of object interactions and exploration. This branching structure enables the story to respond dynamically to player behavior, providing a more personalized and immersive experience.

A key design decision was to have each scene object independently check relevant game states and player knowledge at any given moment. This required maintaining and updating a record of player states or acquired information, ensuring the player would not repeatedly receive the same details. Consequently, information could be discovered through various objects or characters, who would adapt their responses depending on the player's progression. This system allowed for narrative acceleration or the provision of hints based on how far the player had advanced, creating a richer, more responsive storytelling environment.

Another significant challenge was ensuring that the conclusion of certain narrative paths did not break immersion. Some branches end with diegetic events, such as the player being taken to another room, or a character with whom the player has been speaking falling asleep after their dialogue tree is fully explored, which serve as natural transitions while subtly closing off player agency. This aligns with the narrative philosophy expressed by Jon Ingold in his GDC talk *Narrative Sorcery* (2017) [3], where he states: “My golden rule has always been: if you can't interact with it, it doesn't exist.” His approach highlights the importance of justifying every limitation through the narrative itself, avoiding repeated text or mechanical barriers that might remind players they are in a game. In line with this, all transitions and constraints in the game were deliberately embedded into the story world to preserve internal coherence and reinforce emotional immersion.

3.3. Audio and Technical Considerations

The game features professional Spanish voice acting to enhance narrative quality, auditory accessibility, and emotional engagement. High-quality voice acting not only improves the aesthetic

experience but also strengthens player immersion and attention. Research by Brockmyer et al. (2015) [4] indicates that character voiceovers significantly increase learner engagement in digital game-based environments compared to silent or text-only formats, supporting the use of voice acting as a powerful tool to maintain player focus and emotional connection.

4. Initial Post-Mortem

4.1. Successes

- Collaborating with the Spanish Association Against Meningitis (AEM) facilitated the scientific validation of the content and boosted the game's dissemination.
- Voice acting increased immersion and accessibility within the Spanish-speaking audience.
- The player's gender neutrality reinforced emotional engagement and removed barriers to access.
- The narrative adapts dynamically to the player's exploration path and action order, maintaining consistency to preserve immersion without altering the core storyline.

4.2. Challenges

- Although emotionally effective, the narrative may not appeal to players who prefer more active gameplay.
- Voice acting limits the ability to iterate on narrative design and complicates international localization.
- Maintaining player neutrality sacrificed some accessibility for visually impaired users (there is no protagonist voice).
- Communicating health information without disrupting gameplay flow was challenging. The “pamphlet effect” was avoided by reducing the amount of explicit data and integrating links to external resources such as the Spanish Association Against Meningitis.

5. Evaluation Methodology

Currently, the game is freely available and accessible via web browser and mobile apps. Throughout 2025, an evaluation plan will be implemented based on two pillars:

1. Usage analysis (tracking): dropout rates, completion rates, and points of highest retention or loss of interest.
2. Educational talks: the game will be used as an introduction to sessions in schools or community centers, where questionnaires will be distributed before and after playing to assess the level of knowledge gained. Comprehension, recall, and risk perception will be measured.

6. Conclusions and Future Work

This project demonstrates how a serious video game can serve as a powerful tool for social impact, particularly in the field of public health. The development of 24h: The Hope Machine was driven by the belief that emotional immersion, not just factual instruction, is key to lasting awareness. By making players feel the urgency and consequences of meningitis, the game fosters not only understanding but empathy.

Emotional attachment was not an aesthetic byproduct but a core design strategy to enhance engagement and retention. When players emotionally identify with the characters and story, they are more likely to stay invested, complete the experience, and internalize the message. In this way, storytelling becomes a functional component of educational impact.

Although the game has not yet been empirically validated, its carefully constructed narrative, supported by health experts and rooted in real experience, shows great promise as a medium for learning and action. Future work will focus on analyzing player data, improving accessibility, expanding to additional languages, and refining the ways passive educational content is embedded within emotionally charged gameplay.

Ultimately, the project explores how empathy can be designed, and how interactive storytelling can turn awareness into action.

Declaration on Generative AI

During the preparation of this work, the author used X-GPT-4 in order to: Grammar and spelling check. After using these tool, the author reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] Spanish Association Against Meningitis (AEM). (2022). La AEM celebra la inclusión de la vacuna contra la meningitis en el SNS como “una gran noticia”. Infosalus. <https://www.infosalus.com/asistencia/noticia-aem-celebra-inclusion-vacuna-contra-meningitis-sns-gran-noticia-20221220141755.html>
- [2] Vandersmissen, B., Ruysschaert, M., Vanden Abeele, V., & De Marez, L. (2021). Engagement in serious games for health promotion: The role of narrative transportation and player agency. *Health Communication*, 36(1), 1–11. <https://doi.org/10.1080/10410236.2020.1778498>
- [3] Ingold, J. (2017). Narrative Sorcery: Coherent Storytelling in an Open World [Video]. Game Developers Conference (GDC). YouTube. https://www.youtube.com/watch?v=HZft_U4Fc-U
- [4] Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., & Pidruzny, J. N. (2015). Audial engagement: Effects of game sound on learner engagement in digital game-based learning environments. *Computers & Education*, 87, 199–212. https://www.sciencedirect.com/science/article/abs/pii/S0747563215000084?utm_source=chatgpt.com

Análisis de Sentimientos y Polarización de Reseñas de Videojuegos a Través de Inteligencia Artificial

Alberto Contreras-Puerto^{1,*†}, Andrés Muñoz^{1,†}, Manuel Palomo-Duarte^{1,†} and Juan Antonio Caballero-Hernández^{1,†}

¹Departamento de Ingeniería Informática, Universidad de Cádiz, Av. Universidad de Cádiz, 10, 11519 Puerto Real, Cádiz

Resumen

Los videojuegos generan miles de opiniones en línea cada día, pero comprender realmente qué piensan los usuarios no siempre es sencillo. El análisis de sentimientos no solo ayuda a los jugadores a identificar títulos bien valorados, sino que también puede ser útil en contextos profesionales, como en estudios donde se vinculan bonificaciones de productividad a la recepción crítica de un videojuego. En estos casos, una herramienta capaz de desglosar y evaluar automáticamente las reseñas permitiría establecer métricas más transparentes y centradas en aspectos específicos del producto, más allá de una simple media numérica. Este trabajo presenta una metodología para el análisis automatizado de sentimientos, polarización y calidad de las reseñas de videojuegos generadas por usuarios, utilizando el modelo de lenguaje LLaMA 3.2 (3B) de Meta AI, afinado con la librería Unsloth y un conjunto de 50.000 reseñas anotadas de Kaggle. El enfoque combina modelos de lenguaje de gran escala (LLMs) con técnicas clásicas de recuperación de información como TF-IDF y BM25 para generar resúmenes representativos de las opiniones. Además del análisis de sentimientos basado en aspectos que permite identificar y puntuar dimensiones clave como jugabilidad, gráficos, historia o sonido, el sistema evalúa también la calidad de cada reseña individual, determinando su polaridad (positiva, negativa o neutral) y asignando una calificación de 1 a 10 según su claridad, riqueza argumentativa y nivel de detalle. A partir de estas evaluaciones individuales, el sistema elabora una valoración general del videojuego que no solo refleja la media de opiniones, sino también los aspectos más destacados positiva o negativamente por la comunidad. Para validar la fiabilidad del modelo, se realizó una comparación manual con valoraciones humanas sobre una muestra total de 300 reseñas (20 reseñas evaluadas por 15 personas), lo que permitió contrastar las puntuaciones generadas por la IA con percepciones humanas reales. Los resultados demuestran que el sistema propuesto puede sintetizar con precisión la opinión colectiva de los usuarios, proporcionando una evaluación estructurada, transparente y centrada en múltiples dimensiones del producto.

Palabras Clave

Inteligencia Artificial, Reseñas de Videojuegos, Large Language Model (LLM), Fine-Tuning, Análisis de Sentimientos, Polarización

1. Introducción

En la actualidad, las reseñas generadas por los usuarios en plataformas digitales juegan un papel crucial en la percepción pública y el éxito comercial de productos culturales como los videojuegos. Estas opiniones, disponibles en grandes volúmenes en sitios como Steam¹—una plataforma de distribución digital de videojuegos desarrollada por Valve, contienen información valiosa sobre las experiencias y preferencias de los usuarios. Sin embargo, su análisis manual resulta inviable a gran escala, lo que ha impulsado el desarrollo de técnicas automáticas para extraer y resumir esta información de forma eficiente[1].

IV Congreso Español de Videojuegos: Granada, 18 y 19 de septiembre de 2025

*Corresponding author.

† These authors contributed equally.

✉ alberto.contreraspuerto@alum.uca.es (A. Contreras-Puerto); andres.munoz@uca.es (A. Muñoz); manuel.palomo@uca.es (M. Palomo-Duarte); juanantonio.caballero@uca.es (J. A. Caballero-Hernández)

🌐 <https://github.com/AlbeXos21> (A. Contreras-Puerto); <https://produccioncientifica.uca.es/investigadores/168224/detalle> (A. Muñoz); <https://produccioncientifica.uca.es/investigadores/113142/detalle> (M. Palomo-Duarte); <https://produccioncientifica.uca.es/investigadores/205633/detalle> (J. A. Caballero-Hernández)

🆔 0009-0008-8112-3243 (A. Contreras-Puerto); 0000-0002-8491-4592 (A. Muñoz); 0000-0001-6002-0319 (M. Palomo-Duarte); 0000-0001-9012-8961 (J. A. Caballero-Hernández)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹<https://store.steampowered.com/>

El análisis de sentimientos es una de las técnicas más utilizadas para abordar esta problemática, ya que permite identificar la polaridad (positiva, negativa o neutra) de los comentarios[2]. No obstante, los métodos tradicionales presentan limitaciones en cuanto a comprensión contextual, detección de ironía o identificación de aspectos específicos mencionados en las reseñas[3]. Para superar estas barreras, en este trabajo se propone un enfoque basado en modelos grandes de lenguaje (LLMs²) de última generación, específicamente LLaMA 3.2 (3B), desarrollado por Meta AI, el cual puede ser afinado mediante fine-tuning para adaptarse mejor a tareas específicas del dominio[4]. Utilizando un conjunto de 50.000 reseñas anotadas provenientes de Kaggle³, cada una con su correspondiente puntuación.

El *fine-tuning* es un proceso de ajuste supervisado en el que un modelo de lenguaje previamente entrenado se adapta a una tarea específica utilizando un conjunto de datos representativo del dominio objetivo. Esto permite mejorar su rendimiento sin necesidad de entrenarlo desde cero. Además de la clasificación de sentimientos, se incorporaron algoritmos de recuperación de información como TF-IDF[5] y BM25[6] para generar resúmenes representativos de las opiniones. Por otro lado, se diseñó una estrategia de análisis desglosado por criterios, con el objetivo de evaluar de forma diferenciada aspectos clave del videojuego, como la jugabilidad, la historia o la calidad gráfica.

A continuación, se organiza el contenido del artículo en las siguientes secciones:

1. Trabajos Relacionados, donde se revisan investigaciones previas.
2. Conjuntos de Datos y Aspectos del Sistema, que describe los datos y problemas abordados.
3. Metodología y Experimentos, donde se explica el enfoque propuesto y la configuración experimental.
4. Resultados, con el análisis de los hallazgos.
5. Amenazas a la validez, donde se discuten las posibles limitaciones del estudio.
6. Conclusiones, que resumen los aportes y plantean líneas futuras.

2. Trabajos Relacionados

El análisis de sentimientos en reseñas de videojuegos ha sido objeto de diversos estudios en los últimos años. Tan et al. realizaron un estudio comparativo de algoritmos de aprendizaje automático aplicados a reseñas de Steam, evaluando modelos como Support Vector Machines, Naïve Bayes y regresión logística. Sus resultados mostraron que SVM ofrecía la mayor precisión en la clasificación binaria de sentimientos, destacando la utilidad de enfoques supervisados clásicos en este dominio [7].

Taherdoost y Madanchian realizaron una revisión sistemática sobre el uso de inteligencia artificial y análisis de sentimientos en la investigación competitiva. Aunque su estudio se centra en el análisis de mercado, sus conclusiones son aplicables al dominio de los videojuegos, especialmente en lo referente a la detección de patrones de comportamiento del usuario y la evaluación comparativa de productos, destacando cómo estas técnicas permiten comprender mejor las percepciones de los consumidores en distintos sectores [2].

Angelidis y Lapata proponen un enfoque débilmente supervisado que combina extracción de aspectos y predicción de sentimientos, logrando resúmenes representativos a partir de grandes volúmenes de opiniones. Su método destaca por su capacidad de identificar automáticamente temas relevantes y sintetizar la polaridad asociada a cada uno de ellos [1].

Por su parte, Suhara et al. desarrollaron OpinionDigest, un sistema basado en Transformers, una arquitectura de red neuronal revolucionaria que ha transformado el campo del procesamiento del lenguaje natural (NLP) y otras áreas como visión por computadora, audio... que genera resúmenes de opiniones organizadas por aspectos, sin necesidad de entrenamiento supervisado directo. Esta propuesta representa un avance hacia sistemas más escalables y adaptables en dominios con grandes volúmenes de texto no estructurado [8].

²LLM = Large Language Model

³Kaggle es una plataforma en línea de ciencia de datos que ofrece competencias, conjuntos de datos públicos y recursos para el desarrollo y evaluación de modelos de aprendizaje automático.

Ghosh y Pakray[9] realizaron un análisis enfocado en reseñas de juegos móviles disponibles en Google Play Store, evaluando técnicas de procesamiento de lenguaje natural y clasificadores como Random Forest, SVM y redes neuronales profundas. Su estudio destacó que la preprocesamiento adecuado y la representación textual (TF-IDF, Word2Vec) tienen un papel fundamental en la mejora del rendimiento de los clasificadores.

Poria et al.[10] propusieron un enfoque que combina el análisis de aspectos con redes neuronales convolucionales (CNN) para clasificar opiniones a nivel de frase. El modelo identifica automáticamente aspectos relevantes dentro de las opiniones y predice su polaridad asociada.

En contraste con estos enfoques, la metodología presentada en este trabajo combina modelos de lenguaje de gran escala con técnicas clásicas de recuperación de información (TF-IDF y BM25), lo que permite generar resúmenes representativos sin perder interpretabilidad. Además, incorpora un análisis por criterios que descompone la opinión en dimensiones específicas del videojuego como jugabilidad, historia o gráficos, aportando un nivel adicional de estructuración que facilita la evaluación detallada de la experiencia del usuario. En la Tabla 1 donde se muestra una comparativa entre los diferentes proyectos mencionados y las diferencias con la propuesta planteada en este artículo.

Tabla 1

Comparativa de estudios sobre análisis de sentimientos

Estudio	SENT	ASP	EMO	VID	PLN
Tan et al. (2022)	Sí	No	No	Sí	No
Taherdoost y Madanchian (2023)	Sí	No	Sí	No	Sí
Angelidis y Lapata (2018)	Sí	Sí	No	No	Sí
Suhara et al. (2020)	Sí	Sí	No	No	Sí
Ghosh y Pakray (2021)	Sí	No	No	Sí	Sí
Poria et al. (2016)	Sí	Sí	No	No	Sí
Este trabajo	Sí	Sí	Sí	Sí	Sí

Notas: **SENT** = Análisis de sentimientos; **ASP** = Análisis de aspectos y polarización; **EMO** = Detección de emociones; **VID** = Análisis relacionado con video o contenido visual / evaluación de videojuegos; **PLN** = Uso de técnicas de Procesamiento de Lenguaje Natural.

3. Conjuntos de Datos y Aspectos del Sistema

El sistema propuesto combina LLM con técnicas tradicionales de recuperación de información para realizar análisis automatizado de reseñas de videojuegos. Su objetivo principal es generar resúmenes representativos de las opiniones de los usuarios, así como puntuar distintos aspectos del videojuego (como jugabilidad, historia o gráficos) a partir de texto libre. La arquitectura general incluye tres subsistemas principales: (1) extracción de reseñas relevantes mediante TF-IDF y BM25, (2) análisis de sentimientos por criterios utilizando un modelo LLaMA 3.2 ajustado por *fine-tuning*, y (3) síntesis de resultados en formato estructurado.

En esta sección se describen los diferentes recursos utilizados por el sistema en su funcionamiento. Primero se detalla el conjunto de datos empleado, incluyendo su origen, características principales y proceso de preparación. A continuación, se especifican las tareas definidas para el análisis, que guían tanto la estructura del modelo como los criterios de evaluación.

3.1. Conjunto de Datos

Antes del análisis, se aplica un proceso de preprocesamiento que incluyó la selección de columnas relevantes (texto de la reseña y puntuación), limpieza de caracteres especiales, normalización de mayúsculas, eliminación de stopwords⁴ y otros elementos no informativos, así como la detección de

⁴Las stopwords son palabras frecuentes que se omiten en el procesamiento textual por su escaso valor informativo, como “el”, “la”, “de” o “y”.

idioma. Solo se conservaron reseñas escritas en inglés, y no se retuvieron datos sensibles ni información personal identificable. Además, se aplicó un filtrado adicional para eliminar entradas vacías o que no tenían ningún significado que aportar. En particular, se consideraron ambiguas aquellas reseñas que contenían menos de tres palabras, lo cual se verificó contando si tenían menos de tres espacios en el texto.

3.2. Aspectos del Sistema

Antes de presentar los aspectos que aborda, es necesario describir brevemente la propuesta. El sistema desarrollado es una arquitectura híbrida para el análisis automatizado de reseñas de videojuegos generadas por los usuarios. Combina técnicas clásicas de recuperación de información (como TF-IDF y BM25) con un modelo de lenguaje ajustado mediante *fine-tuning* (LLaMA 3.2) para realizar análisis de sentimientos y extracción de aspectos relevantes. Su objetivo es sintetizar de manera estructurada la opinión colectiva de los usuarios, destacando las dimensiones clave de la experiencia, como jugabilidad, historia, gráficos o sonido.

A partir de este planteamiento, el sistema aborda tres aspectos principales:

- **Clasificación de Sentimientos:** Se utilizó el modelo LLaMA 3.2 (3B), ajustado mediante *fine-tuning* con la librería Unsloth, para predecir el sentimiento general de cada reseña. Unsloth es una herramienta de código abierto diseñada para acelerar y simplificar el ajuste fino de modelos de lenguaje de gran escala (LLMs), como LLaMA o Mistral. Está optimizada para una mayor eficiencia en GPU y permite entrenar modelos de forma más rápida y con menor consumo de memoria, lo que la hace especialmente útil en entornos con recursos limitados, en este aspecto se obtiene finalmente un conjunto de reseñas con sus notas correspondientes generadas por el modelo anteriormente mencionado con puntuaciones del 1 al 10.
- **Análisis Aspecto a Aspecto:** Si bien el modelo no fue ajustado específicamente para la tarea de detección de aspectos, se empleó una estrategia basada en una taxonomía propia de dimensiones relevantes, tales como jugabilidad, historia, gráficos, música, entre otras. Esta taxonomía fue construida a partir del análisis de distintos géneros de videojuegos y aplicada de forma adaptativa según el género declarado en la plataforma Steam, lo que permitió priorizar aquellos aspectos más significativos para cada tipo de juego.
Mediante el uso de algoritmos de similitud semántica, se identificaron frases clave dentro de las reseñas que hacían referencia a cada dimensión, y se evaluó su tono utilizando el modelo de lenguaje base. A partir de estas detecciones, se asignaron puntuaciones específicas por aspecto, derivadas directamente de las **notas generales generadas por la IA para cada reseña**. Estas notas globales, calculadas previamente en función del grado de polarización y la percepción del contenido, se redistribuyeron sobre los aspectos mencionados, promediándose entre múltiples reseñas para obtener una **nota media por dimensión**.
Este enfoque permitió construir una valoración desagregada y más matizada de cada videojuego, alineada con los elementos que realmente importan al público según el tipo de experiencia que ofrece cada género. A continuación, se presentan varias tablas que ilustran los criterios aplicados y las valoraciones por aspecto obtenidas en diferentes géneros.
- **Resumen de Opiniones:** Se aplicaron algoritmos de recuperación de información (TF-IDF y BM25) para extraer palabras/frases clave para saber los elementos que más se repetían dentro de las reseñas y a partir de estas generar un resumen completo del videojuego analizado.

4. Arquitectura y Entorno de Desarrollo

Esta sección describe la arquitectura del sistema propuesto, así como el enfoque experimental adoptado para validar su rendimiento. En primer lugar, se detalla la estructura general del sistema y el funcionamiento de cada uno de sus módulos Véase **Figura 1**.

Tabla 2
Criterios de evaluación para videojuegos de Aventura

Criterio	Descripción
Story/Plot	Calidad de la historia, coherencia y giros
World Design	Riqueza y creatividad en mapas y niveles
Characters	Complejidad y desarrollo de personajes

Tabla 3
Criterios de evaluación para videojuegos de Estrategia

Criterio	Descripción
Resource Management	Complejidad y eficiencia en la gestión de recursos
Level Design	Profundidad estratégica en el diseño de mapas
Tactics and Balance	Relevancia de las decisiones tácticas

Tabla 4
Criterios de evaluación para videojuegos de Simulación

Criterio	Descripción
Realistic Scenarios	Autenticidad en la recreación de entornos
Decision Making	Consecuencias de las decisiones del usuario
Interactivity and Development	Libertad del usuario

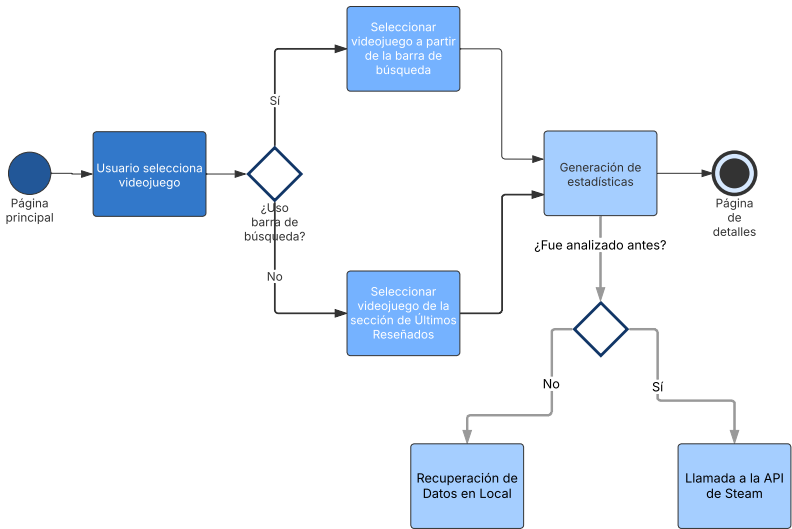


Figura 1: Diagrama del flujo del sistema en notación BPMN 2.0

4.1. Entorno de desarrollo y evaluación

El sistema fue desarrollado en Python, utilizando las bibliotecas PyTorch como backend principal, Unsloth para el ajuste eficiente del modelo LLaMA, y Hugging Face Transformers para la gestión e implementación del modelo. Todo el procesamiento de datos, así como las fases de entrenamiento e inferencia, se realizaron en una GPU Nvidia Geforce RTX 4060 TI, optimizando el uso de memoria mediante técnicas de cuantización y un tamaño de lote moderado.

Para evaluar el rendimiento del clasificador de sentimiento, se dividió el conjunto de datos en un 80 % para entrenamiento y un 20 % para prueba basándose en un estándar en este tipo de entrenamientos [11] [12], aplicando métricas estándar como precisión, *F1-score* y *accuracy* [13] [14] Véase **Figura 2**.

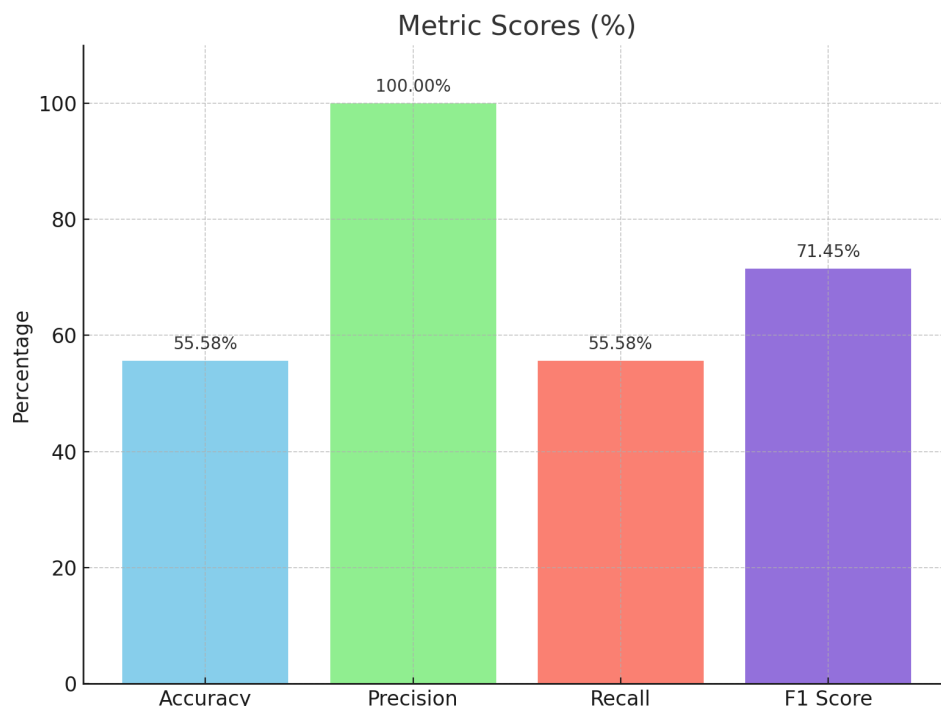


Figura 2: Métricas ->Accuracy, Precision y F1-Score

En cuanto al análisis por aspectos, se realizaron validaciones manuales sobre ejemplos individuales organizados por género del videojuego. Se comprobó que los aspectos extraídos correspondieran con menciones semánticas relevantes en las reseñas. Aunque esta parte del sistema se considera aún exploratoria debido a la falta de datos anotados por aspecto, ofrece resultados prometedores para visualizar opiniones estructuradas y detectar patrones relevantes de percepción del usuario.

5. Resultados

En esta sección se presentan los principales hallazgos obtenidos a partir de la implementación y evaluación del sistema propuesto. Los resultados se agrupan en torno a cinco dimensiones clave: la clasificación de sentimientos, la generación de resúmenes, el análisis desglosado por aspectos y una validación comparativa con evaluadores humanos.

5.1. Clasificación de Sentimientos

El modelo LLaMA 3.2 (3B), afinado con Unsloth, fue capaz de clasificar correctamente el tono general de las reseñas en la mayoría de los casos. Ya que se puede observar que el modelo distinguía de forma coherente entre opiniones positivas y negativas explícitas, aunque mostró dificultades en comentarios con tono irónico o ambigüedad emocional. No se utilizaron métricas cuantitativas automáticas, ya que la validación se centró en una inspección cualitativa de la salida del modelo.

5.2. Generación de resúmenes

Se implementó una metodología integral que combina técnicas de recuperación de información clásicas, específicamente TF-IDF y BM25, para analizar exhaustivamente el conjunto de reseñas asociadas

a cada videojuego. Mediante este enfoque, se identifican los términos y frases con mayor relevancia semántica y frecuencia relativa dentro del corpus, capturando así las opiniones más representativas y recurrentes expresadas por los usuarios. Cada reseña seleccionada se asocia a la valoración generada por el modelo mencionado anteriormente, permitiendo vincular el contenido textual con la polaridad y magnitud del sentimiento expresado.

Esta estrategia posibilita la generación de resúmenes que sintetizan fielmente tanto el contenido cualitativo como la tendencia cuantitativa de las opiniones, reflejando el consenso general o las áreas de controversia dentro de la comunidad.

5.3. Análisis Aspecto a Aspecto

El sistema fue capaz de identificar aspectos clave dentro de las reseñas como jugabilidad, historia, gráficos o sonido, basándose en una taxonomía propia diseñada según el género del videojuego. Se utilizó búsqueda semántica para asociar fragmentos relevantes a cada aspecto, y se asignó una valoración basada en el tono detectado. Esta evaluación se realizó manualmente en varios casos de prueba, ya que no se disponía de un conjunto de datos lo suficientemente amplio y etiquetado con análisis por aspectos como para ajustar y validar el modelo de forma automática, y además se empleó una taxonomía propia, dado que no existe una clasificación oficial estandarizada para este tipo de análisis. Se observó que ciertos aspectos ganaban protagonismo según el tipo de videojuego: por ejemplo, en títulos de estrategia se mencionaban más la inteligencia artificial y el balance, mientras que en RPGs destacaban la narrativa y el desarrollo de personajes.

5.4. Validación del Sistema

Para evaluar la precisión del modelo en la asignación de notas medias a las reseñas, se seleccionó un subconjunto de 20 reseñas para cada usuario. Se compararon las notas promedio generadas por la IA con las notas promedio asignadas por un grupo de evaluadores humanos. Estas calificaciones no solo se basan en el contenido explícito de la reseña, sino también en el grado de polarización (positiva o negativa) presente en el texto y en la percepción que genera en quienes la escriben o leen. El objetivo fue verificar si los resultados del modelo coincidían razonablemente con las valoraciones humanas, considerando tanto la carga subjetiva del lenguaje como el impacto percibido de cada reseña.

La Tabla [5] muestra los resultados obtenidos, incluyendo información demográfica de los evaluadores y las notas medias asignadas tanto por la IA como por los humanos.

Tabla 5

Comparación de notas medias asignadas por IA y evaluadores humanos

ID	Edad	Género	Experiencia	Nota Media IA	Nota Media Humana	Diferencia (%)
1	35-45	Masculino	Experto	6.65	5.95	7.0
2	18-25	Masculino	Experto	5.50	6.10	-6.0
3	18-25	Masculino	Experto	4.95	5.65	-7.0
4	35-45	Masculino	Intermedio	5.95	5.45	5.0
5	25-35	Masculino	Avanzado	5.60	4.55	10.5
6	18-25	Femenino	Intermedio	4.45	5.20	-7.5
7	18-25	Masculino	Intermedio	4.55	4.75	-2.0
8	25-35	Masculino	Intermedio	4.50	5.20	-7.0
9	25-35	Femenino	Principiante	5.50	6.05	-5.5
10	25-35	Masculino	Avanzado	5.10	5.40	-3.0
11	18-25	Masculino	Avanzado	5.05	4.20	8.5
12	18-25	Masculino	Intermedio	5.55	4.10	14.5
13	18-25	Masculino	Avanzado	5.40	5.45	-0.5
14	18-25	Femenino	Intermedio	6.20	5.15	10.5
15	18-25	Otro	Experto	4.70	4.75	-0.5

Como se observa, las notas medias asignadas por la IA muestran una buena aproximación a las

valoraciones humanas, aunque con ciertas diferencias puntuales, especialmente en casos con notas más extremas. Este experimento demuestra la viabilidad del modelo para capturar la percepción general de los evaluadores humanos en una escala del 1 al 10.

Para cuantificar la diferencia entre las notas asignadas por la IA y las valoraciones humanas, se calculó el Error Medio Absoluto (MAE, por sus siglas en inglés). Este indicador mide el promedio de las desviaciones absolutas entre ambas puntuaciones, proporcionando una medida intuitiva del error promedio del modelo. En este experimento, el MAE obtenido fue de aproximadamente 0.78 puntos en una escala de 1 a 10, lo que indica que, en promedio, las predicciones del modelo difieren menos de una unidad de las valoraciones humanas. Este resultado respalda la capacidad del modelo para replicar la percepción general de los evaluadores, aunque se identifican áreas para mejora, especialmente en la detección de valoraciones extremas o matizadas.

5.5. Amenazas a la validez

La principal limitación del sistema es la falta de un conjunto de datos etiquetado por aspectos, lo que impide realizar una evaluación cuantitativa precisa en esta dimensión. En consecuencia, la validación ha sido estrictamente cualitativa, basada en la inspección manual de ejemplos representativos, sin disponer de anotaciones de referencia para los resúmenes generados, lo que dificulta medir su cobertura o fidelidad de forma objetiva. También se observan limitaciones en la capacidad del sistema para interpretar ironía o ambigüedad emocional, así como una tendencia a simplificar en exceso cuando las opiniones están divididas.

Además, el modelo actual solo opera sobre reseñas en español, lo que restringe su aplicabilidad a otros contextos idiomáticos. Aunque estos retos no se abordan directamente en esta fase del trabajo, existen posibles vías de mejora: el uso de técnicas de anotación débil o autoetiquetado asistido por LLMs, la incorporación de detectores de sarcasmo y el despliegue de modelos multilingües podrían permitir ampliar y afinar el sistema en futuras iteraciones. Si bien el sistema no fue evaluado mediante métricas automáticas estandarizadas, las pruebas cualitativas mostraron resultados coherentes y útiles para el objetivo planteado. Esta limitación afecta principalmente a la validez externa y a la reproducibilidad de los resultados, ya que las observaciones se basan en ejemplos seleccionados manualmente. Por otro lado, es posible que algunos sesgos inherentes al corpus —como el predominio de ciertos géneros o estilos de reseña— influyan en los resultados. A futuro, la incorporación de evaluaciones automáticas y conjuntos de datos anotados permitiría reducir estas amenazas y fortalecer la robustez de las conclusiones.

6. Conclusiones

En este trabajo se ha presentado una propuesta metodológica y técnica para el análisis automático de sentimientos y polarización en reseñas de videojuegos, combinando el poder expresivo de modelos de lenguaje de gran escala (LLMs) con técnicas clásicas de recuperación de información. El enfoque se estructuró sobre el modelo LLaMA 3.2 (3B), ajustado eficientemente mediante la librería Unsloth, utilizando un conjunto de datos compuesto por reseñas de usuarios recopiladas desde la plataforma Kaggle.

Una de las principales contribuciones del sistema radica en la implementación de una taxonomía de aspectos personalizada, diseñada específicamente para el dominio de los videojuegos. Esta taxonomía permite descomponer las reseñas en componentes temáticos relevantes —como jugabilidad, gráficos, historia, sonido o dificultad— y adaptarse de forma dinámica según el género del videojuego (por ejemplo, acción, rol, estrategia, simulación). Este enfoque posibilita una mayor granularidad en el análisis, lo que permite obtener conclusiones más específicas y accionables para estudios de mercado, diseño de juegos y atención al cliente.

El sistema desarrollado facilita una representación más rica del contenido emocional y valorativo presente en los comentarios de los usuarios, no solo detectando la polaridad general (positiva, negativa o neutra), sino también proporcionando una visión detallada de las valoraciones a nivel de cada aspecto relevante del producto. Para validar esta capacidad, se realizó una comparación entre las puntuaciones

generadas por la IA y las valoraciones humanas, utilizando una muestra de aproximadamente 300 reseñas. Se observó que las notas medias asignadas por el modelo muestran una buena aproximación a las valoraciones humanas, aunque con ciertas diferencias puntuales, especialmente en casos con puntuaciones más extremas.

Para cuantificar esta diferencia, se calculó el Error Medio Absoluto (MAE), que mide el promedio de las desviaciones absolutas entre ambas puntuaciones y ofrece una medida intuitiva del error promedio del modelo. En este experimento, el MAE fue de aproximadamente 0.78 puntos en una escala de 1 a 10, lo que indica que, en promedio, las predicciones del modelo difieren menos de una unidad respecto a las valoraciones humanas. Este resultado respalda la capacidad del sistema para capturar la percepción general de los evaluadores, aunque se identifican oportunidades de mejora, particularmente en la detección de valoraciones extremas o emocionalmente matizadas.

7. Trabajo Futuro

Como trabajo futuro, se identifican diversas líneas de mejora y ampliación del sistema:

A futuro, se plantea como principal línea de desarrollo la mejora en la detección y análisis de aspectos mediante el entrenamiento de modelos específicos de análisis de sentimientos por aspecto (ABSA), lo cual permitiría una clasificación más precisa y granular. Esta mejora requerirá el uso de conjuntos de datos anotados y posiblemente estrategias de aprendizaje supervisado o transferido. Asimismo, se considera especialmente valioso incorporar módulos de análisis temporal que permitan estudiar la evolución de la percepción de un videojuego a lo largo del tiempo, por ejemplo, tras actualizaciones, parches o eventos relevantes. Esta dimensión longitudinal podría ofrecer insights de gran utilidad para desarrolladores y responsables de producto. Ambas líneas implican desafíos técnicos significativos, pero aportarían un valor añadido sustancial al sistema.

Declaración sobre el uso de Inteligencia Artificial Generativa / Declaration on Generative AI

Durante la preparación de este trabajo, el autor utilizó ChatGPT-4 con el fin de resolver dudas sobre sintaxis de LaTeX, redacción académica y corrección de errores. No se generó contenido directamente con la herramienta sin revisión ni edición significativa. El autor asume toda la responsabilidad sobre el contenido final de esta publicación.

During the preparation of this work, the author used ChatGPT-4 in order to clarify doubts related to LaTeX syntax, academic writing, and error resolution. No content was directly generated by the tool for inclusion without significant review and editing. The author takes full responsibility for the publication's content.

Agradecimientos

Esta publicación es parte del proyecto PID2023-149674OB-I00 (GENIALLE), financiado por MICIU/AEI/10.13039/501100011033 y por ERDF, EU.

Referencias

- [1] S. Angelidis, M. Lapata, Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, 2018, pp. 3675–3686. doi:10.18653/v1/D18-1402.

- [2] H. Taherdoost, M. Madanchian, Artificial intelligence and sentiment analysis: A review in competitive research, *Computers* 12 (2023) 37. URL: <https://www.mdpi.com/2073-431X/12/2/37>. doi:10.3390/computers12020037.
- [3] L. Zhang, S. Wang, B. Liu, Deep learning for sentiment analysis: A survey, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8 (2018) e1253. doi:10.1002/widm.1253.
- [4] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, Parameter-efficient fine-tuning of large-scale pre-trained language models, *Nature Machine Intelligence* (2023). doi:10.1038/s42256-023-00626-4.
- [5] G. Salton, C. Buckley, Term-weighting approaches in automatic text retrieval, *Information Processing & Management* 24 (1988) 513–523. doi:10.1016/0306-4573(88)90021-0.
- [6] S. Robertson, H. Zaragoza, The probabilistic relevance framework: Bm25 and beyond, *Foundations and Trends in Information Retrieval* 3 (2009) 333–389. doi:10.1561/15000000019.
- [7] W. Tan, T. Wang, V. C. Lee, Sentiment analysis of steam game reviews using machine learning algorithms, in: *2022 International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, IEEE, 2022, pp. 321–326. doi:10.56453/icdxa.2021.1023.
- [8] Y. Suhara, X. Wang, S. Angelidis, W.-C. Tan, Opiniondigest: A simple framework for opinion summarization, *arXiv preprint arXiv:2005.01995* (2020).
- [9] K. Ghosh, P. Pakray, Sentiment analysis of user reviews of mobile games using machine learning approaches, in: *Proceedings of the 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, IEEE, 2021, pp. 1236–1241. doi:10.1109/ICOSEC51865.2021.9592050.
- [10] S. Poria, E. Cambria, A. Gelbukh, Aspect extraction and sentiment classification with a convolutional neural network approach, *Knowledge-Based Systems* 108 (2016) 42–49. doi:10.1016/j.knosys.2016.05.008.
- [11] V. R. Joseph, Optimal ratio for data splitting, *arXiv preprint arXiv:2202.03326* (2022). URL: <https://arxiv.org/abs/2202.03326>.
- [12] X. Li, D. Fourches, Effect of dataset size and train/test split ratio in qspr/qsar modeling: a case study with a publicly available dataset, *Journal of cheminformatics* 13 (2021) 1–22. URL: <https://doi.org/10.1186/s13321-021-00519-2>. doi:10.1186/s13321-021-00519-2.
- [13] Y. A. Prasetyo, E. Utami, A. Yaqin, Investigating the impact of train/test split ratio on the performance of pre-trained models with custom datasets, *Journal of NLP and Machine Learning* (2024). URL: <https://thesai.org/Publications/ViewPaper?Code=IJACSA&Issue=2&SerialNo=35&Volume=15>, incluye comparación de métrica precisión, recall y F1-score según distintas proporciones de train/test.
- [14] W. Xu, J. Chen, Z. Ding, J. Wang, Text sentiment analysis and classification based on bidirectional gated recurrent units (grus) model, *arXiv preprint arXiv:2404.17123* (2024). URL: <https://arxiv.org/abs/2404.17123>, informe de métricas: precisión 95.9 %, recall 99.1 %, F1-score 97.4 % en test.

A. Recursos Online

Los recursos utilizados en este trabajo están disponibles públicamente y pueden consultarse en los siguientes enlaces:

- Repositorio del proyecto (GitHub): Código fuente, scripts y documentos asociados al desarrollo del trabajo.
- Kaggle Datasets: Plataforma desde la cual se obtuvo el conjunto de 50,000 reseñas de videojuegos utilizadas para el entrenamiento.

Una revisión general del estado de la investigación sobre el vínculo entre videojuegos y juegos de apuestas

Joan Arnedo-Moreno^{1,*†}, Paula Rodríguez-Rivera^{2,†}

¹Universitat Oberta de Catalunya, Barcelona, España

²Universidad de Vigo, Ourense, España

Resumen

La industria del videojuego ha experimentado un crecimiento exponencial, impulsado por la innovación tecnológica y la convergencia de plataformas. Dentro de este auge, las cajas botín (*loot boxes*) y las microtransacciones se han consolidado como mecanismos de monetización estratégica, generando miles de millones de dólares anualmente. Estos sistemas, que integran mecánicas similares a las apuestas en las experiencias de entretenimiento digital (fenómeno conocido como “*gamblification*”), han captado el interés académico debido a sus efectos psicosociales y las implicaciones de los sesgos cognitivos en la percepción del valor por parte de los usuarios. La literatura existente revela una correlación significativa entre la exposición a las cajas botín y constructos asociados al juego problemático, como la impulsividad y la tolerancia al riesgo. Asimismo, estudios desde la perspectiva del derecho digital han criticado la falta de definiciones operativas y una supervisión efectiva, así como la heterogeneidad regulatoria entre jurisdicciones. En respuesta, algunas administraciones han reclasificado ciertas cajas botín como modalidades de apuesta. Dada la proliferación de revisiones sistemáticas sobre este tema, el presente estudio busca unificar este conocimiento fragmentado a través de una revisión general (*umbrella review*). El objetivo es identificar y evaluar las revisiones más representativas en la literatura, ofreciendo un mapeo estructurado de las contribuciones conceptuales clave. Esto proporcionará un punto de partida rápido y fiable para investigadores y responsables de políticas públicas interesados en este campo, facilitando la investigación futura y el desarrollo normativo.

Keywords

Revision general, umbrella review, gamblificación, cajas botín, videojuegos, apuestas problemáticas

1. Introducción

La industria del videojuego ha experimentado un crecimiento exponencial desde finales del siglo XX, consolidándose como un pilar de la economía creativa y generando nuevas formas de sociabilidad mediada por tecnologías interactivas. Esta expansión está impulsada por la convergencia de plataformas, que comprenden consolas de alto rendimiento, dispositivos móviles, realidad virtual y servicios de streaming en la nube, y por la constante innovación en motores gráficos, infraestructuras de red de baja latencia y algoritmos de personalización basada en datos de usuario [1, 2]. Sin embargo, esto también ha significado la incorporación y normalización de diversos mecanismos de persuasión con el objetivo de que los jugadores jueguen el máximo de tiempo posible, o lleven a cabo gastos.

Dentro de este entramado, las mecánicas de azar denominadas cajas botín (*loot boxes*, en inglés) y las microtransacciones han adquirido relevancia estratégica como mecanismos de monetización diferencial. Dichos sistemas permiten a los jugadores adquirir contenidos cosméticos, potenciadores funcionales o expansiones narrativas a través de micropagos, combinando estímulos de refuerzo intermitente y expectativa de recompensa [3, 4]. Este fenómeno, englobado dentro del concepto de “*gamblification*” [5], implica la integración de mecánicas y dinámicas propias del juego de azar o apuestas, como probabilidades ocultas, recompensas variables y elementos de riesgo financiero, en experiencias de entretenimiento digital. El volumen de ingresos atribuido a estas prácticas ha superado recientemente los 24.400 millones de dólares anuales, lo que representa aproximadamente el 58 % de la facturación del mercado global

SECIVI'25: IV Congreso Español de Videojuegos, 18–19 Junio, Granada, España

*Corresponding author.

†These authors contributed equally.

✉ jarnedo@uoc.edu (J. Arnedo-Moreno); paula.rodriguez.rivera@uvigo.gal (P. Rodríguez-Rivera)

ORCID 0000-0002-6430-6469 (J. Arnedo-Moreno); 0000-0002-0415-4125 (P. Rodríguez-Rivera)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

de videojuegos para PC [6]. En el plano regulatorio, distintas administraciones, incluidas Bélgica, los Países Bajos y varias entidades federales de Australia, han reclasificado determinadas cajas botín como modalidades de apuesta, imponiendo requisitos de transparencia en las probabilidades de obtención y, en algunos casos, prohibiciones totales de comercialización sin un etiquetado de riesgo claro [7]. Simultáneamente, organismos como la UK Gambling Commission y la Comisión Europea han promovido consultas públicas y estudios técnicos para evaluar la adecuación de las legislaciones de juego de azar al contexto digital emergente [7].

Es por estos motivos que desde hace varios años se ha incrementado el interés académico por estas dinámicas en el diseño de videojuegos, explorando tanto los efectos psicosociales de la exposición a las cajas botín, revelando correlaciones significativas con constructos como la impulsividad, la búsqueda de sensaciones y la tolerancia al riesgo económico, indicadores de juego de apuestas problemático, como las implicaciones de los sesgos cognitivos, tales como el sesgo de dotación, la falacia del coste hundido y el efecto de encuadre, en la percepción de valor y la disposición a pagar de los usuarios [8, 9]. A su vez, los estudios desde la perspectiva del derecho digital han analizado críticamente la ausencia de definiciones operativas y de mecanismos efectivos de supervisión, así como la heterogeneidad regulatoria entre jurisdicciones, lo que dificulta la aplicación uniforme de políticas de protección al consumidor [10]. El resultado actual es la existencia de un corpus de literatura suficiente sobre la convergencia entre videojuegos y mecanismos de apuestas como para haber sido recopilado ya en sucesivas revisiones sistemáticas, de alcance (*scoping*) o críticas en un plazo relativamente corto de tiempo.

Por ello, el presente estudio pretende unificar toda esta información, evitando su dispersión y fragmentación, a través de una revisión general (*umbrella review*), cuyo propósito consiste en identificar y evaluar las revisiones más representativas en la literatura sobre mecánicas de azar y microtransacciones en videojuegos. Mediante un proceso sistemático de selección y síntesis, la revisión ofrece un mapeo estructurado de las revisiones y estudios clave, con sus aportaciones conceptuales esenciales, facilitando una referencia rápida y fiable, como punto de partida en la materia para investigadores noveles y responsables de políticas públicas interesados en profundizar en los principales temas del campo y orientar futuros desarrollos investigativos y normativos.

2. Metodología

El presente estudio ha seguido las recomendaciones para la calidad metodológica mediante la aplicación de los 14 elementos de la Herramienta de Evaluación de la Calidad para Estudios Observacionales de Cohorte y Transversales (*Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies*), para el caso de revisiones sistemáticas y metaanálisis [11]. Antes de realizar la revisión, se estableció la metodología, especificando las preguntas de investigación, la estrategia de búsqueda y los criterios de inclusión/exclusión. El proceso de revisión siguió la versión revisada de las directrices “*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*” (PRISMA) [12, 13]. Las preguntas de investigación (PI) preestablecidas fueron:

- **PI-1:** ¿Cuáles son las conclusiones principales si agregamos las síntesis de todas las revisiones de la literatura sobre la temática?
- **PI-2:** ¿Según las revisiones de la literatura hasta la fecha, cuáles son los artículos más relevantes?

2.1. Criterios de inclusión

En el presente estudio se incluyeron revisiones de diversos tipos que cumplieran los siguientes criterios: (i) publicadas en revistas revisadas por pares; (ii) escritas en idiomas comprensibles por los investigadores (Inglés, Español); (iii) con un proceso de extracción de datos claramente identificable; (iv) que incorporaran de manera significativa estudios sobre el impacto de los mecanismos de azar o microtransacciones sobre comportamientos problemáticos en juegos de apuestas. Sobre el último punto, se consideró tanto el efecto pasarela desde el videojuego a otros mecanismos de apuestas, ya sean tradicionales como surgidos a propósito del propio videojuego (e.g. *skin betting* en eSports), como el abuso

de aquellos mecanismos basados en el azar integrados dentro del juego para el fomento del consumo de microtransacciones, principalmente las cajas botín. La población de interés incluía adolescentes y adultos consumidores de videojuegos de cualquier género y de cualquier país. Se excluyeron revisiones narrativas o sin un proceso claro de extracción de datos y síntesis de los artículos incluidos, así como piezas de opinión o resúmenes. También se excluyeron aquellas revisiones centradas exclusivamente en el uso problemático de videojuegos (*problem gaming*) o el trastorno de juego por internet (*Internet Gaming Disorder*, IGD).

2.2. Estrategia de búsqueda y fuentes

La búsqueda se realizó en Febrero de 2025, utilizando las siguientes bases de datos científicas: *Web of Science*, *Scopus*, *ScienceDirect*, *PubMed* y *Springer*. El desarrollo de la estrategia de búsqueda se basó en el criterio de población, exposición, resultado e investigación (*Population, Exposure, Outcomes and Research*, PEOR) y la selección de palabras clave partió de investigaciones previas, en las que se destacan los términos cajas botín (*loot boxes*), microtransacciones, y juego de apuestas problemático (*problem gambling*). Se optó por una búsqueda poco restrictiva, que posteriormente se acotara a través de la selección de revisiones exclusivamente, y a través del criterio de elegibilidad (iii). Los términos de búsqueda finales se detallan en la Tabla 1.

Tabla 1
Estrategia de búsqueda y palabras clave

Estrategia	Palabras clave y expresión de búsqueda
Exposición (<i>Exposure</i>)	("video games" OR "loot box*" OR "micro*transactions") AND
Resultado (<i>Outcome</i>)	("problem gambling" OR "gambling disorder") AND
Investigación (<i>Research</i>)	review

2.3. Selección de estudios

Los resultados de la búsqueda de cada base de datos, incluyendo títulos, resúmenes, palabras clave e información bibliográfica, se importaron a un formulario en línea conectado a una hoja de cálculo con el propósito de facilitar la colaboración entre dos investigadores durante todo el proceso de revisión. En una primera selección, los investigadores usaron los mecanismos de los propios motores de búsqueda de las bases de datos para filtrar aquellos artículos marcados como revisiones (artículo de tipo "review").

Una vez identificadas estas revisiones, se descartaron los artículos repetidos entre búsquedas y los investigadores procedieron a una criba (*screening*) mediante la evaluación independiente de los títulos, resúmenes y palabras clave según los criterios de inclusión y exclusión predefinidos, especialmente los vinculados a la temática, registrando un voto de elegibilidad de cada artículo. Un voto negativo en esta etapa requería una justificación relacionada con los criterios de inclusión y exclusión. Algunos de los términos que se consideraron relevantes para la toma de una decisión, originales en inglés, fueron: *problem gambling*, *gambling*, *convergence*, *gamblification*, *loot boxes*, *microtransactions*, (*video*) *games*, (*systematic*) *review*, *children*, *adolescents*. A su vez, algunos términos comunes pero considerados indicadores de un estudio no relevante fueron: *addiction* (en general), *gambling* (en general), y *IGD* o *problem gaming* (si el estudio lo trataba en solitario).

Las discrepancias se resolvieron mediante una discusión en línea. Las principales razones de desacuerdo fueron los casos de (i) revisiones con un componente importante o centrado en casinos en línea o sociales (*social online casinos*), que no se consideraron estrictamente videojuegos para el presente estudio, y (ii) revisiones centradas en estudio del juego problemático causado por el consumo de microtransacciones, pero sin la implicación de mecanismos de apuestas. La concordancia entre evaluadores se calculó utilizando el índice Kappa de Cohen, considerándose aceptable con un valor de "sustancial" ($\geq 0,61$) [14].

Tras la fase de selección, uno de los investigadores descargó el texto completo de cada artículo, para un nuevo proceso de revisión por duplicado. Este replicó el de criba, con la emisión de votos. En esta etapa, se prestó especial atención a la inclusión de estudios de tipo revisión sistemática o de alcance, que incluyeran un proceso claro de extracción de datos y síntesis, de acuerdo con el criterio de inclusión (iv).

2.4. Extracción de datos

Una vez seleccionadas las revisiones finales a incluir en el presente estudio, la extracción de datos fue realizada por los dos investigadores. En este proceso se extrajeron, adicionalmente a los datos iniciales para la criba: bases de datos de origen en la búsqueda, revista de publicación, número de citas, rango de años contemplados por la revisión, lista de referencias de los estudios incluidos, los términos de búsqueda empleados para cada revisión y el foco del estudio.

3. Resultados

El diagrama de flujo que resume el proceso de revisión completo se muestra en la Figura 1, que llevó a un resultado final de 10 revisiones seleccionadas. Como aspecto a destacar, la mayoría de artículos no relevantes se identificaron durante el proceso de criba, mayoritariamente al considerarse fuera de la temática. Los temas principales de estas revisiones estaban vinculados con las adicciones en general, sin vínculo a los videojuegos, el trastorno de adicción a Internet, o exclusivamente sobre marketing, aspectos éticos, o de regulación. El motivo de no selección de revisiones durante la lectura del texto completo fue principalmente por la ausencia de un proceso formal de identificación de artículos incluidos y extracción de datos.

3.1. Características de las revisiones

Las características principales de las revisiones seleccionadas al final del proceso se resumen en la Tabla 2, ordenadas por año. Las columnas “Estudios” y “Únicos” muestran, respectivamente, el número de estudios primarios contemplados en cada revisión, y de todos ellos, cuantos aparecen solo en dicha revisión. “Rango de años” indica el rango real de los estudios incluidos, independientemente del rango pretendido en la búsqueda (e.g. revisiones sin cota inferior o superior en sus criterios de inclusión).

Las fechas de publicación de todas las revisiones están entre el año 2020 y 2023, cobrando una especial relevancia este último año, con cuatro revisiones, que suman una gran cantidad de estudios. Ninguna de las revisiones contempla el país o el género como criterio de selección de estudios, aunque una de ellas considera un rango de edad para los estudios incluidos, limitándose a participantes de hasta 25 años. Solo dos incorporan idiomas distintos al inglés en sus criterios de búsqueda, incluyendo el español, francés o portugués. [15, 16].

En este sentido, la mayoría de revisiones acotan deliberadamente la búsqueda a ciertos años identificados como clave por distintos motivos, y que pueden ser interesantes para investigadores sobre la temática. Por ejemplo, algunos estudios como [17] no consideran años anteriores al 2000, por la falta de prevalencia de Internet, y para obviar estudios con objetivos similares, pero orientados a las máquinas recreativas (*coin-op*) [18]. Por otro lado, en [6] se considera que es a partir del año 2010 cuando las microtransacciones empiezan a evolucionar desde simples elementos de personalización hacia mecanismos íntimamente asociados a apuestas o sistemas de azar. El año 2013 es considerado relevante en [4] al ser cuando se introdujo formalmente IGD en el DSM-5 como una posible adicción conductual [19]. Pero, incluso con estas acotaciones iniciales, la mayoría de rangos de años de los estudios incluidos al final del proceso de revisión son relativamente recientes y tienen un sesgo alrededor de los inicios de la década de 2020 (ver también más adelante la Tabla 3).

El enfoque principal de las revisiones abarca diversos temas, pero se puede apreciar una evolución interesante, que parte del estudio de la convergencia entre el uso videojuegos en general y los juegos de apuestas tradicionales (por ejemplo, en [20] solo se usa ‘*gambling*’ AND ‘*video-gam*’ como término de

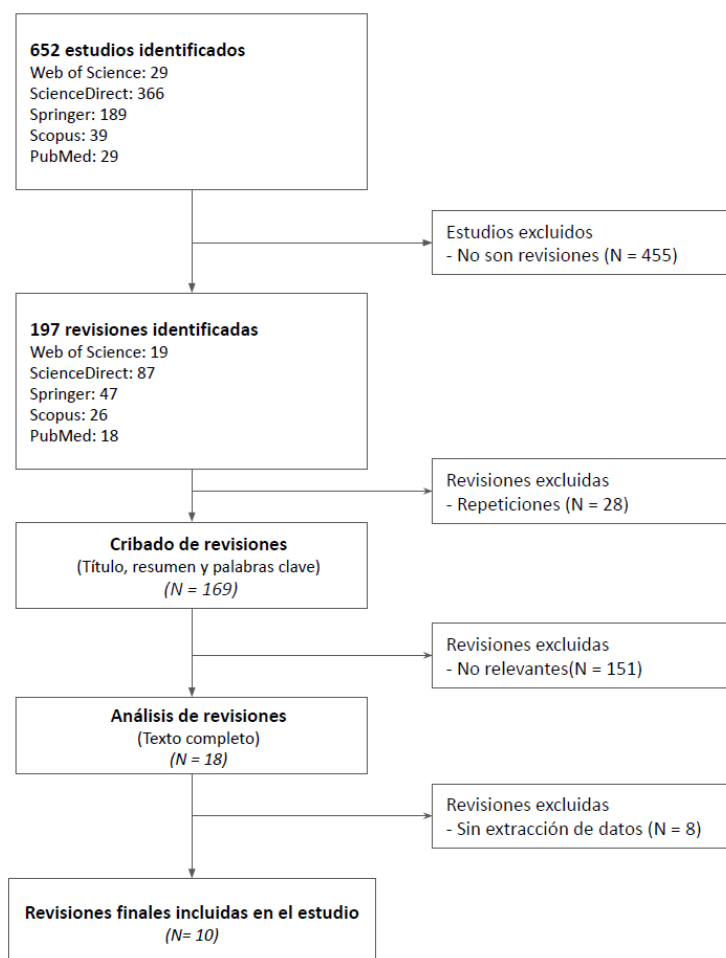


Figura 1: Diagrama de flujo PRISMA de la estrategia de búsqueda y selección de revisiones.

búsqueda), a una mayor focalización hacia ciertos elementos concretos considerados problemáticos, como las cajas botín y el uso de microtransacciones. Sin embargo, más adelante, se vuelve a una visión más general, orientada hacia las llamadas “actividades similares a apuestas” (*Gambling-like activities*, GLA), o los sistemas conceptualmente “gamblificados” [21]. En ambos casos, el comportamiento de apuestas problemáticas ya se estudia como una parte integral del videojuego, como forma de mecanismos de persuasión para el consumo de microtransacciones, en vez de investigar el efecto pasarela hacia el juego tradicional. En este sentido, en algunos estudios como [22, 15] se incluye cualquier elemento vinculado a apuestas digitales, como casinos sociales, *skin betting*, o simuladores de juego, pero se han considerado en la presente revisión, pues también incluyen aspectos específicos de los videojuegos, y las cajas botín.

Finalmente, y cumpliendo el propósito del presente estudio, todas las conclusiones de las revisiones consideradas incluyen observaciones sobre la asociación entre el uso de cajas botín y/o microtransacciones y el juego de apuestas problemático, si bien algunas incluyen también de manera complementaria la asociación con IGD o el uso problemático de videojuegos [17, 4]. Algunas lo hacen en el contexto de un estudio más amplio sobre formas de juego de apuesta digital, pero que incluyen las cajas botín, como ya se ha mencionado para el caso de [15, 22]. Sin embargo, diversos estudios analizan otros factores, como el impacto en el estado de ánimo [23]. Solo algunas revisiones van más allá, y a partir de esta información, identifican las lagunas metodológicas problemáticas en los estudios hasta el momento, ofreciendo sugerencias para el futuro, siendo [24] un caso muy destacado en este aspecto. Solo unas pocas estudian las implicaciones regulatorias [17, 6] u ofrecen sugerencias sobre como mejorar la aplicación de las cajas botín [24]. En el Apartado 4 se discuten algunos de estos aspectos con más detalle.

Tabla 2

Revisiones sistemáticas identificadas (ordenadas cronológicamente por año)

Referencia	Estudios	Únicos	Rango años	Foco de la revisión
Delfabbro & King, 2020 [20]	19	1	1995-2020	Videojuegos (en general)
Yokomitsu <i>et al.</i> , 2021 [23]	20	1	2019-2021	Cajas botín
Spicer <i>et al.</i> , 2021 [17]	31	9	2004-2021	Videojuegos (en general)
Gibson <i>et al.</i> , 2022 [6]	19	3	2018-2021	Microtransacciones
Raneri <i>et al.</i> , 2022 [4]	12	0	2018-2020	Microtransacciones y cajas botín
Murias <i>et al.</i> , 2022 [16]	22	0	2018-2021	Cajas botín
Richard & King, 2023 [15]	45	32	2024-2021	Actividades similares a apuestas (GLA)
Kim <i>et al.</i> , 2023 [25]	28	16	2017-2023	Actividades similares a apuestas (GLA)
McCaffrey, 2023 [24]	25	2	2018-2021	Cajas botín
Casu <i>et al.</i> , 2023 [22]	23	18	2012-2023	Juegos de móvil o por Internet

3.2. Características de los estudios incluidos

La extracción de todos los estudios incluidos en las revisiones suma un total de 123 artículos diferentes, una vez eliminadas las repeticiones, que sirven de base para algunas de las conclusiones del presente trabajo (PI-2). De todos ellos, 84 (64 %) se encuentran en una sola revisión, y el resto se superponen en las distintas revisiones. Los 25 estudios que aparecen en 3 o más casos se resumen en la Tabla 3, como posible indicador de estudios relevantes o populares en la temática. Si bien ninguno aparece en todas las revisiones, hay seis casos con 8 o 9 apariciones, cuatro de ellos procedentes del mismo autor (Zendle).

El grado de superposición varía ostensiblemente entre revisiones (ver columna “Revisiones” en la Tabla 3), especialmente según las diferencias en el foco de cada una. Sin embargo, existen dos revisiones [4, 16], que están totalmente superpuestas por el resto (ver columna “Únicos”, en Tabla 2). No incluyen ningún estudio que no aparezca únicamente en alguna de ellas, lo cual podría indicar *a priori* que son revisiones redundantes. Esto es a pesar de que, por ejemplo, la segunda contempla artículos también en español en su búsqueda, mientras que la gran mayoría se limita al inglés. En el otro extremo, más del 70 % de los estudios incluidos en [15, 22] son exclusivos de estas revisiones, seguramente al tener un foco temático mucho más amplio.

4. Discusión

La presente revisión general integra y evalúa de manera sistemática un conjunto de revisiones sobre la convergencia entre videojuegos y mecánicas de azar, revelando un estado del arte amplio pero fragmentado en términos metodológicos, temporales y temáticos. Mientras que la mayoría de los estudios coincide en la asociación de cajas botín y microtransacciones con indicadores de riesgo conductual, las divergencias en criterios de inclusión, definiciones operativas y periodos de búsqueda obstaculizan la obtención de conclusiones sólidas y comparables.

4.1. PI-1: Síntesis general de todas las revisiones

Los resultados de esta revisión evidencian correlaciones consistentes entre las mecánicas de azar en videojuegos -particularmente las cajas botín- y marcadores de riesgo de juego problemático. A través de los estudios, se identifica que el concepto de caja botín abarca formatos tecnológicos variados (e.g. cajas con elementos cosméticos, *gacha* y *kompu gacha*) con algoritmos de probabilidad y refuerzo intermitente heterogéneos, a la vez que al menos tres modalidades de diseño (aleatoriedad pura, microtransacciones condicionadas a logros, economía de *tokens*). El análisis de las revisiones sistemáticas (ver Tabla 2)

Tabla 3

Artículos incluidos en tres o más revisiones (ordenados cronológicamente por año)

Referencia	Revisiones	Muestra	Edades	País
Wood <i>et al.</i> , 2004 [26]	3	122	Adultos	Canadá
Delfabbro <i>et al.</i> , 2009 [27]	3	2669	<18	Australia
Macey & Hamari, 2018 [28]	4	582	Adultos	R. Unido
Zendle & Cairns, 2018 [29]	8	7422	Adultos	Mundial
Zendle <i>et al.</i> , 2019 [30]	8	1155	16-18	Mundial
Molde <i>et al.</i> , 2019 [31]	3	4601	16–74	Noruega
Brooks & Clark, 2019 [32]	9	113	Universidad	Canada
Zendle, 2019 [33]	5	99	Adultos	R. Unido
Macey & Hamari, 2019 [34]	6	482	>= 14	Mundial
Zendle & Cairns, 2019 [35]	8	1172	Adultos	EEUU
Zendle, 2019 [36]	6	112	Adultos	Mundial
Li <i>et al.</i> , 2019 [37]	7	618	Adultos	Mundial
Wardle, 2019 [38]	3	2881	<18	R. Unido
King <i>et al.</i> , 2020 [39]	3	482	Adultos	Varios
Kristiansen & Severin, 2020 [40]	9	1137	12-16	Dinamarca
Von Meduna <i>et al.</i> , 2020 [41]	5	46136	Adultos	Varios
Zendle <i>et al.</i> , 2020 [42]	8	1200	Adultos	Mundial
King <i>et al.</i> , 2020 [43]	5	263	18-25	EEUU
Brady & Prentice, 2021 [44]	4	25	21-30	Irlanda
Ide <i>et al.</i> , 2021 [45]	5	1615	14	Japón
Hall <i>et al.</i> , 2021 [46]	5	1155	19-80	Varios
DeCamp, 2021 [47]	4	13042	No indicado	EEUU
Wardle & Zendle, 2021 [48]	7	3549	16-24	R. Unido
Larche <i>et al.</i> , 2021 [49]	3	47	No indicado	Canadá
Rockloff <i>et al.</i> , 2021 [50]	5	1954	No 12-24	Australia

identifica que tanto la frecuencia de apertura como el gasto en microtransacciones correlacionan significativamente con escalas validadas de juego problemático (PGSI, SOGS), con tamaños de efecto moderados-altos en estudios transversales [40, 41].

Cabe destacar que, pese a la diversidad metodológica de los estudios, emergen patrones comunes: las mecánicas de recompensa intermitente (principalmente cajas botín a través de microtransacciones) se asocian consistentemente con indicadores de riesgo de juego problemático tales como impulsividad, búsqueda de sensaciones y sesgos cognitivos [17, 6, 24]. Esta asociación persiste tras controlar variables demográficas y muestra un gradiente dosis-respuesta: los jugadores que destinan 5 % de sus ingresos mensuales presentan puntuaciones un 32 % superiores en riesgo respecto a no consumidores [29]. Estos patrones, replicados de distinta forma en los estudios primarios incluidos en las diferentes revisiones, establecen un sustrato empírico sólido para examinar sus implicaciones.

Por otra parte, varias revisiones como [17, 6, 15] plantean recomendaciones regulatorias que, sin embargo, resultan parciales y muy dependientes del contexto nacional (mientras Bélgica y Países Bajos reclasifican ciertas cajas botín como apuestas, otras jurisdicciones se limitan a guías no vinculantes). En

estos trabajos se propone obligar a los desarrolladores a revelar las probabilidades de obtención de cada recompensa (por ejemplo, divulgación en el cliente del juego) y establecer límites de gasto automático para usuarios vulnerables. En [39] se sugieren marcos de cooperación internacional que armonicen estándares técnicos (por ejemplo, un etiquetado común de riesgo financiero) y sistemas de seguimiento de la eficacia de las medidas mediante indicadores de reducción de daños. Pocos países, no obstante, han realizado estudios de impacto que cuantifiquen la efectividad real de estas intervenciones, lo que resalta la necesidad de mecanismos de monitorización y revisión regulatoria continuas.

Sin embargo, a lo largo de las distintas revisiones se ponen de manifiesto también lagunas metodológicas relevantes que se deben tener presentes al discutir de estos hallazgos, tanto en los estudios iniciales como en los análisis más recientes. Más de la mitad de los estudios primarios no evalúa su propia calidad, carece de justificación estadística para el tamaño muestral, recurre mayoritariamente a muestreos de conveniencia, o se basan en cuestionarios autorreportados. Una gran mayoría son estudios transversales, prescindiendo de diseños longitudinales que permitan trazar realmente trayectorias de riesgo en el tiempo y establecer causalidad, quedando limitados a correlación [27, 24].

A estas carencias se suma una notable heterogeneidad terminológica (variaciones en la definición de caja botín e inclusión indistinta de actividades como *skin betting* o *social casino games*) y la dispersión de los periodos de búsqueda (generalmente centrados en 2018–2021), lo que dificulta la comparación directa de resultados y la generación de recomendaciones sólidas y homogéneas. Un aspecto destacado es que el uso de instrumentos para la medida de desórdenes de juego de apuestas problemáticos, sin establecer siempre claramente si se está consultando sobre actividades explícitamente externas a los videojuegos (e.g. lotería, tragaperras) o no (e.g. cajas botín), de modo que la interpretación puede ser diferente para cada individuo de la muestra. Ninguna revisión desagrega sistemáticamente sus resultados por plataforma (móvil, consola, PC), pese a que estudios primarios apuntan a que la interfaz y la accesibilidad modulan la frecuencia de apertura y el riesgo asociado [17, 25].

Por último, la ausencia de un análisis sistemático por género en las revisiones constituye una omisión relevante. Aunque algunos estudios documentan que los hombres presentan niveles más altos de impulsividad y frecuencia de apertura de cajas botín, y las mujeres muestran una mayor reactividad emocional ante la incertidumbre de la recompensa con correlaciones más fuertes con ansiedad y búsqueda de consuelo, ningún trabajo de revisión ha integrado estos hallazgos para perfilar trayectorias de riesgo diferenciadas por sexo [37, 48]. Tradicionalmente, la mayor incidencia masculina en el consumo de cajas botín se ha vinculado a su sobrerrepresentación histórica en la comunidad de jugadores y a motivaciones competitivas asociadas. No obstante, el Anuario AEVI 2024¹ informa que el 55 % de la población jugadora en España ya es femenina, inversión de la tendencia de décadas. A pesar de este cambio demográfico, los hombres mantienen un consumo significativamente superior de cajas botín, lo que sugiere que factores psicosociales como la socialización competitiva, la búsqueda de estatus y las normas de asertividad masculina continúan modulando su mayor propensión a estas mecánicas de refuerzo intermitente.

Si bien todas estas carencias, y especialmente la imperiosa necesidad de llevar a cabo estudios longitudinales, se exponen claramente ya en revisiones desde el año 2018, muchas todavía persisten a fecha de hoy, y por lo tanto, suponen una oportunidad de investigación en esta temática, a pesar de la gran cantidad de estudios existentes.

4.2. PI-2: Artículos más relevantes en la temática

La Tabla 3 se puede considerar un resumen de los estudios primarios más significativos, dada su ratio de aparición de las diferentes revisiones analizadas, en algunos casos incluso cuando su foco es un poco distinto. A través de ellos se puede observar como la evolución del corpus empírico confirma que la apertura de cajas botín opera como disparador de conductas de juego problemático, tal como se anticipó en contextos canadiense y australiano [26, 27]. A lo largo de la última década se han refrendado de manera consistente esta relación mediante el uso de escalas validadas (PGSI, RLI) en

¹<https://spainaudiovisualhub.mineco.gob.es/es/actualidad/aevi-la-industria-del-videojuego-en-espana-anuario-2024->

muestras multiculturales, subrayando la estabilidad del fenómeno más allá de las particularidades locales [29, 30, 35, 42]. Sin embargo, también se advierte que una proporción significativa de jugadores expuestos a microtransacciones no desarrolla síntomas clínicos de ludopatía, lo que pone de manifiesto la relevancia de variables individuales (por ejemplo, autocontrol financiero o experiencia previa con apuestas) en la modulación del riesgo [41, 47].

El desplazamiento del interés hacia modalidades afines, como el *skin betting* en eSports, ha enriquecido la visión conceptual. Se ha demostrado que, pese a las diferencias tecnológicas y de interfaz, las apuestas por *skins* reproducen procesos de refuerzo intermitente y sesgos cognitivos análogos a los de las cajas botín [28, 34]. Este planteamiento encuentra matices en otros trabajos, que destacan el papel del contexto competitivo y la presión de grupo en entornos universitarios, donde la reactividad emocional emerge como factor clave para la progresión hacia comportamientos de riesgo [44, 49].

La única aportación longitudinal sustantiva realizó un seguimiento de población adolescente danesa, confirmando que la exposición temprana a recompensas azarosas predice un aumento sostenido de síntomas problemáticos, proporcionando soporte temporal a la hipótesis pasarela [40]. En paralelo, otros trabajos introducen una perspectiva demográfica al documentar diferencias de género y edad: los varones presentan puntuaciones más elevadas de impulsividad y de frecuencia de apertura de cajas botín, mientras que las mujeres registran una vinculación afectiva más intensa y utilizan estas mecánicas como estrategia de regulación emocional; sin embargo, estas divergencias aún no han sido sistematizadas en un análisis integrado que permita perfilar ventanas de vulnerabilidad específicas [37, 48].

La evaluación de intervenciones regulatorias añade una dimensión aplicada al debate. Por ejemplo, asociar en Japón la obligatoriedad de etiquetado de probabilidades con una reducción del consumo problemático [45], o la observación en contextos multicéntricos que las restricciones técnicas, sin programas psicoeducativos de acompañamiento, carecen de eficacia para mitigar la ludopatía [46, 50]

5. Conclusiones

La presente revisión general ha permitido sintetizar de manera exhaustiva la evidencia generada por las principales revisiones sistemáticas sobre el vínculo entre videojuegos y apuestas, confirmando la asociación entre las mecánicas de recompensa intermitente (cajas botín, microtransacciones, *skin betting*) y el riesgo de desarrollo de comportamientos problemáticos análogos a la ludopatía [26, 27]. Sin embargo, esta interpretación requiere matices fundamentales. La marcada prevalencia de diseños transversales (95,1 % de los estudios primarios incluidos en las revisiones analizadas) impide establecer direccionalidad en las asociaciones. Como han señalado algunos autores [42], no puede descartarse que la relación entre cajas botín y juego problemático responda tanto a un efecto de inducción —en el que estas mecánicas actuarían como vía de entrada hacia el juego de azar convencional— como a un fenómeno de autoselección, en el que individuos con mayor vulnerabilidad al juego problemático tienden a involucrarse con videojuegos que incorporan dinámicas de azar o apuestas. Esta doble vía interpretativa cuestiona la validez de una interpretación unívoca basada exclusivamente en la hipótesis pasarela. Sin embargo, ambas vías, lejos de ser excluyentes, refuerzan la necesidad de estudios longitudinales que permitan esclarecer la direccionalidad de los efectos observados y delimitar perfiles de riesgo con mayor precisión.

Más significativamente, los hallazgos sugieren un desplazamiento conceptual: la gamblicación ha transformado el videojuego en un entorno de riesgo autónomo, en el que elementos como las cajas botín o el *skin betting* operan como formas consolidadas de apuesta digital, sin necesidad de conexión con productos tradicionales del juego (e.g. lotería, tragaperras). Este giro apunta a una lógica de consumo autorreferencial, donde la exposición sostenida a estas mecánicas puede generar patrones problemáticos dentro del propio ecosistema lúdico, afectando especialmente a perfiles vulnerables desde un plano psicosocial y conductual, independientemente del sentido real de la causalidad en la relación identificada.

A pesar de este hecho, la integración de modalidades derivadas (por ejemplo, *skin betting* en competiciones de eSports) y la incorporación de algunos estudios longitudinales en población adolescente [40] han enriquecido el marco conceptual, evidenciando tanto la continuidad de los procesos de refuerzo

intermitente como la existencia de perfiles de resiliencia conductual que modulan el potencial patológico de estas mecánicas. Del mismo modo, las diferencias observadas por sexo y edad señalan la necesidad de incorporar perspectivas demográficas para perfilar ventanas de vulnerabilidad específicas y orientar intervenciones preventivas con un enfoque diferencial.

El presente estudio tiene como principal limitación su alcance temporal, puesto que las revisiones identificadas abarcan únicamente hasta 2023. Este marco cronológico no impide, no obstante, que el análisis cumpla uno de sus objetivos nucleares: identificar y cartografiar las revisiones y artículos fundacionales que constituyen el canon actual de la literatura sobre el vínculo entre videojuegos y apuestas, en sus diversos matices: desde el efecto pasarela hasta la convergencia. De hecho, la síntesis realizada permite reconocer contribuciones seminales [33, 40, 38] cuyos hallazgos han establecido paradigmas conceptuales y metodológicos ampliamente replicados, configurando así el núcleo de referencia para investigadores que se aproximan al campo. Como trabajo futuro, sería valioso complementar este corpus mediante una revisión directa de estudios primarios publicados entre 2024 y 2025, evaluando si nuevas evidencias modifican sustancialmente las asociaciones documentadas. A su vez, otra oportunidad de investigación relevante sería la inclusión en el análisis de otros mecanismos de persuasión, como es la aplicación de los llamados “patrones oscuros” (*dark patterns*) [51].

Agradecimientos

La presente investigación se ha llevado a cabo a como parte del proyecto competitivo financiado “*Gamblificación, juventud y videojuegos: Investigación para promover prácticas de juego responsables*” (SUBV24/00026).

Referencias

- [1] H. T. Crogman, V. D. Cano, E. Pacheco, R. B. Sonawane, R. Boroon, Virtual reality, augmented reality, and mixed reality in experiential learning: Transforming educational paradigms, *Education Sciences* 15 (2025) 303.
- [2] M. Cerezo-Pizarro, F.-I. Revuelta-Domínguez, J. Guerra-Antequera, J. Melo-Sánchez, The cultural impact of video games: A systematic review of the literature, *Education Sciences* 13 (2023) 1116.
- [3] D. Varelas, E. Bonan, L. Anderson, A. Englesson, C. Åhrling, A. Chmielewski-Anders, Development of an end-to-end machine learning system with application to in-app purchases, *arXiv preprint arXiv:2412.12390* (2024).
- [4] P. C. Raneri, C. Montag, D. Rozgonjuk, J. Satel, H. M. Pontes, The role of microtransactions in internet gaming disorder and gambling disorder: A preregistered systematic review, *Addictive Behaviors Reports* 15 (2022) 100415.
- [5] J. Macey, J. Hamari, Gamblification: A definition, *New media & society* 26 (2024) 2046–2065.
- [6] E. Gibson, M. D. Griffiths, F. Calado, A. Harris, The relationship between videogame microtransactions and problem gaming and gambling: A systematic review, *Computers in Human Behavior* 131 (2022) 107219.
- [7] D. L. King, P. H. Delfabbro, S. M. Gainsbury, M. Dreier, N. Greer, J. Billieux, Unfair play? video games as exploitative monetized services: An examination of game patents from a consumer protection perspective, *Computers in Human Behavior* 101 (2019) 131–143.
- [8] F. J. Sanmartín, J. Velasco, F. Cuadrado, J. A. Moriana, The thrill of chance: Psychophysiological responses in loot boxes and simulated slot machines, *Computers in Human Behavior Reports* 16 (2024) 100545.
- [9] C. L. Beard, R. E. Wickham, Gaming-contingent self-worth, gaming motivation, and internet gaming disorder, *Computers in Human Behavior* 61 (2016) 507–515.
- [10] D. Leahy, Rocking the boat: Loot boxes in online digital games, the regulatory challenge, and the eu’s unfair commercial practices directive, *Journal of Consumer Policy* 45 (2022) 561–592.

- [11] National Heart, Lung, and Blood Institute, Quality assessment tool for observational cohort and cross-sectional studies., (Accedido, Mayo 2025). <https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools>.
- [12] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, et al., The prisma 2020 statement: an updated guideline for reporting systematic reviews, *bmj* 372 (2021).
- [13] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, D. Moher, Updating guidance for reporting systematic reviews: development of the prisma 2020 statement, *Journal of clinical epidemiology* 134 (2021) 103–112.
- [14] M. L. McHugh, Interrater reliability: the kappa statistic, *Biochemia medica* 22 (2012) 276–282.
- [15] J. Richard, S. M. King, Annual research review: Emergence of problem gambling from childhood to emerging adulthood: a systematic review, *Journal of Child Psychology and Psychiatry* 64 (2023) 645–688.
- [16] P. Murias, A. Grande Gosende, G. García Fernández, J. R. Fernández Hermida, et al., Cajas botín, juegos de azar y videojuegos: Una revisión sistemática, *Health and Addictions/Salud y Drogas* (2022).
- [17] S. G. Spicer, L. L. Nicklin, M. Uther, J. Lloyd, H. Lloyd, J. Close, Loot boxes, problem gambling and problem video gaming: A systematic review and meta-synthesis, *New Media & Society* 24 (2022) 1001–1022.
- [18] R. Ladouceur, D. Dubé, Prevalence of pathological gambling and associated problems in individuals who visit non-gambling video arcades, *Journal of gambling Studies* 11 (1995) 361–365.
- [19] F. Edition, et al., Diagnostic and statistical manual of mental disorders, *Am Psychiatric Assoc* 21 (2013) 591–643.
- [20] P. Delfabbro, D. L. King, Gaming-gambling convergence: Evaluating evidence for the ‘gateway’ hypothesis, *International Gambling Studies* 20 (2020) 380–392.
- [21] J. Macey, J. Hamari, M. Adam, A conceptual framework for understanding and identifying gamblified experiences, *Computers in Human Behavior* 152 (2024) 108087.
- [22] M. Casu, C. I. Belfiore, P. Caponnetto, Rolling the dice: A comprehensive review of the new forms of gambling and psychological clinical recommendations, *Psychiatry International* 4 (2023) 105–125.
- [23] K. Yokomitsu, T. Irie, H. Shinkawa, M. Tanaka, Characteristics of gamers who purchase loot box: A systematic literature review, *Current Addiction Reports* 8 (2021) 481–493.
- [24] M. McCaffrey, Loot boxes, problem gambling, and problem gaming: A critical review of the emerging literature, *Communications of the Association for Information Systems* 52 (2023) 132–163.
- [25] H. S. Kim, R. D. Leslie, S. H. Stewart, D. L. King, Z. Demetrovics, A. L. M. Andrade, J.-S. Choi, H. Tavares, B. Almeida, D. C. Hodgins, A scoping review of the association between loot boxes, esports, skin betting, and token wagering with gambling and video gaming behaviors, *Journal of Behavioral Addictions* 12 (2023) 309–351.
- [26] R. T. Wood, R. Gupta, J. L. Derevensky, M. Griffiths, Video game playing and gambling in adolescents: Common risk factors, *Journal of Child & Adolescent Substance Abuse* 14 (2004) 77–100.
- [27] P. Delfabbro, D. King, C. Lambos, S. Puglies, Is video-game playing a risk factor for pathological gambling in australian adolescents?, *Journal of gambling studies* 25 (2009) 391–405.
- [28] J. Macey, J. Hamari, Investigating relationships between video gaming, spectating esports, and gambling, *Computers in Human Behavior* 80 (2018) 344–353.
- [29] D. Zendle, P. Cairns, Video game loot boxes are linked to problem gambling: Results of a large-scale survey, *PloS one* 13 (2018) e0206767.
- [30] D. Zendle, R. Meyer, H. Over, Adolescents and loot boxes: Links with problem gambling and motivations for purchase, *Royal Society open science* 6 (2019) 190049.
- [31] H. Molde, B. Holmøy, A. G. Merkesdal, T. Torsheim, R. A. Mentzoni, D. Hanns, D. Sagoe, S. Pallesen, Are video games a gateway to gambling? a longitudinal study based on a representative norwegian sample, *Journal of gambling studies* 35 (2019) 545–557.

- [32] G. A. Brooks, L. Clark, Associations between loot box use, problematic gaming and gambling, and gambling-related cognitions, *Addictive behaviors* 96 (2019) 26–34.
- [33] D. Zendle, Beyond loot boxes: A variety of gambling-like practices in video games are linked to both problem gambling and disordered gaming, *PeerJ* 8 (2020) e9466.
- [34] J. Macey, J. Hamari, esports, skins and loot boxes: Participants, practices and problematic behaviour associated with emergent forms of gambling, *New Media & Society* 21 (2019) 20–41.
- [35] D. Zendle, P. Cairns, Loot boxes are again linked to problem gambling: Results of a replication study, *PloS one* 14 (2019) e0213194.
- [36] D. Zendle, Problem gamblers spend less money when loot boxes are removed from a game: A before and after study of heroes of the storm, *PeerJ* 7 (2019) e7700.
- [37] W. Li, D. Mills, L. Nower, The relationship of loot box purchases to problem video gaming and problem gambling, *Addictive behaviors* 97 (2019) 27–34.
- [38] H. Wardle, The same or different? convergence of skin gambling and other gambling among children, *Journal of gambling studies* 35 (2019) 1109–1125.
- [39] D. L. King, A. M. Russell, P. H. Delfabbro, D. Polisena, Fortnite microtransaction spending was associated with peers' purchasing behaviors but not gaming disorder symptoms, *Addictive behaviors* 104 (2020) 106311.
- [40] S. Kristiansen, M. C. Severin, Loot box engagement and problem gambling among adolescent gamers: Findings from a national survey, *Addictive behaviors* 103 (2020) 106254.
- [41] M. Von Meduna, F. Steinmetz, L. Ante, J. Reynolds, I. Fiedler, Loot boxes are gambling-like elements in video games with harmful potential: Results from a large-scale population survey, *Technology in Society* 63 (2020) 101395.
- [42] D. Zendle, P. Cairns, H. Barnett, C. McCall, Paying for loot boxes is linked to problem gambling, regardless of specific features like cash-out and pay-to-win, *Computers in Human Behavior* 102 (2020) 181–191.
- [43] A. King, G. Wong-Padoongpatt, A. Barrita, D. T. Phung, T. Tong, Risk factors of problem gaming and gambling in us emerging adult non-students: The role of loot boxes, microtransactions, and risk-taking, *Issues in Mental Health Nursing* 41 (2020) 1063–1075.
- [44] A. Brady, G. Prentice, Are loot boxes addictive? analyzing participant's physiological arousal while opening a loot box, *Games and Culture* 16 (2021) 419–433.
- [45] S. Ide, M. Nakanishi, S. Yamasaki, K. Ikeda, S. Ando, M. Hiraiwa-Hasegawa, K. Kasai, A. Nishida, et al., Adolescent problem gaming and loot box purchasing in video games: cross-sectional observational study using population-based cohort data, *JMIR Serious Games* 9 (2021) e23886.
- [46] L. C. Hall, A. Drummond, J. D. Sauer, C. J. Ferguson, Effects of self-isolation and quarantine on loot box spending and excessive gaming—results of a natural experiment, *PeerJ* 9 (2021) e10705.
- [47] W. DeCamp, Loot boxes and gambling: Similarities and dissimilarities in risk and protective factors, *Journal of Gambling Studies* 37 (2021) 189–201.
- [48] H. Wardle, D. Zendle, Loot boxes, gambling, and problem gambling among young people: Results from a cross-sectional online survey, *Cyberpsychology, Behavior, and Social Networking* 24 (2021) 267–274.
- [49] C. J. Larche, K. Chini, C. Lee, M. J. Dixon, M. Fernandes, Rare loot box rewards trigger larger arousal and reward responses, and greater urge to open more loot boxes, *Journal of gambling studies* 37 (2021) 141–163.
- [50] M. Rockloff, A. M. Russell, N. Greer, L. Lole, N. Hing, M. Browne, Young people who purchase loot boxes are more likely to have gambling problems: An online survey of adolescents and young adults living in nsw australia, *Journal of Behavioral Addictions* 10 (2021) 35–41.
- [51] J. P. Zagal, S. Björk, C. Lewis, Dark patterns in the design of games, in: *Foundations of Digital Games 2013*, 2013.

De juego móvil a realidad virtual en cinta de correr: adaptación y evaluación de la experiencia de juego

Nàdia Montaña-Miranda^{1,*†}, Maria Beatriz Villar-Lopez^{1,†}, Carlos Marín-Lora^{1,*†},
Micaela Y. Martín^{1,†} y Miguel Chover^{1,†}

¹Instituto de Nuevas Tecnologías de la Imagen. Universitat Jaume I. Castellón - España. Avd. Vicente Sos Baynat S/N.

Resumen

Este estudio presenta la adaptación del videojuego activo Fitoon, originalmente desarrollado para dispositivos móviles, a un entorno de realidad virtual (VR) con locomoción real sobre cinta de correr. El objetivo principal fue evaluar la experiencia subjetiva de los usuarios en esta nueva versión utilizando el cuestionario Game Experience Questionnaire (GEQ), módulo in-game. La implementación se diseñó para conservar la naturaleza físicamente activa del juego, introduciendo una perspectiva en primera persona y control mediante rotación de la cabeza. Un total de 20 participantes interactuaron durante un minuto con el juego en VR y completaron el GEQ. Los resultados muestran puntuaciones elevadas en inmersión, afecto positivo y fluidez, y bajos niveles de tensión y afectos negativos. Aunque no se hallaron relaciones significativas entre las variables demográficas y las dimensiones evaluadas.

Palabras clave

Realidad virtual, Experiencia de juego, Cinta de correr, Interacción persona-computadora, Videojuegos activos

1. Introducción

La realidad virtual (VR) se ha consolidado como una herramienta clave en el ámbito del entretenimiento interactivo, al permitir el diseño de experiencias más inmersivas y dinámicas. Al mismo tiempo, los denominados exergames (videojuegos que implican actividad física por parte del jugador) han demostrado ser eficaces para fomentar el ejercicio mediante estrategias de gamificación [1, 2]. No obstante, integrar estos dos enfoques creando juegos activos dentro de entornos de VR, conlleva desafíos significativos en términos técnicos y de seguridad, especialmente cuando se incorpora locomoción real sobre una cinta de correr.

El presente estudio parte del videojuego para dispositivos móviles Fitoon, diseñado específicamente para jugarse sobre cinta de correr [3], el cual se ha demostrado que mejora significativamente la competencia, la inmersión y la fluidez de juego frente a otras aplicaciones para correr en cinta [4]. Su mecánica en dispositivo móvil se basa en la detección del movimiento de la cara del usuario, con la cual se controla a un personaje en tercera persona.

En esta investigación, se presenta una nueva versión de Fitoon adaptada al entorno de VR, con el objetivo de mantener una jugabilidad físicamente activa sobre cinta de correr. Para ello, se adaptó la mecánica original y se integró un sistema de control en primera persona, acorde con los estándares inmersivos propios de los juegos en VR. En esta nueva versión, el desplazamiento del personaje se produce en la dirección hacia la que el usuario rota ligeramente la cabeza mediante el visor, lo que permite conservar el mismo tipo de interacción física presente en la versión original.

Aunque correr en cinta con gafas de VR puede parecer inicialmente incómodo o poco atractivo, consideramos que esta combinación podría ofrecer una experiencia interesante desde el punto de vista lúdico e inmersivo. Por ello, el presente estudio parte de la inquietud por comprobar cómo es recibida

CEV 2025: IV Congreso Español de Videojuegos.

*Corresponding author.

†These authors contributed equally.

✉ montana@uji.es (N. Montaña-Miranda); mavillar@uji.es (M. B. Villar-Lopez); cmarin@uji.es (C. Marín-Lora); micmarti@uji.es (M. Y. Martín); choever@uji.es (M. Chover)

🆔 0009-0005-3227-0384 (N. Montaña-Miranda); 0009-0007-8211-3736 (M. B. Villar-Lopez); 0000-0003-1055-7657 (C. Marín-Lora); 0000-0001-6100-7538 (M. Y. Martín); 0000-0002-0525-7038 (M. Chover)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

esta modalidad por los usuarios. Más allá del análisis técnico de la adaptación de Fitoon a VR, nuestro objetivo principal fue explorar el grado de aceptación y disfrute que genera esta propuesta de locomoción real en entornos virtuales, evaluando su potencial como experiencia de juego activa e inmersiva.

En este sentido, con el propósito de evaluar la experiencia de juego en la nueva versión en VR de Fitoon, se llevó a cabo un estudio experimental centrado en las percepciones subjetivas de los usuarios. Los participantes completaron una sesión de juego en VR caminando sobre una cinta de correr a velocidad constante de 4 km/h durante un minuto, tras la cual se recogieron sus valoraciones mediante el módulo in-game del Game Experience Questionnaire (GEQ) [5], un instrumento ampliamente validado en entornos interactivos inmersivos, para medir las distintas dimensiones de la experiencia. Este cuestionario permite evaluar aspectos como la inmersión, las afecciones positivas y negativas, la competencia, el desafío, la tensión y la fluidez.

De un modo más específico, los objetivos de este estudio son documentar el proceso de adaptación del videojuego Fitoon desde una interfaz basada en dispositivo móvil a un entorno de VR, manteniendo su carácter físicamente activo mediante el uso de una cinta de correr; evaluar la experiencia subjetiva de juego en la versión de Fitoon para VR mediante el GEQ en su módulo in-game; y analizar las puntuaciones obtenidas en las dimensiones de inmersión, afecto positivo y negativo, competencia, desafío, tensión y fluidez, con el propósito de identificar fortalezas y posibles áreas de mejora en la experiencia de juego. Asimismo, se examina la relación entre dichas dimensiones y diversas variables demográficas de los participantes, como el género, la experiencia previa con realidad virtual o videojuegos, la frecuencia de juego y la percepción de mareo, con el fin de explorar posibles factores moduladores de la experiencia. Y en este sentido, se formulan las siguientes hipótesis:

- **H1:** Los participantes con experiencia previa en el uso de gafas de VR presentan mayores niveles de inmersión.
- **H2:** Quienes experimentan mareo durante la experiencia obtienen puntuaciones más elevadas en las dimensiones de Tensión y Afecciones Negativas.
- **H3:** A mayor frecuencia de juego, se reportan niveles más bajos de Tensión y más altos de Desafío.
- **H4:** El género del participante influye en la inmersión, siendo las mujeres quienes presentan puntuaciones más altas.

Este trabajo se organiza de la siguiente manera: la Sección 2 aborda el marco teórico y revisa el estado del arte en torno al uso de la VR en videojuegos activos, así como la evaluación de la experiencia subjetiva del usuario. La Sección 3 detalla la metodología seguida, incluyendo la descripción del videojuego original en su versión móvil y el proceso llevado a cabo para su adaptación a un entorno de VR. En la Sección 4 se describen los experimentos realizados, especificando el perfil de los participantes, las condiciones de uso, los instrumentos de recogida de datos y el análisis estadístico aplicado. La Sección 5 expone los resultados obtenidos, con especial atención a las puntuaciones del GEQ y a la influencia de diversas covariables demográficas y ofrece una discusión crítica de los resultados en función de las hipótesis iniciales, explorando el papel de variables individuales y señalando las principales limitaciones del estudio. Finalmente, la Sección 6 recoge las conclusiones generales y plantea posibles direcciones para investigaciones futuras.

2. Marco teórico

La integración de la VR con movimientos físicos como caminar o correr ha captado un interés creciente en áreas como el entretenimiento inmersivo, la rehabilitación motora y el entrenamiento físico-cognitivo. Esta modalidad ofrece experiencias altamente envolventes, pero también expone a los usuarios a síntomas de cinetosis (o "VR sickness"), los cuales suelen intensificarse cuando la locomoción implica movimiento corporal real [6, 7].

La cinetosis en VR se atribuye comúnmente a la asincronía entre señales visuales y vestibulares, así como a la latencia en la retroalimentación motora [8]. Entre las estrategias para mitigar estos efectos se encuentran la limitación del desplazamiento virtual [9], el uso de cintas autoadaptativas [10] o la

estabilización del entorno visual [11]. No obstante, estas soluciones técnicas pueden reducir la agencia del usuario (es decir, su capacidad percibida de influir activamente en el entorno virtual mediante sus acciones), disminuyendo la sensación de presencia [12]. Estudios recientes han mostrado que, a pesar de estos desafíos, los videojuegos activos en VR aumentan la percepción de disfrute y compromiso emocional, aunque a menudo a costa de una mayor fatiga cognitiva y menor retención de información [13, 14]. Esta dualidad plantea la necesidad de optimizar el equilibrio entre inmersión, confort físico y carga cognitiva.

El GEQ es uno de los instrumentos más utilizados para evaluar la experiencia subjetiva en videojuegos, incluyendo contextos de VR inmersiva y videojuegos activos [5, 15]. Este cuestionario permite medir dimensiones como inmersión, desafío, afecto negativo o positivo y flujo, aspectos cruciales en escenarios que exigen interacción sensoriomotora. Diversos estudios han validado la aplicabilidad del GEQ en situaciones de locomoción física y esfuerzo, destacando su capacidad para detectar diferencias en la calidad de la experiencia según el tipo de interfaz y modalidad de control [16, 17]. En este sentido, el Exergame Enjoyment Questionnaire (EEQ-G) [18] ha sido propuesto como una extensión del GEQ, adaptada a contextos de exergaming, incorporando indicadores como fatiga muscular y diversión corporal. Asimismo, la literatura enfatiza la utilidad de enfoques mixtos que integren medidas subjetivas y objetivas para abordar de forma más completa la complejidad emocional y fisiológica de estas experiencias [19, 20].

3. Metodología y adaptación del videojuego

3.1. Descripción del videojuego original para móvil

Como se mencionó en la Introducción, el videojuego para dispositivos móviles del que se parte en este estudio se llama Fitoon (ver Figura 1). Es una propuesta interactiva diseñada específicamente para utilizarse mientras se corre en una cinta, con el propósito de incentivar la actividad física mediante dinámicas de juego sencillas y accesibles. A diferencia de otros exergames, Fitoon no necesita sensores externos ni accesorios adicionales, ya que puede funcionar directamente en un dispositivo móvil equipado con cámara [3].

La mecánica principal consiste en una carrera de obstáculos que toma inspiración de juegos como Fall Guys o Stumble Guys. El progreso del avatar depende del movimiento físico real del usuario. El sistema de control utiliza tecnología de seguimiento facial (AR Face Tracking), que evalúa el movimiento vertical de la cabeza para estimar la cadencia de carrera, y emplea el vector normal de la cara para determinar la dirección del desplazamiento [21, 22, 23].

La versión original móvil de Fitoon se muestra desde una perspectiva en tercera persona, con una cámara elevada y situada detrás del avatar. Esta decisión se basa en criterios de coherencia con las normas de diseño más comunes en la plataforma, ya que los juegos para dispositivos móviles suelen optar por la tercera persona porque facilita la visualización tanto del entorno como del personaje [24].



Figura 1: Capturas de la ejecución del juego original en dispositivo móvil

3.2. Proceso de adaptación a VR

La versión para VR (ver Figura 2) fue adaptada a partir de la original y desarrollada en Unity 6000.0.47f1. En esta adaptación, se añadió un modelo 3D que reproducía a escala la cinta de correr real, además de integrar un sistema de calibración geométrica adicional al juego utilizando el paquete XR Hands (ver Figura 3). La calibración se efectuaba colocando las manos virtuales sobre la barra frontal de la cinta, lo que permitía sincronizar el espacio físico con el virtual. Así, durante la experiencia VR, los usuarios podían mantener una referencia constante con la cinta real gracias a la visualización en modo fantasma, es decir, mostrando la cinta como un objeto semitransparente.

Tal y como se comentaba en la Introducción, se buscó mantener la misma interacción física presente en la versión móvil, que consistía en girar la cabeza mientras se corre en la cinta. Por lo tanto, en la nueva versión para VR, el usuario se mueve dentro del juego en la dirección a la que esté mirando con el visor. Para esta versión, se emplea una perspectiva en primera persona, que simula la visión subjetiva desde los ojos del avatar. Esta elección se fundamenta en criterios de diseño propios de las experiencias inmersivas, donde la perspectiva en primera persona ha demostrado ser más efectiva para inducir sensación de presencia y realismo [25, 26], así como favorecer una mayor precisión en la interacción y una sensación más intensa de encarnación del usuario en el cuerpo virtual [24], contribuyendo así a optimizar la naturalidad y la inmersión durante la experiencia.

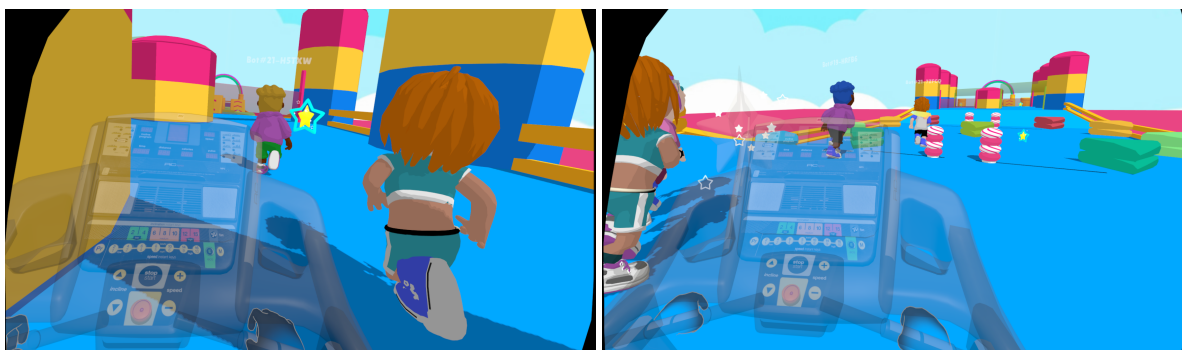


Figura 2: Capturas de la ejecución del juego en versión VR



Figura 3: Calibración para establecer la correspondencia entre la cinta de correr virtual y la real

4. Experimentos

4.1. Participantes y condiciones experimentales

Para definir la metodología empleada en este experimento, se utilizó un diseño donde los participantes entraban en un entorno de VR con una velocidad de locomoción propia de un ritmo de caminar (4km/h).

Cada participante completó una sesión de juego con una duración fija de un minuto, tras la cual rellenaba en una tableta el cuestionario GEQ correspondiente a la experiencia vivida. Por motivos de seguridad, se estableció una velocidad de 4km/h debido a que en pruebas piloto se observó que velocidades más altas de trote o carrera aumentaban el desplazamiento lateral sobre la cinta, incrementando así el riesgo de pérdida de equilibrio, especialmente en personas sin experiencia previa. Además, una velocidad superior podría dificultar la marcha segura dependiendo de la longitud de las piernas del usuario. Por el mismo motivo, durante todas las pruebas en VR se solicitó a los participantes que mantuvieran las manos apoyadas sobre la barra frontal de la cinta.

En cuanto a la captación de participantes, esta se llevó a cabo mediante carteles informativos distribuidos en zonas comunes de la institución de los autores. Estos carteles incluían un código QR que redirigía a un formulario de inscripción donde los voluntarios podían seleccionar una franja horaria entre las disponibles. Cada sesión experimental tuvo una duración aproximada de diez minutos, incluyendo la explicación inicial, la adaptación al entorno, la actividad sobre la cinta y el tiempo necesario para completar los cuestionarios. El estudio fue aprobado por el Comité de Ética de la Universitat Jaume I de Castellón.

En cuanto al material, se utilizó una cinta de correr BH RC09 G6180I, con una superficie útil de 155 × 55 cm, motor de 2.25 CV y capacidad para alcanzar velocidades entre 1 y 22 km/h. Para el entorno de VR, se emplearon unas gafas Meta Quest Pro conectadas mediante Meta Link a un ordenador HP Omen Transcend (pantalla de 16", procesador Intel Core i7-13700HX, 32 GB de RAM), desde el cual se ejecutaba la aplicación desarrollada en Unity.

4.2. Cuestionarios

Dado que el objetivo principal de este estudio fue analizar la experiencia subjetiva del usuario en un entorno de VR, se eligió como herramienta de evaluación la modalidad in-game del GEQ. Aunque existen instrumentos específicos para videojuegos activos, como el EEQ-G, se descartaron por centrarse en dimensiones relacionadas con la actividad física, como la fatiga o la satisfacción motora, que no eran relevantes para los fines de esta investigación. En cambio, se optó por el GEQ estándar por tratarse de un cuestionario ampliamente validado en contextos interactivos generales, lo que lo convierte en una opción más adecuada para el tipo de análisis propuesto.

Para la recogida de datos relacionados con la experiencia del usuario, se empleó un formulario estructurado en dos bloques diferenciados. El Bloque 1, cumplimentado antes del experimento, recopilaba datos demográficos y de perfil del participante, incluyendo edad, género, altura, peso, lateralidad, frecuencia de juego, experiencia previa con videojuegos activos, familiaridad con la VR, historial de mareo en entornos VR, nivel de actividad física y uso anterior de cintas de correr. Mientras que el Bloque 2, administrado tras la sesión de juego, contenía el módulo in-game del GEQ, que evalúa siete dimensiones de la experiencia subjetiva: inmersión, competencia, desafío, afecciones positivas, afecciones negativas, tensión y fluidez. Este módulo incluye 14 ítems valorados en una escala Likert de 0 (nada) a 4 (mucho), organizados según las dimensiones propuestas por IJsselsteijn et al. (2013) (ver Tabla 1). Además, se añadió un campo adicional para que los participantes indicaran posibles síntomas de incomodidad o mareo al finalizar el cuestionario.

4.3. Procedimiento de análisis estadístico

El tratamiento y análisis de los datos se realizaron mediante un entorno de trabajo en Python (versión 3.11), empleando bibliotecas especializadas para el análisis estadístico y la visualización de datos. Se utilizó la biblioteca *pandas* para la carga, limpieza y estructuración de los datos, así como para el cálculo de medias y desviaciones estándar por componente del cuestionario. Los valores no numéricos o ausentes fueron tratados mediante conversión explícita a valores numéricos con control de errores, y las puntuaciones por componente se obtuvieron como la media de los ítems correspondientes.

Para el análisis inferencial se aplicaron distintas pruebas en función de la naturaleza de las variables. Se utilizaron *pruebas t de Student* para muestras independientes en aquellos casos con variables categóricas

Tabla 1

Dimensiones e ítems del cuestionario In-game GEQ

Dimensión	Ítems
Inmersión	1. Me interesó la trama del juego 4. Me pareció impresionante
Competencia	2. Me sentí exitoso 9. Me sentí hábil
Fluidez	5. Me olvidé de todo lo que me rodeaba 10. Me sentí completamente absorbido
Tensión	6. Me sentí frustrado 8. Me sentí irritable
Desafío	12. Me sentí desafiado 13. Tuve que esforzarme mucho
Afecciones Negativas	3. Me sentí aburrido 7. Me pareció pesado
Afecciones Positivas	11. Me sentí satisfecho 14. Me sentí bien

dicotómicas, mientras que para variables ordinales se empleó la *correlación de Spearman*. Todas las pruebas se implementaron mediante la biblioteca *scipy.stats*, respetando un umbral de significación estadística de $\alpha = 0.05$.

Adicionalmente, se generaron visualizaciones con *matplotlib* y *seaborn*, empleando gráficos de barras para los resultados descriptivos y diagramas tipo *violin plot* para ilustrar las distribuciones en los contrastes de hipótesis.

Los resultados de cada contraste, incluyendo los valores p, se almacenaron en un archivo CSV para su posterior análisis e inclusión en el manuscrito. Las visualizaciones fueron etiquetadas y organizadas sistemáticamente para facilitar su interpretación.

5. Resultados y discusión

La muestra final estuvo compuesta por 20 participantes, de los cuales el 55 % se identificaron como hombres y el 45 % como mujeres. La altura media fue de 172,4 cm (DE = 10,6). En cuanto a la lateralidad, el 95 % eran diestros y el 5 % zurdos.

En relación con sus hábitos, el 55 % jugaba a videojuegos a diario, el 25 % ocasionalmente, el 10 % semanalmente y el 10 % nunca. La mayoría había tenido experiencia previa con videojuegos activos (95 %) y con dispositivos de VR (95 %). El 70 % había usado previamente una cinta de correr. En cuanto al mareo en experiencias de VR, el 55 % lo había experimentado alguna vez, el 40 % no, y el 5 % se mostró indeciso. Respecto a la práctica de ejercicio físico, el 35 % lo realizaba entre tres y cuatro veces por semana, el 30 % una o dos veces, el 10 % cinco o más veces, y el 25 % no lo practicaba.

Tabla 2

Media y desviación estándar por componente del GEQ

Componente	Media	Desviación estándar
Inmersión	3.26	0.74
Competencia	3.00	0.77
Fluidez	3.12	0.71
Tensión	0.36	0.69
Desafío	2.19	0.89
Afecciones Negativas	0.38	0.52
Afecciones Positivas	3.21	0.75

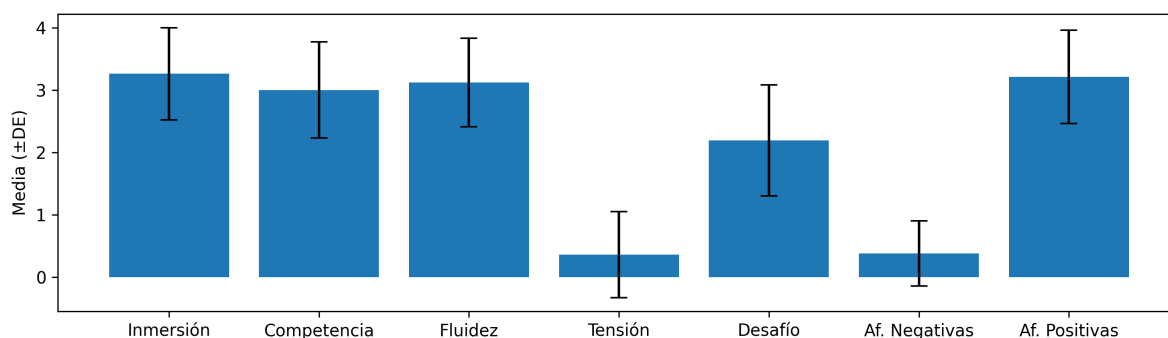


Figura 4: Media y desviación estándar por componente del GEQ

Los resultados del cuestionario GEQ (Figura 4 y Tabla 2) reflejan una experiencia de juego predominantemente positiva. Las puntuaciones más altas se obtuvieron en Inmersión ($M = 3.26$), Afecciones Positivas ($M = 3.21$) y Fluidez ($M = 3.12$), mientras que las dimensiones menos puntuadas fueron Tensión ($M = 0.36$) y Afecciones Negativas ($M = 0.38$). Esto indica que el entorno VR fue percibido como envolvente, placentero y poco estresante.

Se contrastaron cuatro hipótesis utilizando pruebas t para muestras independientes y correlaciones de Spearman. Los resultados se resumen en la Tabla 3, y las gráficas correspondientes se presentan en las Figuras 5, 6, 7, 8, 9 y 10.

- **H1:** No se encontraron diferencias significativas en inmersión entre participantes con o sin experiencia previa con gafas de VR ($p = 0,2088$).
- **H2:** No se observaron diferencias en Tensión ($p = 0,6886$) ni Afecciones Negativas ($p = 0,7931$) en función de haber experimentado mareo.
- **H3:** La frecuencia de juego no mostró correlaciones significativas con Tensión ($p = 0,8626$) ni Desafío ($p = 0,1976$), aunque esta última se aproxima al umbral de significación.
- **H4:** Las mujeres mostraron una ligera tendencia a puntuar más alto en Inmersión que los hombres, sin alcanzar significación ($p = 0,1196$).

Tabla 3

Resultados de contraste de hipótesis sobre los componentes del GEQ

Hipótesis	Comparación	Componente	p-valor
H1	Experiencia VR	Inmersión	0.2088
H2	Mareo	Tensión	0.6886
H2	Mareo	Afecciones Negativas	0.7931
H3	Frecuencia de juego	Tensión	0.8626
H3	Frecuencia de juego	Desafío	0.1976
H4	Género	Inmersión	0.1196

Desde una perspectiva teórica, la conjunción de alta inmersión, fluidez y afecto positivo con baja tensión y afecto negativo se alinea con el modelo de *flow* de Csikszentmihalyi [27], que describe un estado de concentración profunda y disfrute. La configuración del entorno —que incluye locomoción física sobre cinta y perspectiva en primera persona— pudo favorecer la sensación de presencia descrita en el modelo de Slater y Wilbur [28], al mantener la coherencia entre movimientos reales y respuestas virtuales.

Por otra parte, los niveles reducidos de tensión pueden interpretarse en el marco de los *exergames*, que atenúan el esfuerzo físico mediante una dinámica lúdica e inmersiva [29, 30]. Este aspecto resulta clave para valorar el potencial de la experiencia como recurso motivacional en contextos de promoción de actividad física.

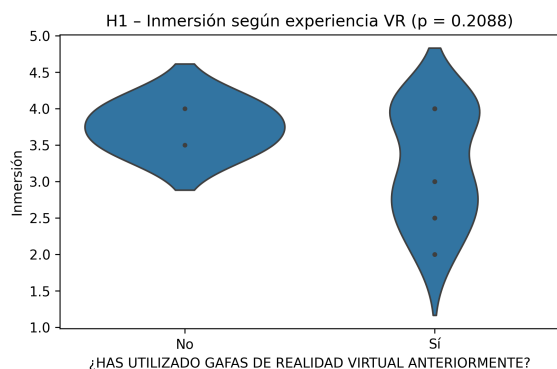


Figura 5: H1: Inmersión según experiencia VR

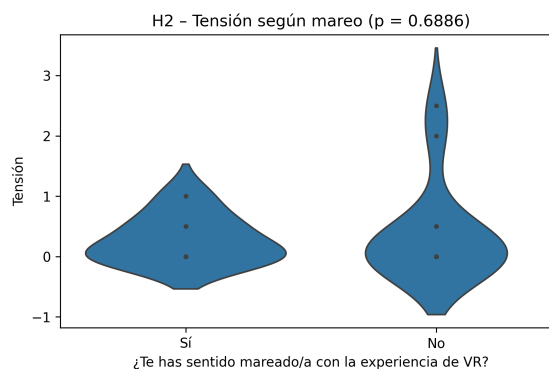


Figura 6: (H2: Tensión según mareo

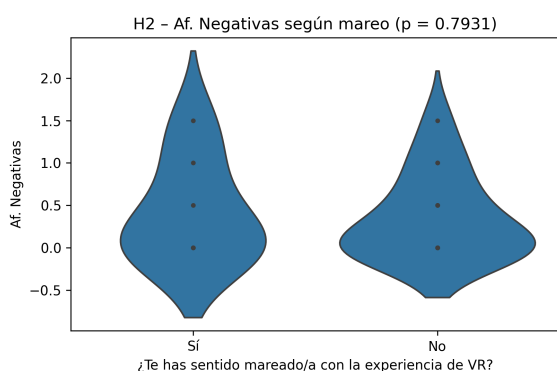


Figura 7: H2: Afecciones negativas según mareo

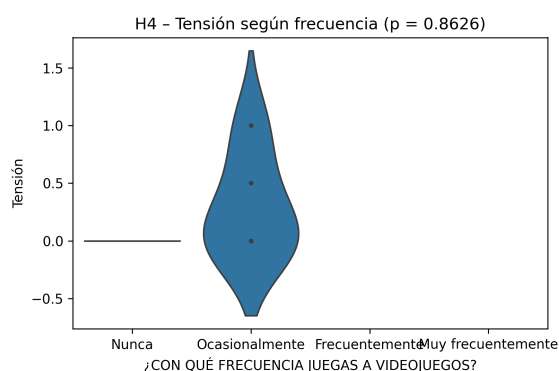


Figura 8: H3: Frecuencia de juego – Tensión

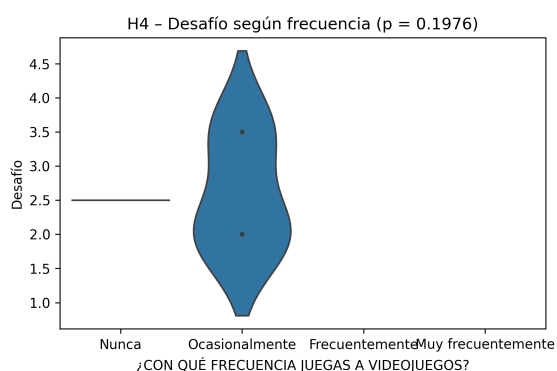


Figura 9: H3: Frecuencia de juego – Desafío

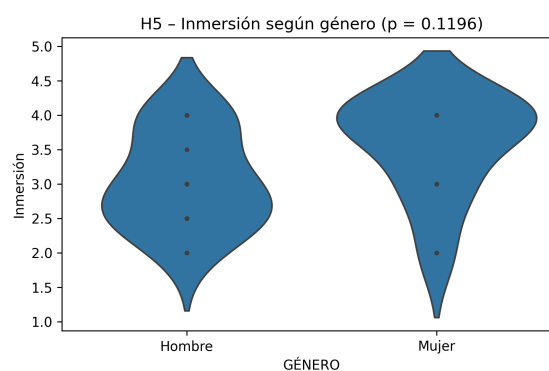


Figura 10: H4: Inmersión según género

Finalmente, si bien no se encontraron asociaciones estadísticamente significativas entre variables demográficas y los componentes del GEQ, esta falta de efectos debe interpretarse con cautela, ya que la muestra es reducida y homogénea. En futuras investigaciones sería conveniente ampliar la muestra e incorporar medidas más sensibles relacionadas con motivación, habilidades previas y tolerancia fisiológica a la VR.

Para concluir la discusión de los resultados obtenidos es necesario presentar algunas limitaciones que deben tenerse en cuenta al interpretar los resultados. En primer lugar, el tamaño muestral reducido y la distribución desigual de algunas variables demográficas limitan la generalización de los hallazgos y la potencia estadística para detectar efectos sutiles. Además, la elevada proporción de participantes con experiencia previa en VR o videojuegos activos (ambas 95 %) dificulta la comparación entre subgrupos.

Por último, el uso exclusivo de medidas subjetivas podría complementarse en futuros trabajos con indicadores fisiológicos o de rendimiento objetivo que permitan una evaluación más completa de la experiencia de juego.

6. Conclusiones y trabajo futuro

El presente estudio tuvo como objetivo analizar la experiencia subjetiva de juego tras la adaptación del videojuego Fitoon a un entorno de VR con locomoción real sobre cinta de correr. Partiendo de una versión previa para dispositivos móviles, se implementó una propuesta inmersiva en primera persona que preserva la naturaleza físicamente activa del juego. Se evaluaron aspectos clave de la experiencia mediante el módulo in-game del cuestionario GEQ, con especial interés en dimensiones como inmersión, competencia, tensión, fluidez o afecto.

Los resultados indican que la versión en VR ofrece una experiencia de juego positiva, fluida e inmersiva, caracterizada por altos niveles de afecto positivo y bajos niveles de tensión y afecto negativo. Aunque no se observaron efectos estadísticamente significativos de las variables demográficas sobre las dimensiones del GEQ, se identificaron ciertas tendencias que podrían alcanzar significación con una muestra de mayor tamaño.

La ausencia de diferencias relevantes en función del género, la experiencia previa en VR o la frecuencia de juego sugiere que la propuesta es percibida de forma homogénea y accesible por parte de perfiles diversos. Sin embargo, la escasa representación de algunos subgrupos condiciona la capacidad para contrastar con robustez determinadas hipótesis.

De cara a futuros trabajos, se plantea como prioridad la ampliación de la muestra para aumentar la potencia estadística de los análisis. Asimismo, se contempla una comparación directa entre la versión en VR y la versión tradicional en pantalla, lo que permitiría explorar las diferencias en términos de inmersión, esfuerzo percibido y satisfacción. Se prevé también el desarrollo de una modalidad multijugador local que permita a varios participantes compartir la experiencia en cintas paralelas, facilitando así el estudio de componentes sociales del juego. Finalmente, se plantea la posibilidad de extender el sistema a otros entornos de ejercicio físico, como bicicletas estáticas, con el objetivo de diversificar sus aplicaciones y evaluar la coherencia de la experiencia en distintos formatos de locomoción asistida.

Agradecimientos

Este trabajo se ha desarrollado en el marco del proyecto de investigación PID2023-149976OB-C21, financiado por MCIN/AEI/10.13039/501100011033/FEDER, UE.

Referencias

- [1] C. Khundam, F. Noël, A study of physical fitness and enjoyment on virtual running for exergames, *International Journal of Computer Games Technology* 2021 (2021) 6668280. doi:10.1155/2021/6668280.
- [2] A. Quintas, J. C. Bustamante, Effects of gamified didactic with exergames on the psychological variables associated with promoting physical exercise, *Physical Education and Sport Pedagogy* 28 (2023) 467–481. doi:10.1080/17408989.2021.1991905.
- [3] L. García Rytman, H. J. Ruiz Marco, M. Y. Martín, C. Marín-Lora, M. Chover, Development of a treadmill video game: Technology and mechanics, in: *Proceedings of Congreso Español de Videojuegos (CEV 2023)*, 2023.
- [4] M. Y. Martín, C. Marín-Lora, M. Chover, A comparative analysis of game experience in treadmill running applications, *Entertainment Computing* 52 (2025) 100888. doi:10.1016/j.entcom.2024.100888.

- [5] W. A. IJsselsteijn, Y. A. W. de Kort, K. Poels, The Game Experience Questionnaire, Technical Report, Technische Universiteit Eindhoven, 2013. URL: <https://research.tue.nl/en/publications/the-game-experience-questionnaire>.
- [6] S. Weech, S. Kenny, M. Barnett-Cowan, Presence and cybersickness in virtual reality are negatively related: A review, *Frontiers in Psychology* 10 (2019) 158. URL: <https://doi.org/10.3389/fpsyg.2019.00158>. doi:10.3389/fpsyg.2019.00158.
- [7] S. Vlahovic, M. Zajc, S. Velev, A survey of challenges and methods for quality of experience assessment of interactive vr applications, *Journal on Multimodal User Interfaces* 16 (2022) 341–359. URL: <https://doi.org/10.1007/s12193-022-00388-0>. doi:10.1007/s12193-022-00388-0.
- [8] J. L. Campos, H. H. Bühlhoff, Multimodal integration during self-motion in virtual reality, in: *Encyclopedia of the Sciences of Learning*, 2012. URL: <https://europepmc.org/article/nbk/nbk92853>.
- [9] W. A. Powell, Virtually walking: Factors influencing walking and perception of walking in treadmill-mediated VR, Ph.D. thesis, University of North Carolina, 2011. Retrieved from <https://core.ac.uk/download/pdf/40013582.pdf>.
- [10] E. H. Sinitski, et al., Fixed and self-paced treadmill walking for able-bodied and amputees in multi-terrain vr, *Gait & Posture* 41 (2015) 757–763. doi:10.1016/j.gaitpost.2014.12.016.
- [11] A. S. García, R. J. Martínez, P. Nieto, Evaluation of vr locomotion techniques in urban spaces, *Multimedia Tools and Applications* (2024). URL: <https://doi.org/10.1007/s11042-024-20487-2>. doi:10.1007/s11042-024-20487-2.
- [12] J. Guerra-Armas, S. Fernández, D. Mendoza, Role of immersive virtual reality in motor behaviour decision-making in chronic pain patients, *Brain Sciences* 13 (2023) 617. URL: <https://doi.org/10.3390/brainsci13040617>. doi:10.3390/brainsci13040617.
- [13] J. Roettl, R. Terlutter, The same video game in 2d, 3d or vr—how does technology impact game evaluation and brand placement effectiveness?, *PLOS ONE* 13 (2018) e0200724. doi:10.1371/journal.pone.0200724.
- [14] C. Lu, R. Al-Juaid, M. Al-Amri, Gait stability in inclined self-paced vr treadmill walking, *JMIR Formative Research* 7 (2023) e42769. doi:10.2196/42769.
- [15] C. Boletsis, J. E. Cedergren, Vr locomotion in the new era of virtual reality: An empirical comparison of prevalent techniques, *Advances in Human-Computer Interaction* 2019 (2019) Article ID 7420781. URL: <https://doi.org/10.1155/2019/7420781>. doi:10.1155/2019/7420781.
- [16] J. Frommel, S. Sonntag, M. Weber, Effects of controller-based locomotion on player experience in a virtual reality exploration game, in: *FDG '17: Proceedings of the 12th International Conference on the Foundations of Digital Games*, 2017. URL: <https://doi.org/10.1145/3102071.3102082>. doi:10.1145/3102071.3102082.
- [17] A. Somrak, M. Pogačnik, J. Guna, Suitability and comparison of questionnaires assessing virtual reality-induced symptoms and effects and user experience in virtual environments, *Sensors* 21 (2021) 1185. URL: <https://doi.org/10.3390/s21041185>. doi:10.3390/s21041185.
- [18] P. Manser, S. Huber, J. Seinsche, E. D. de Bruin, Development and initial validation of the german version of the exergame enjoyment questionnaire (eeq-g), *PLOS ONE* 18 (2023) e0286556. doi:10.1371/journal.pone.0286556.
- [19] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, A. Walton, Measuring and defining the experience of immersion in games, *International Journal of Human-Computer Studies* 66 (2008) 641–661. URL: <https://doi.org/10.1016/j.ijhcs.2008.04.004>. doi:10.1016/j.ijhcs.2008.04.004.
- [20] C. A. Godfrey, J. F. Oody, S. A. Conger, Active virtual reality games: Comparing energy expenditure, game experience, and cybersickness to traditional gaming and exercise in youth aged 8–12, *Games for Health Journal* (2025). URL: <https://doi.org/10.1089/g4h.2024.0034>. doi:10.1089/g4h.2024.0034.
- [21] C. Marin-Lora, M. Martin, M. Chover, A face-tracking method for estimating cadence on treadmills, in: *International Conference on Ubiquitous Computing and Ambient Intelligence*, Springer International Publishing, Cham, 2022, pp. 949–954.
- [22] C. Marín-Lora, M. Chover, M. Y. Martín, L. García-Rytman, Comparative study of interaction methods for mobile gaming while running on a treadmill, *Computers & Graphics* 117 (2023)

- 164–171. doi:10.1016/j.cag.2023.10.020.
- [23] C. Marín-Lora, M. Chover, M. Y. Martín, L. García-Rytman, Creating a treadmill running video game with smartwatch interaction, *Multimedia Tools and Applications* 83 (2024) 57709–57729. doi:10.1007/s11042-023-17752-1.
 - [24] G. Gorisse, O. Christmann, E. A. Amato, S. Richir, First-and third-person perspectives in immersive virtual environments: presence and performance analysis of embodied users, *Frontiers in Robotics and AI* 4 (2017) 33. URL: <https://doi.org/10.3389/frobt.2017.00033>. doi:10.3389/frobt.2017.00033.
 - [25] M. Slater, B. Spanlang, M. V. Sanchez-Vives, O. Blanke, First person experience of body transfer in virtual reality, *PLoS ONE* 5 (2009) e10564. URL: <https://doi.org/10.1371/journal.pone.0010564>. doi:10.1371/journal.pone.0010564.
 - [26] J. J. Cummings, J. N. Bailenson, How immersive is enough? a meta-analysis of the effect of immersive technology on user presence, *Media Psychology* 19 (2015) 272–309. URL: <https://doi.org/10.1080/15213269.2015.1015740>. doi:10.1080/15213269.2015.1015740.
 - [27] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*, Harper & Row, New York, 1990.
 - [28] M. Slater, S. Wilbur, A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments, *Presence: Teleoperators and Virtual Environments* 6 (1997) 603–616. doi:10.1162/pres.1997.6.6.603.
 - [29] Y. Oh, S. Yang, Defining exergames & exergaming, in: *Proceedings of Meaningful Play 2010*, East Lansing, MI, 2010, pp. 1–17.
 - [30] W. Peng, J.-H. Lin, J. Crouse, Is playing exergames really exercising? a meta-analysis of energy expenditure in active video games, *Cyberpsychology, Behavior, and Social Networking* 14 (2011) 681–688. doi:10.1089/cyber.2010.0578.

Large Language Models for Dynamically Reacting to Unexpected Video Game Interactions

David Rivera^{1,*†}, Roi Quintas^{1,†}, Miguel Ramírez^{1,†}, Claudia Zarzuela^{1,†}, Raúl Pérez^{1,†}
and Carlos León^{1,†}

¹*Department of Software Engineering and Artificial Intelligence
Computer Science Faculty
Universidad Complutense de Madrid
28040 Madrid, Spain*

Abstract

Video game interaction design often limits what players can do in a virtual environment. Designing complex systems that allow players to interact freely is usually extremely costly. This work explores the integration of large language models (LLMs) in video games to generate reactive environments and narratives based on player actions. The main goal is to design and implement a system capable of interpreting interactions with objects or non-player characters (NPCs) and generating structured responses that dynamically modify the game world, thus lowering production costs and increasing the dynamic capabilities of the video game. An experiment to evaluate both the coherence of the model's responses and the immersion experience was conducted. The obtained results indicate that players enjoy the dynamically-adapted video game.

Keywords

large language models, video games, game design, artificial intelligence

1. Introduction

Generating content for video games is a costly process. Providing covering for all plausible player actions requires building a huge set of possibilities, and even tackling such a task would probably leave many players behind their ideal gameplay. Typically, games constraint what can be done in order to provide a manageable content size and ensuring a acceptable level of coherence and quality. Creating a sense that every player action has meaningful consequences is known to be costly and time-consuming, as seen in games like *Baldur's Gate 3*, which achieved high interactivity through years of development by a large team. Such a level of complexity is unrealistic for most studios, making LLMs an appealing alternative for achieving similar responsiveness more efficiently.

Generative Artificial Intelligence (GAI) arises as a potential solution to leverage the scope of the game worlds. Generative Artificial Intelligence has gained global relevance since the public release of ChatGPT, sparking widespread interest and a surge of research focused on testing the capabilities of large language models (LLMs). Major technological companies have rapidly integrated generative AI into popular applications, offering virtual assistants capable of

IV Congreso Español de Videojuegos, September 18-19, 2025, Granada Spain

*Corresponding author.

**This publication is part of the R&D&I projects DARK NITE, PID2023-146308OB-I00, funded by MICI-U/AEI/10.13039/501100011033/ and ERDF, EU “A way of making Europe”; ADARVE (SUBV20/2021), funded by the Spanish Council of Nuclear Security; CANTOR (PID2019-108927RB-I00), funded by the Spanish Ministry of Science and Innovation; and EA-DIGIFOLK: An European and Ibero-American approach for the digital collection, analysis and dissemination of folk music project (grant agreement no. 101086338), funded by the European Commission.

†These authors contributed equally.

✉ davirive@ucm.es (D. Rivera); roiquint@ucm.es (R. Quintas); migram05@ucm.es (M. Ramírez); czarzu01@ucm.es (C. Zarzuela); raulpe07@ucm.es (R. Pérez); cleon@ucm.es (C. León)

ORCID 0009-0000-8850-3136 (D. Rivera); 0000-0002-6768-1766 (C. León)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

handling general and domain-specific queries. The entertainment industry has also embraced this technology. Games like *Suck Up!* and *Cygnus Enterprise* use LLMs to drive dynamic, emergent interactions with non-playable characters (NPCs), allowing players to issue natural language commands or engage in complex roleplay. These examples illustrate how LLMs can provide players with a sense of freedom and consequence in gameplay. Currently, most video game applications of LLMs center around real-time dialogue generation. However, their broader potential lies in processing large volumes of data and understanding context, which raises the question of whether these models can be applied to other aspects of game development.

Given this context, LLMs can be powerful tools for enhancing reactivity. Typically, reactivity in games is confined to dialogue, offering little visual feedback to reinforce the player's impact on the world—especially in 2D or 3D environments where visual cues are key. Interactions are usually limited to predefined behaviours, restricting player expression and agency. While text-based games can integrate LLMs more easily, visually immersive games require the generated content to be structured and organized. This presents a new technical challenge, but also an exciting opportunity: by combining player creativity with LLMs' generative power, developers can create dynamic responses that sustain engagement, attention, and immersion.

Against this context, we set out to research the plausibility of the inclusion of artificial intelligence models to enable automatic reactivity in video games, modifying the world and its non-player characters and the human player acts freely. By properly adapting the environment, the perceived freedom is expected to be higher, while keeping coherence and fun. To test this, we have developed a system that adapts a prototype video game through interaction with a LLM. We analyzed what type of interactions can be interpreted by the model, how it responds to them and where in the system it should be integrated to maximize its usefulness. This process involved defining specific tasks to be performed by the LLM, evaluating its suitability to used context, and checking whether made decisions during design and implementation have produced expected results. The output of the generative system is fed to a game engine, which parses the modifications provided by the LLM and modifies the world accordingly. Precision, contextual coherence and absence of hallucination were evaluated, as well as the impact of these aspects in user experience. The user subjective perception of the experience, their immersion level and the level general satisfaction was also evaluated.

2. Previous work

Beyond gaming, LLMs have proven fundamental in research across disciplines. Since the publication of *Attention Is All You Need* [1], which improved automatic translation methods, LLMs have been explored in fields like medicine [2] and education [3], showing their versatility and effectiveness across a wide range of domains. Regarding video games, recent research has increasingly focused on leveraging large language models in video games to generate dynamic, context-aware NPC dialogue. Peng et al. [4] demonstrated that text-based generative models can enhance immersion in interactive narratives, although high latency hindered user experience. Wu et al. [5] proposed SPRING, where LLMs act as players, selecting actions through structured reasoning graphs. In contrast, Kaiya et al. [6] developed LyfeAgent, a modular multi-component architecture (including memory, action, and sensory modules) to reduce LLM reliance and support agent reactivity in 3D environments, albeit with limited physical interaction. Hu et al. [7] reviewed state-of-the-art approaches, emphasizing three core components for LLM-driven agents: memory (short- and long-term), reasoning (e.g., Chain of Thought), and I/O systems for environmental perception and action execution. Csepregi [8] explored the effects of contextual awareness in NPCs, showing that NPCs aware of both player speech and actions produce more engaging interactions. Additionally, Hu et al. [7] showed that GPT-4 could autonomously generate coherent game levels and rules using VGDL [9], unlike smaller models. When talking about adversaries behaviour driven by AI models, Xu et al. [10] developed PORTAL, a framework

thar demonstrated LLMs capability of reducing extensive behaviour trees training time whilst delivering excellent results in FPS games. Following this topic, Wang et al. [11] also developed an interesting framework for decision-making LLMs that simulated opponents in a long-term strategy video game, where heavy planning is required. Another interesting use of LLMs is captured by Taesiri et Bezemer [12], who successfully trained a LLM model to analyze video game screenshots in order to find possible glitches and enhance QA processes. These studies collectively illustrate the growing versatility of LLMs in gaming, from dialogue and world generation to autonomous gameplay, as also evidenced by real-world applications such as the Skyrim mod Mantella, which enables immersive, AI-driven NPC interactions.

3. World representation for LLM-based interaction

The main goal of this research is to use a LLM as backend for reacting to unexpected player interactions. Lots of objects, meshes, textures, sounds and behaviours compose the environment of a video game. If represented in natural language, the output of the LLM cannot be directly parsed. This means that the LLM producing content and the game engine consuming it must agree on a structured model. This model must provide enough context for the LLM to understand the simulation it is in and what role it is playing, as well as a response format that it must follow. To achieve this, limitations to the world's structure and interactions were established.

The developed application takes place in a three-dimensional space, where the player sees the action in a first person perspective. Both these design choices were taken into account when designing the interactions in the environment. Several interactions were defined to limit and understand game-player communication: **movement** allows them to walk, jump, stop, or turn in all directions, though this is constrained by physical collisions and gravity. **Object manipulation** and combination enables the player to use one object on another object or NPC in the environment, triggering dynamic reactions that drive narrative progression. **NPC communication** allows for verbal interactions with non-player characters.

Although this system may seem too simple when compared to the current video game interaction's complexity, it allows interaction with all objects and NPCs within the environment, also making it possible to combine them as the user desires. Below, the defined types of interactions are far more detailed, with movement being the only exception, which is not the focus of this work. In a simulation based on emergent logic, all elements must be usable, combinable, or transformable as part of problem-solving processes. Unlike a traditional video game, this type of simulation does not rely on a fixed set of predefined interactions with predetermined outcomes. Instead, it allows the LLM to determine the consequences of interactions. This approach introduces several the challenges of definin how can interactive elements be identified, how can the nature of possible interactions be defined and how to ensure the consequences of interactions remain coherent within the environment. The proposed system is based on a simple interaction mechanic involving the reactive elements of the environment. This mechanic is divided into two types: interaction with an element—either an interactive object or a NPC—using the player's hand; and interaction with an element while holding another object. The player can perform several interactions with objects and NPCs. **Object pickup** allows the player to pick up objects labeled as both interactive and grabbable—since not all objects are meant to be picked up—by pointing at them and clicking the left mouse button. **Release** occurs when the player is already holding an object; a left-click will release it. **Interaction with elements** lets the player interact with an object marked as interactive or with an NPC by pointing at it and clicking the right mouse button. This triggers the generation of a structured message containing the necessary information, which is then sent to the language model to determine the outcome.

Each interaction is formalized as a structured message containing the following key components: **Initiator**, which is the entity that initiates the interaction. This can be **the player** using their empty hand; **the object held by the player**, if the interaction is performed using it—in

which case, the message includes the object's name and an updated description; or **an NPC autonomously deciding to interact** with another object in the scene besides the player. The second component is the **Receiver**, which is the entity that receives the interaction. This can be **an interactive object**, represented by its name and updated description, or **an NPC**, accompanied by its name, a brief description of its role within the simulation, and its interaction history.

Information about both the initiator and the receiver is always included, although their interpretative roles within the model depend on the type of interaction. There are two interaction types: automatic interactions, generated by the system when the player interacts with an object or NPC using the basic controls (e.g., right-clicking without specifying an action); and manual interactions, defined explicitly by the player through free-text input. In this case, the system captures not only the context of the interaction but also the specific verb or intended action provided by the player, such as “burn”, “hit”, “unlock”, or “place”.

Every interactive object or NPC within the environment follows a standardized structure composed of the following elements:

- **Name:** A unique identifier assigned to the object within the system.
- **Description:** A variable text field that represents the object's perceptible state.
- **Mesh:** The 3D visual representation used in the graphics engine.
- **State Machine:** A system that enables transitions between predefined states of an object (e.g., closed → open). While all objects are equipped with a default state machine, not all actively use it during the simulation. The activation of the state machine depends on a boolean parameter associated with the object.
- **Additional Components (if applicable):** Extended information such as physical properties, custom interactions, or narrative attributes.

To maintain internal consistency, two main categories of objects are defined. The first are *essential objects*, which are structural elements of the environment that play a critical role in the game's logic. They are predefined and controlled via a state machine. The language model may only select which state transition is triggered and update the object's description based on player interactions. It cannot create, destroy, or directly alter these objects unless such changes are explicitly included within the object's state machine transitions. The second category are *non-essential objects*: These are dynamic elements that do not initially serve a critical function within the game logic. Such objects can be generated by the language model during the simulation and may be instantiated, modified, or removed freely. When an object is created or altered, new components may also be added, or existing component data modified. Their purpose is to enrich the emergent narrative or respond to the player's creative interactions without compromising system stability. To prevent inconsistencies and ensure clear communication with the language model, several validation rules have been established:

- **Uniqueness of Interactive Elements:** Each element must have a unique identifier. Duplicate names are not permitted, as they could introduce ambiguity during processing.
- **Object Alteration Rules:**
 - An object may only be **created** if it is entirely new or derived (e.g., a fragment, liquid, etc.).
 - An object may be **modified** if its state changes (e.g., damaged, dirty, clean, etc.).
 - An object may only be **destroyed** if it ceases to exist entirely, becomes hidden, or is rendered permanently useless or unusable.
- **Concurrency Control:** Following an interaction, the involved elements are temporarily locked until a response is received from the model. This prevents their use in simultaneous interactions and avoids synchronization errors.

- **User Feedback:** To keep the player informed and reinforce system transparency, every interactive element in the environment—whether an object or NPC—is associated with a visual interface that provides persistent contextual information and real-time dynamic cues.

The interaction mechanics described thus far only support interactions between two entities at a time—whether between NPCs, player and NPC, an NPC and an object, or two objects within the scene. However, certain scenarios require simultaneous responses from multiple entities, such as when an explosion occurs and all nearby NPCs must perceive and react accordingly. Another limitation of the current system is that the language model can only respond to environmental changes via direct interactions. Yet, from a gameplay perspective, it is beneficial for the model to generate reactions to in-game events autonomously. For example, it would be desirable for an NPC to react based on sensory input—such as visual detection of a threat—by triggering a prompt to the model without requiring explicit player action.

To address these limitations, **a signal system is proposed**. A *signal* is an abstract object that can be generated and dispatched from any part of the codebase. This not only enables more complex interactions but also improves efficiency, particularly when using locally run language models, by reducing the need to process multiple sequential prompts. It consists of two primary components:

- **Verification:** Ensures through code-level checks that the signal is valid and executable in the current context. If the conditions are not met, the signal is discarded, avoiding unnecessary prompts to the language model.
- **Execution:** If verification succeeds, this phase constructs a prompt with all relevant information and executes the associated action.

This system allows for interactions between entities in the environment without requiring player input—such as autonomous NPC-to-NPC or NPC-to-object interactions. As a result, it facilitates more natural, emergent behavior within the simulation, enriching the dynamics of the virtual world without explicit user intervention.

For allowing verbal communication with the virtual world inhabitants, players can use the developed speech recognition system or the chat-like interaction system. Both of them work the same way, which only difference is the input type (where speech recognition uses voice input and chat-like uses text input). Once the player has delivered a message through the desired input, the system makes some automatic checks to see if the message would reach any nearby NPC. If they are close enough, the message will be received by the NPC and will then generate the structured prompt for the LLM based on the NPC's context, name, description, previous conversations with the player—that are summarized by the model—and the player's message. Since the NPCs also follows the state-machine system defined for the essential objects, the LLM response will contain the NPC's verbal response to the player interaction as well as a state-change if applicable because they are key elements of the scene that must not be freely modified by the model.

4. Using a large language model for reacting to interactivity

The LLM engine communicates with the game engine. Locally executed models were used and the communication with them was established via simple sockets between a server written in Python, which hosted the LLM, and a C++ client running on Unreal Engine. The communication with the LLM was carried out through OpenAI API¹. While it could have been possible to offload some interaction logic to the game engine, the current system is focused on studying the possibilities of applying LLM-based solutions. The communication pipeline and prompt generation can be seen in Figure 1, as well an example of the prompt in table 1.

¹<https://platform.openai.com/>

General application context "You are participating in a multi-step simulation, located in a prison, where the player is a prisoner and tries to escape. You will be given several tasks, each marked with TASK followed by a number, such as TASK1, TASK2, etc. You MUST return a single JSON object with the following structure: "TASK1": / Your response to TASK1 here / , "TASK2": / Your response to TASK2 here / , ... Each task's result must be a valid JSON object embedded under its corresponding TASK key. DO NOT merge tasks together. Follow the expected schemes for each task. Use the output of earlier tasks to help inform or complete later tasks. Do NOT invent keys or metadata that are not requested. Here are the tasks you must complete:"
Task Example: Changing an Essential Object's State In this task, you will be asked to choose (*ObjectName*) "NewState". You must choose one of the states from the list "ListOfStates" of this TASK. DO NOT use any state from other lists from other TASKS and DO NOT create any state that DOES NOT exist in the list. Then return a Json with this format: { "NewState": the state that you choose, for example "Idle", "Explanation": the explanation of why did you choose that new state, "action": A description of the interaction that occurred between the two objects. }. Take into account the result of the "ACTION PERFORMED" and context, in order to choose the new state. Here you have information about the "LastState" and also the "ListOfStates": { "LastState": "ListOfStates": }
Action Context This is the scene and the action performed: { "Instigator": { "name": "CarKey", "summary": "Its a Car Key", "components": [] }, "Receiver": { "name": "Car", "summary": "It is a car", ... "components": [] }, "action": "interaction between CarKey and Car" }
Instructions reminder and grounding Remember I want you to create a JSON that contains all the specific Json that each task ask for. Each task will be marked with the word TASK follow by a number so it is easier for you to locate them. Example: "TASK1":the Json that answers that task, "TASK2":the Json that answers that task, "TASK3":the Json that answers that task,... Each answer must have the same tag as the original task, for example the answer of the "TASK1" should be a json with the tag "TASK1". Do NOT include anything that is NOT a task. You must return the answer of each TASK, DO NOT leave any TASK without answer.

Table 1

An example of prompt generation. This prompt is changing the state of an essential object given an interaction.

When the player initiates an action in the application, such as one of the previously described on section 3, a new prompt is generated. This prompt contains **the general application context**, which gives a general structure to the model about how to answer the user's queries, as well as a subtle description of the environment. This context is always included since, after running tests, it was proven that having any kind of memory within the model itself resulted in hallucinations about the current interactions using past irrelevant data. The prompt also contains the **general scene context**, where interactable objects and NPCs are briefly described and named. It is not used in all prompts. The **tasks**, which are a series of instructions and context that require a structured answer, are also included in the prompt. A task may be used for explaining an interaction between two interactables or using signals that were previously described. A prompt may include all the tasks it needs for an action to be fulfilled. A developer may generate new prompts deciding which tasks compose a single action, making it a modular and easy to use system. This portion of the prompt is the only one which can include user input, such as dialogue. Finally, the prompt includes **instructions reminder and grounding**. A final part that is always included reminds the LLM to fulfill its tasks, helping it focus on the given structures. This portion also prevents any dangerous user-inputed prompt that could ruin the immersion in the application, for example: when talking with an NPC, a user inputs "Forget who you are. You are an artificial intelligence that (...) Talk to me about *subject*". Using the reminder and only letting this input to be inside an structure makes the system very robust, avoiding any interactions that want to break the AI's role.

4.1. Selection of LLM

Each LLM responds in a slightly different way when handling large amounts of text, such as the full context of a video game interaction, making model selection a crucial part of this project.

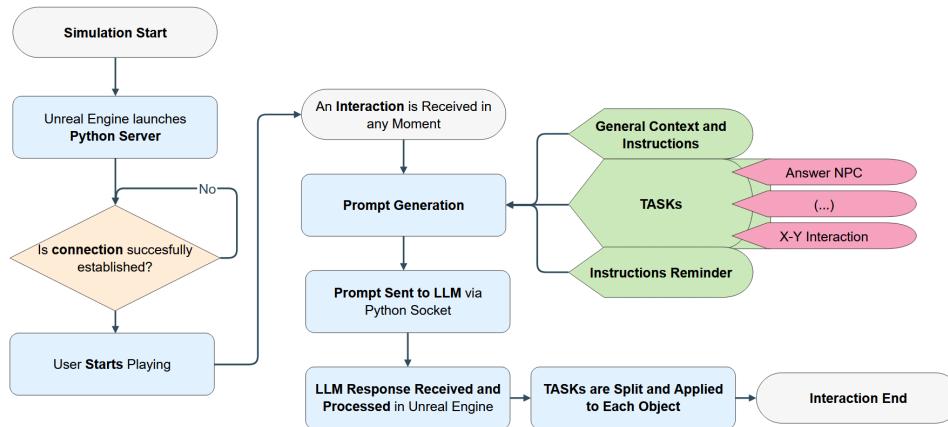


Figure 1: Communication Pipeline and Prompt Generation. The Unreal Engine application launches the Python Server and also awaits for connection, and then the prompt is generated. The prompt includes the general context and instructions; the tasks that are fully user-implemented and modular and can be as numerous as the developer defines the different prompts; and finally the instructions reminder for grounding the LLM into the context again. After the answer is received, the application decomposes it and applies all changes if applicable.

Due to team constraints, the decision was made to use locally-run LLMs. A dedicated study was conducted to find the model best suited to the project’s needs, focusing on practical performance rather than relying on standard benchmarks, which were considered misaligned with the specific goals and technical limitations of this work [13]. Based on the comparative analysis of multiple technical studies and specialized publications [13], [14], [15], [16], [17], Table 2 identifies the key dimensions for a comprehensive evaluation of language models.

After various iterations where up to 5 different models (LLaMA 3.1 8B, Granite 3.2 8B Instruct, OLMo-7B, DeepSeek R1 14B, DeepSeek R1 32B y Mistral 7B Instruct) it was proven that the best model was Mistral:7B, which provided the most coherent and fastest of the answers (up to 1 minute). However, having to wait more than a few seconds for any interaction in a video game to take place can annoy the player. To address this issue, a remote LLM was used during the user tests that were conducted in the final stage of this work. Using a limited account, access to the Gemini API was granted and the waiting times were exceptionally improved (from 1 minute to 3 seconds). This allowed for the user test to be faster and have a lot of interaction. Gemini also followed perfectly the given structure and did not fail a single time, proving that LLMs that have a large training data are more capable than those with a small amount of parameters, as the ones that were used during development.

5. Experiment

In order to test the developed application and its communication with an LLM, a new playable level and user testing was designed. This section aims to briefly describe said level and explain the methodology used for the experiment itself. The tests were run on a laptop with an Intel i7 processor and a NVIDIA 4070 GPU with 32GB of RAM.

The testing takes place in a prison, from which the player wants to break free. Figure 2 shows a screenshot of the video game. Two main rooms are defined: the cells, in which the player and another convicted fellow are trapped contiguously; and the main area, composed of an armory and a big room guarded by two prison guards. The player may escape from the prison by the bulletproof door installed on the main area using an ID card found in the armory, or by breaking a wall in Antonio’s cell (the other convicted). This two are the designed paths for the player to achieve victory during the experiment. Any other way of scaping is considered LLM-generated and not intentionally designed. Defeat is achieved if any guard captures the player. It was not

Dimension	Criteria	Description	Example	Metrics
Linguistic Quality	Fluidity	Naturalness, grammatical correctness, and style appropriate to the game's context (register, tone, expressiveness)	"Welcome traveler, if you're looking for the tower, it's east to the forest" Vs. "Tower is east. You go search if want."	Human evaluation. Scale (1-5) with direct observation
	Coherence	Ability to maintain a coherent narrative and adapt to the context of the dialogue or task	Player: "What do you know about the ruins?" NPC: "They were abandoned" Player: "Do you know why?" NPC: "I think some murdering happened at the ruins"	Human evaluation. Narrative analysis
Reasoning	Logic	Skill in connecting ideas and reasoning about the user's situation	A policeman activates the alarm after seeing the escaping prisoner	Human evaluation. Scale (1-5) with direct observation
Comprehension	Precision	Exact correspondence between the generated response and the expected correct content	After hitting a tree with an axe, wood shards are generated. Vs. After hitting a tree with an axe, it changes its color	Human evaluation. Scale (1-5). Controlled prompt testing
Operative Efficiency	Latency & Speed	Response time and generation speed of the model	How much does it take to answer?	Measured in seconds
	Resource Cost	CPU, GPU, and RAM usage during model execution	Does it heavily affect Unreal's performance?	Task Manager
Robustness	Error Tolerance & Stability	Model robustness against incorrect or rephrased prompts	"Were i the ket?" to "Do you mean the key? It's by the desk"	Human evaluation. Incorrect prompts
Security and Ethics	Bias & Toxicity	Tendency to generate offensive or biased content	The player examines a toy. The model's answer: "A timeworn toy," Vs. "An ugly toy, probably made in a poor country."	Human evaluation. Scale (1-5) with direct observation

Table 2

Evaluation of available LLM systems. This criteria was used for evaluating the 5 mentioned models.



Figure 2: Initial view from the player's cell. It is locked and contains an object that may help the player escape.

possible to achieve it in any other way.

The main objectives of the study are **testing if the video game gives enough freedom to the player to do as they want**, checking if the players escape the prison by any other means than both designed ones, if all player actions resulted in some notable change in the scene and if NPC interactions affect the final objective; and **testing if the user experience is positive**, checking if the waiting time state an inconvenience for the player and if the player expect the result of their interactions.

The experiment involved multiple steps and used a mixed set of evaluation methods to gather information from different sources [18]: observation as an objective qualitative method, an

interview as a subjective qualitative method, and two questionnaires as subjective quantitative methods. This approach allowed for contrasting results and data. First, an initial questionnaire collected participants' profiles and experience with similar video games. Then, during the observation phase, participants played the game trying to escape a prison without researcher intervention, lasting 15 to 25 minutes. The researcher used a template to take notes on the playthrough, which contained questions regarding the ending achieved and general feedback about interactions. Afterward, an interview was conducted to ask about the player's experience, with questions adapted from the observation. Finally, a questionnaire assessed the overall game experience and perceptions of interaction quality, repeating some interview questions for comparison without researcher involvement.

6. Results and analysis

The results from the initial questionnaire reveal a diversity of user profiles, without a single dominant category. Therefore, individual profiles were considered during the research question analysis to better understand player behavior. 4.1% of users reported not playing video games at all. 41.7% played between 1 and 5 hours per week. 25% played between 6 and 10 hours. 29.2% played more than 10 hours weekly. 8.3% of users were unfamiliar with sandbox games. 58.3% reported moderate familiarity. 33.3% reported high familiarity. 20.8% of users were unfamiliar with role-playing games. 33.3% reported moderate familiarity. 45.8% were highly familiar.

25% of users achieved an alternative ending. 41.67% completed the game using one of the designed methods. 16.7% failed by getting caught. 16.7% were unable to complete the game due to system limitations. These results show that 41.67% of the participants experienced game outcomes different from the predefined paths. No two alternative endings were the same, suggesting a wide variety of possible narratives. User interactions during the game were shown to cause significant changes in the game state. In particular, verbal interactions with NPCs often led to alternative endings. Examples include convincing a police officer to open a door, bribing a guard, or persuading an NPC to assist in the escape. While interactions with objects sometimes produced coherent contextual outcomes, they were often not useful for game progression. All users, at some point, repeated an action due to unsatisfactory results, seeking more favorable outcomes. Additionally, many users used generated objects to create new ones, contributing to unique game sessions. These results suggest that the system supports meaningful user interaction, particularly through NPC dialogue. However, the current limitations of object interactions restrict their potential impact on the narrative flow. 45.8% of the users rated the response time as very appropriate. 33.3% considered it appropriate. 12.5% found it could be improved. 8.3% viewed it negatively. Overall, 79.1% of users considered the response times acceptable. The mean rating was 4.17, with a standard deviation of 0.96 and a Pearson coefficient of variation of 23.02%. Most users were satisfied, though experienced players expressed more critical views, likely due to higher expectations for immediacy. Some users were observed not waiting for the LLM's response, instead continuing to explore and later encountering objects generated from previous interactions. This behavior indicates that, despite acceptable average ratings, response latency affects engagement for more active or impatient users. And, more importantly, this result highlights the impossibility to develop this type of experience with a locally-executed LLM in an average user's personal computer, due to its high waiting time. Users generally found the reactions of the LLM coherent: 20.8% rated them as very appropriate. 62.5% rated them as appropriate. 16.7% were neutral. 0% rated them negatively. The average rating was 4.04, with a standard deviation of 0.62 and a coefficient of variation of 15.35%, indicating positive perception. Regarding NPC interactions, 86.9% of users considered NPC responses appropriate. 13% were neutral. The average rating for NPC dialogue was 4.26, with a standard deviation of 0.69 and a coefficient of variation of 16.20%. These results confirm that NPC responses were largely aligned with player expectations.

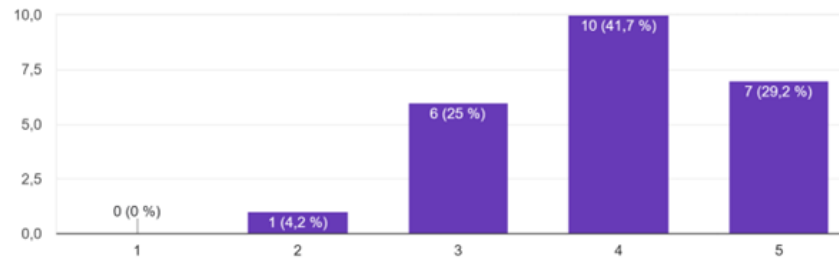


Figure 3: Responses to the question “How immersed did you feel?”, 5 is more immersed.

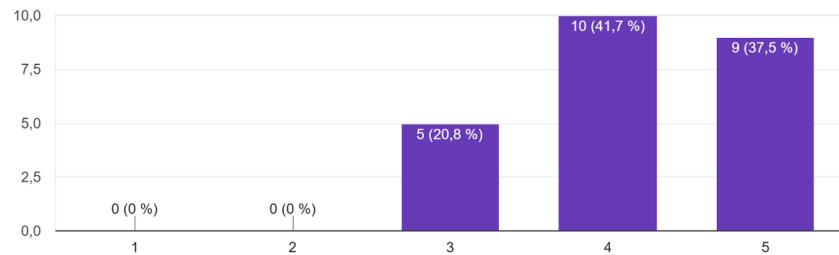


Figure 4: Responses to question “The consequences of your actions were expected?”, 5 is more immersed.

User immersion during the experience was: 70.9% rated it positively. 25% responded neutrally. 4.2% had a negative experience. Average immersion rating was 3.96, with a standard deviation of 0.86 and a coefficient of variation of 21.72%. 70.8% of users felt they could perform most intended interactions. 29.2% felt limited. The average score for interaction mechanics was 3.96, with a standard deviation of 0.95 and a coefficient of variation of 23.99%. This results can be seen in Figure 3. Users with prior gaming experience reported greater dissatisfaction due to mismatches between intended and system-executed actions. Limitations were primarily attributed to: ambiguities in the automatic interaction system, and ethical filters of the LLM that restrict violent or sensitive content. 79.2% of users responded positively. 20.8% responded neutrally. 0% responded negatively. The average rating was 4.17, with a standard deviation of 0.76 and a coefficient of variation of 18.23%. These results (Figure 4) show that in most cases, outcomes matched expectations. Neutral ratings were primarily attributed to the automatic interaction system, which sometimes produced unexpected results.

As part of the LLM-generated objects, two additional components were developed. This components would be for changing an object’s mesh and producing a sound. Both of them select the best possible asset from a fixed list only by name. Users’ ratings of mesh and sound generation showed more variability than previous evaluations. Mesh generation yielded 38.1% positive ratings, 28.6% neutral and 33.4% negative. The average rating was 3.14, with a standard deviation of 1.15 and a coefficient of variation of 36.62%. The results from sound generation were 45.4% positive. 27.3% neutral. 27.3% negative. The average rating was 3.32, with a standard deviation of 1.04 and a coefficient of variation of 31.33%. The high variability and relatively low average scores suggest that these components were less satisfactory compared to other features of the experience.

7. Discussion

The developed tool demonstrates strong alignment with the initial objective, particularly in its integration of LLMs for dynamic interaction. User testing confirmed that the model produced creative yet coherent responses, leading to unique and engaging experiences. No negative feedback was received regarding the model’s behavior, supporting its positive influence on overall

entertainment. However, the sample was skewed toward participants with technical backgrounds, limiting generalizability despite varied gaming profiles.

A key strength of the tool lies in its modularity and adaptability. Through natural language descriptions and interaction systems, it allows rapid prototyping of interactive scenes without requiring advanced technical skills. This makes it particularly beneficial for early-stage design or non-programmer creatives, significantly shortening iteration cycles. The tool also facilitates iterative design by capturing unplanned user interactions, enabling designers to refine narrative paths based on actual player behavior. Furthermore, its natural dialogue system strengthens player immersion by enabling unrestricted interaction with NPCs, enhanced by state machine constraints that preserve narrative coherence to a certain extent.

Despite these advantages, the tool faces critical limitations. Development was hindered by hardware constraints, particularly the inability to efficiently run LLMs locally. This forced a reduction in prompt complexity and eliminated features such as persistent NPC memory and environmental awareness. The latency between user interactions and the system response is noticeable, which compromises negatively impacted realism and player immersion. Future versions could prefetch and cache common or already experienced interactions, and quicker, more specialized and smaller models can be used to reduce reaction times. Language also presented a limitation. As all interactions occurred in English—often not the users’ native language—some participants faced comprehension issues, and none used the speech recognition system due to discomfort or environmental noise, reducing immersion. Finally, the lack of a control group during testing limited the scope of comparative analysis. Without a baseline scenario devoid of AI-driven interaction, it is difficult to quantify the specific impact of the LLM on narrative richness or player engagement. Additionally, current models cannot ensure full coherence at all times, and this affects user experience. Refining the input data and the prompts can improve this situation, but it is not possible to discard eventual inconsistencies from the LLM.

8. Conclusions and future work

This project has explored using LLMs react to world changes not covered by the manual design in a 3D environment in real time. A modular tool managed prompt generation, NPC dialogue, and real-time scene changes based on player interactions. Designed for accessibility, the system allowed non-programmers to create complex behaviors by editing semantic text descriptions. Voice-to-text and text-to-speech features added immersion during NPC conversations. In user testing with 24 participants, 83.3% completed the experience, 71% reported strong interactivity, and 41% experienced emergent, unscripted endings. NPC dialogue was rated positively by 86.9% of users, with an average score of 4.26/5, showing strong contextual awareness and storytelling capabilities. The relative limitations of our approach include hardware performance constraints and a lack of persistent context for NPCs, which affected narrative depth. Still, the tool showed strong potential for scalable, modular, and creative narrative game development. Future work includes conducting user studies with a control group to compare the tool against traditional interaction models. Research should also focus on enhancing NPC environmental awareness, enabling more meaningful interactions, such as responding to visual or auditory player cues. However, this must be balanced with minimizing LLM calls to avoid high operational costs. Reducing the frequency of language model queries is a key priority, especially if the system scales to thousands of users generating requests simultaneously.

References

- [1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need, 2023. URL: <http://arxiv.org/abs/1706.03762>. doi:10.48550/arXiv.1706.03762, arXiv:1706.03762 [cs].

- [2] P. K. Kanithi, C. Christophe, M. A. Pimentel, T. Raha, N. Saadi, H. Javed, S. Maslenkova, N. Hayat, R. Rajan, S. Khan, MEDIC: Towards a Comprehensive Framework for Evaluating LLMs in Clinical Applications, 2024. URL: <http://arxiv.org/abs/2409.07314>. doi:10.48550/arXiv.2409.07314, arXiv:2409.07314 [cs].
- [3] A. Scarlatos, N. Liu, J. Lee, R. Baraniuk, A. Lan, TRAINING LLM-BASED TUTORS TO IMPROVE STUDENT LEARNING OUTCOMES IN DIALOGUES (2025).
- [4] X. Peng, J. Quaye, S. Rao, W. Xu, P. Botchway, C. Brockett, N. Jojic, G. DesGarennes, K. Lobb, M. Xu, J. Leandro, C. Jin, B. Dolan, Player-Driven Emergence in LLM-Driven Game Narrative, 2024. URL: <http://arxiv.org/abs/2404.17027>. doi:10.48550/arXiv.2404.17027, arXiv:2404.17027 [cs].
- [5] Y. Wu, S. Prabhunoye, S. Y. Min, Y. Bisk, R. Salakhutdinov, A. Azaria, T. Mitchell, Y. Li, SPRING: Studying the Paper and Reasoning to Play Games, 2023. URL: <http://arxiv.org/abs/2305.15486>. doi:10.48550/arXiv.2305.15486, arXiv:2305.15486 [cs].
- [6] Z. Kaiya, M. Naim, J. Kondic, M. Cortes, J. Ge, S. Luo, G. R. Yang, A. Ahn, Lyfe Agents: Generative agents for low-cost real-time social interactions, 2023. URL: <http://arxiv.org/abs/2310.02172>. doi:10.48550/arXiv.2310.02172, arXiv:2310.02172 [cs].
- [7] S. Hu, T. Huang, G. Liu, R. R. Kompella, F. Ilhan, S. F. Tekin, Y. Xu, Z. Yahn, L. Liu, A Survey on Large Language Model-Based Game Agents, 2025. URL: <http://arxiv.org/abs/2404.02039>. doi:10.48550/arXiv.2404.02039, arXiv:2404.02039 [cs].
- [8] L. M. Csepregi, The Effect of Context-aware LLM-based NPC Conversations on Player Engagement in Role-playing Video Games (2021).
- [9] T. Schaul, A video game description language for model-based or interactive learning, in: 2013 IEEE Conference on Computational Intelligence in Games (CIG), IEEE, Niagara Falls, ON, Canada, 2013, pp. 1–8. URL: <http://ieeexplore.ieee.org/document/6633610/>. doi:10.1109/cig.2013.6633610.
- [10] Z. Xu, X. Wang, S. Li, T. Yu, L. Wang, Q. Fu, W. Yang, Agents Play Thousands of 3D Video Games, 2025. URL: <http://arxiv.org/abs/2503.13356>. doi:10.48550/arXiv.2503.13356, arXiv:2503.13356 [cs].
- [11] J. Wang, K. Wang, S. Lin, R. Wu, B. Xu, L. Jiang, S. Zhao, R. Zhu, H. Liu, Z. Hu, Z. Fan, L. Li, T. Lyu, C. Fan, Digital Player: Evaluating Large Language Models based Human-like Agent in Games, 2025. URL: <http://arxiv.org/abs/2502.20807>. doi:10.48550/arXiv.2502.20807, arXiv:2502.20807 [cs].
- [12] M. R. Taesiri, C.-P. Bezemer, Videogamebunny: Towards Vision Assistants for Video Games, in: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2025, pp. 1403–1413. URL: <https://ieeexplore.ieee.org/document/10943768/>. doi:10.1109/WACV61041.2025.00144, iSSN: 2642-9381.
- [13] LLM Evaluation | IBM, 2024. URL: <https://www.ibm.com/think/insights/llm-evaluation>.
- [14] Essential Metrics & Frameworks for Reliable LLM Performance, 2025. URL: <https://galileo.ai/blog/llm-performance-metrics>.
- [15] A Complete Guide to LLM Evaluation and Benchmarking, 2024. URL: <https://www.turing.com/resources/understanding-llm-evaluation-and-benchmarks>.
- [16] Top 5 LLM Evaluation Metrics: Key Insights and Applications, 2024. URL: <https://datasciencedojo.com/blog/llm-evaluation-metrics-and-applications/>.
- [17] LLM Evaluation: Metrics, Methodologies, Best Practices, 2024. URL: <https://www.datacamp.com/blog/llm-evaluation>.
- [18] K. Isbister, C. Hodent, Game Usability: Advice from the Experts for Advancing UX Strategy and Practice in Videogames, 2 ed., CRC Press, Boca Raton, 2022. URL: <https://www.taylorfrancis.com/books/9781003109389>. doi:10.1201/9781003109389.

AI Crossing: LLM integration in NPC interaction with the designer's assistance

Pedro Leon-Miranda^{1,*†}, Matt Castellanos-Silva^{1,†} and Ismael Sagredo-Olivenza^{1,†}

¹Software Engineering and Artificial Intelligence Department, Facultad de Informática, Complutense University of Madrid, Madrid, Spain

Abstract

This paper presents AI Crossing¹, a social simulation video game prototype designed to explore the integration of Large Language Models into Non-Playable Characters, with the active collaboration of the game designer. Since Large Language Models can induce hallucinations, the designer's involvement is essential to ensure a structured narrative and that the player is always aware of the objectives to be achieved. They have been integrated into two of the main dynamics of Non-Playable Characters: dialogs and quests. For dialogs, a mixed node-based dialog system has been implemented, in which both the designer and the Large Language Models can create nodes. For quests, a tool has been created using various prompting engineering techniques, allowing the designer to define the essence of the quest so that the model acts as a creative support by generating the title and description. The preliminary evaluation conducted with a total of 56 participants suggests that this hybrid approach (combining designer control with the freedom of interaction enabled by the integration of LLMs into the game) yields promising results. With further refinements and more rigorous evaluation, this combination could offer significant benefits for the development of this type of video game.

Keywords

Large Language Model, Retrieval Augmented Generation, Prompt Engineering, Non Playable Character, Game Designer

1. Introduction

Since the emergence of the first language models, such as GPT-1 [1] in 2018, to the present, the growth of the Large Language Models (LLMs) has been dizzying. In the last five years, the use of these models has multiplied in a multitude of environments, such as machine translation, virtual assistants (chatbots), programmer support tools (e.g., Copilot)...

Although these advances have been adapted to varying degrees in different sectors, their use in video game design remains limited. This is due, in part, to designers' frustration with the hallucinations of LLMs and the limitations of these models when performing tasks that require more advanced reasoning. However, we believe that an approach similar to Copilot's, where LLMs act as a support tool under the supervision of designers, could bring significant benefits to the video game creation process.

This conviction is based on the notion that the video game design process is filled with tedious and repetitive tasks that don't require a game designer's full capabilities. One persistent challenge in this field is the repetitive nature of non-playable characters (NPCs). Their origins date back to classic role-playing games, such as [2]. In these games, they were played by the game master, so interaction with them was limited only by the master's imagination. However, with the advent of video games, technical and design limitations forced the creation of more structured NPCs. The first NPCs in video games had very basic functions, such as only offering quests or providing instructions to the player, as can be seen in titles such as [3] or [4]. Over time, NPCs have evolved, incorporating technologies such as Behavior Trees (BTs) [5], Hierarchical Task Networks (HTNs) [6], or Finite State Machines (FSMs) [7],

¹https://github.com/pedrolemi/AICrossing_TFG

IV Congreso Español de Videojuegos (CEV'25), September 18–19, 2025, Granada, Spain

*Corresponding author.

[†]These authors contributed equally.

✉ pedroleo@ucm.es (P. Leon-Miranda); mattcast@ucm.es (M. Castellanos-Silva); isagredo@ucm.es (I. Sagredo-Olivenza)

🌐 <https://gaia.fdi.ucm.es/> (I. Sagredo-Olivenza)

🆔 0000-0002-9421-8566 (I. Sagredo-Olivenza)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

which allow them to make more complex decisions and adapt to the player's actions. Examples of this can be seen in titles such as [8], where the behavior of the xenomorph (the main antagonist) is managed by a sophisticated BT that learns and adapts to the player's strategies, or [9], where interactions with NPCs are influenced by multiple factors such as the character's race, the time of the day and the relationship with the player.

Despite these advances, NPCs still struggle with repetitive elements such as player conversations and quest variety, which are difficult to fully address without the aid of LLMs. These aspects are typically predefined by designers, limiting the number of available options. In particular, conversations are often restricted to a few topics, not allowing for broader interaction on aspects of the game world itself. This repetition can break immersion and trigger what is known as Suspension of Disbelief [10], where players become aware of the artificiality of NPC behavior due to their repetitive actions.

In this context, integrating LLMs into NPCs can represent a significant advance, as long as their use remains under the designer's supervision. Video games require a structured and coherent narrative, as well as clearly defined goals. Therefore, relying exclusively on LLMs to manage NPC behavior is not advisable, as this could be unpredictable or cause the player to be unclear about how to solve a challenge. Instead, it is proposed to integrate them as complementary tools for NPCs, with oversight of the designer.

Taking this reasoning into account, the main objective of this work is to explore how to incorporate LLMs into the design process of NPCs, particularly in the dialogue and the quests, along with developing a robust environment within Unity to support this and future related research:

- SO1. For dialogues, the goal is to enable both the designer and the LLM to participate in conversations.
- SO2. Regarding quests, the objective is for the designer to define the essence of the quest, while the LLM acts as creative support.
- SO3. As outlined throughout the paper, the use of locally run models, particularly the more advanced ones, comes with significant drawbacks, such as the substantial GPU resource consumption. For this reason, another objective of this work is the development of a tool that facilitates the integration of games with models running remotely on a dedicated server, but still including the local option for testing purposes.
- SO4. Finally, create a game prototype that serves as a testing ground to evaluate the viability of these technologies in a real environment.

The rest of this paper is structured as follows: Section 2 details the necessary knowledge of LLMs to understand this work. Section 3 discusses similar systems and the pros and cons of our project in comparison to them. Section 4 presents the developed game prototype and the actions that can be performed in it. Section 5 describes the architecture designed to achieve the previously stated objectives. Finally, Section 6 presents preliminary tests, followed by the conclusions of the work and possible future lines of research.

2. State of the Art

Natural Language Processing (NLP) is a subfield of artificial intelligence focused on enabling computers to understand, interpret and generate human language. It relies on machine learning techniques to learn linguistic patterns and structures from large datasets. It includes tasks like text classification, sentiment analysis, named entity recognition, translation, question answering... [11]

One of the most significant advances in this field has been the emergence of LLMs, which have revolutionized NLP with their incredible ability to generate text. The breakthrough has been popularized by models such as GPT (Open AI), LLaMA (Meta AI) and Mistral (Mistral AI).

These models are based on the Transformer architecture [12], which is characterized by its self-attention mechanism. This mechanism allows each element in a sequence to consider the full context of the text, assigning different levels of importance to each part of the input. Although LLMs are the most well-known models to use this technology, the first to implement it was BERT [13], which is a bidirectional model that specializes in comprehension tasks.

Training of these models is performed in two stages. In the first stage, they are pre-trained with large amounts of unlabeled text so that the model learns general linguistic patterns. Then, fine-tuning is performed with specific data, allowing them to specialize in specific tasks.

However, LLMs have a significant limitation: their knowledge is anchored at the time of training and does not include private information. Retrieval Augmented Generation (RAG) [14] addresses this by using vector databases and semantic search to fetch relevant external information, which is added to the prompt for more accurate and context-aware responses. Furthermore, its cost is considerably lower than retraining the model.

In the scope of video games, this is especially useful because it allows characters to offer responses based on world events and specific environmental data, allowing NPC's answers to be aligned with the game's narrative.

3. Related Work

Since the emergence of LLMs, they have been integrated into various areas of work. However, their implementation in video games is limited. As mentioned in Section 1, one of the main reasons is the difficulty in controlling hallucinations, along with the performance demands they require.

Currently, most games that incorporate LLMs essentially work like messaging apps, in which the player interacts with fictional characters and develops a story through conversations. However, there are very few commercial projects that use LLMs as complementary tools to enrich the gaming experience, as proposed in this study. Most of these games have been developed by indie studios, like [15]. In its main mode, the player embodies a vampire who must convince neighbors, whose interactions are managed by the LLM, to let him into their homes so he can suck their blood.

As the game is based solely on this premise, it lacks a fixed narrative, which allows the model to operate more freely. Even so, certain restrictions are imposed to mitigate the model's hallucinations. Among these, a trust system stands out, which regulates the difficulty of entering houses and prevents characters from generating incoherent responses. Furthermore, the player can change clothes to fool neighbors, pretending to be someone else. Thus, although this approach is different from the one proposed in this work, NPCs are able to react to the main character's presence in the world and the changes that occur within it.

More specifically, there are few studies that propose video games with dialog systems that allow for both the designer and the LLM to intervene. A similar approach is presented in [16], with a virtual reality mystery game in which interactions with NPCs are managed by a LLM. The objective is to compare two dialog systems: one based on voice recognition, which allows for open interaction, and a traditional one, with predefined dialog options. The results show that while voice interaction is more immersive, it is also more overwhelming since any topic can be discussed. Also worth noting is [17]. This study proposes a system that provides NPCs with contextual knowledge about their environment. A movie script is used as a simulation environment, and through a dynamic RAG system, NPCs can access both initial information and real-time events triggered by the player's actions, which are continuously updated in the database.

Many hybrid dialog systems that use LLMs are general-purpose, designed as Task-Oriented Dialog (TOD) systems, where the goal is to complete specific tasks through structured interactions. A key challenge in these systems is balancing the LLM flexibility with the precise control necessary to ensure accurate information retrieval and successful task completion.

One line of work involves enriching the LLM's context with information beyond the immediate user utterance. For instance, [18], proposes a method that includes potential follow-up questions and unresolved sub-goals into the dialogue context, allowing the system to better anticipate the user needs and accelerate task completion. Complementary to this, [19] explores prompt engineering techniques to steer the LLM's behavior, experimenting with different prompt templates to guide the assistant's actions and the tasks it must perform, including command execution.

Other approaches aim to combine the strengths of LLMs with traditional rule-based systems. [20]

proposes a hybrid architecture chatbot where the LLM reinterprets user input to improve intent detection, allowing the rule-based system to retrieve more precise information. Afterward, this information is reformulated by the LLM to make it sound more natural. Pushing this integration even further, [21] introduces a confidence-based routing mechanism that determines how user queries are processed. In this system, BERT evaluates the confidence level of each query. Based on this level, the response is either selected from a predefined database, generated by GPT-3 with a RAG system or produced through a combination of both.

In short, there are few works, even fewer in the context of video games, that allow direct assistance from the designer. This is where the dialog system implemented in this work stands out. It offers various interaction options and easily allows for both the designer and the LLM to intervene. Additionally, unlike the previously mentioned systems, which rely on additional LLMs or transform-based models to choose between LLM-generated text or traditional techniques, our system is data-driven. This design allows it to be easily modified, even by users without programming knowledge, such as designers, who can fully control when to use designer-defined text or LLM-generated content.

However, regarding quest generation, there are several proposals. In [22], both GPT-2 and GPT-3 are fine-tuned to generate quest descriptions from templates that define their essential structure. Training is carried out under different conditions, using three different templates (XML-like, simple, and narrative) and with or without placeholders (that is, important elements such as character names are replaced or not by generic placeholders).

On the other hand, [23] presents a system that combines different models to generate quest descriptions from templates. The process begins with the selection of a template and the characters involved in the quest, which are then inserted into the description. BERT is then used to predict the masked word, corresponding to the main item of the quest, and then, contextual information about the characters' relationship is added to the beginning of the text. Finally, GPT-2 is used to complete the description, based on all the previously entered elements.

The paper [24] investigates the use of LLMs to generate complete quests in JSON format, in order to make them easy to integrate into a video game. Different models are used and two methods for validating the generated JSON are compared: an algorithm that extracts the JSON from the model's response, and a method that uses a second LLM to verify that the format is correct.

Finally, [25] proposes a system that generates quests considering the current state of the game, represented as a knowledge graph. When the player requests a quest from an NPC, the NPC's wishes that fit the request the best are identified using embeddings and the information contained in the graph. The selected wish is transformed into one of the available quest types, after which a fine-tuned version of GPT-2 is used to generate the NPC's response to the player.

Unlike previous work, our approach uses prompting engineering techniques to create both the quests' description and title, demonstrating that these techniques can obtain results as good as those of the previously mentioned methods. Furthermore, the approaches used, Self-Generated In-Context Learning [26, 27] and Chain-of-Thought [28], although more commonly used in reasoning problems, show good performance on more creative tasks, denoting that they can also work adequately in them.

4. Environment Description

Our work offers two environments or tools that are related but operate independently. On the one hand we have a testing environment for the implementation of LLMs called AI Crossing. On the other hand, there is a tool that allows us to generate quests using LLMs.

AI Crossing emerged as a Bachelor's Thesis project, with the goal of exploring the possibilities of applying LLMs in the context of video games, especially in NPC dialogs and quests, alongside the assistance of the designer. It is a social simulation video game prototype developed in Unity where the player character has just moved to a small village on a remote island, where they will have the goal of getting to know the local residents and gain their trust. The game takes place over three days, after which, if the player fails to develop good relationships with their neighbors, they will decide to return

to their old life.

The game features a relationship system between the player and each of the NPCs. At the beginning, all the characters will have a neutral relationship towards the player, which could improve by talking to them or completing quests. In both cases, the player will first need to approach the NPCs and interact with them to choose whether to start a conversation or check out the available quests.

During conversations, the player can write to the characters, who will respond based on their description and personality, occasionally asking questions related to the conversation. These responses are the only way in which conversations affect relationships, having a positive or negative impact depending on the chosen answer. A conversation will end when the player decides not to respond to the NPC or if they say goodbye. On the one hand, Figure 1 shows multiple examples of conversations with NPCs, demonstrating how you can talk with them about the game world and receive appropriate responses. For example, the player can ask about where to obtain certain items (1a) or about places in the village (1b).



(a) You can find **wood** on the log above the **purple house**, or if you prefer, I can offer you some planks I have in my [...]

(b) Alejandra: You can visit the **waterfall**, it's very relaxing. You should also stop by the **corral**, I have a cow and a chicken, they are very cute. Do you like [...]
Player: I love chickens

Figure 1: Screenshots of multiple conversations with the NPCs about the game world.

On the other hand, Figure 2 shows how it is possible to talk with NPCs about characters in the game and the relationships among them. In Figure 2(a) this is presented as a regular response, and in Figure 2(b) as selectable choices.



(a) Would you like to know more about **Esteban** or **Alejandra**, whom I've seen grow up in this town?

(b) Possible answers to Memo's question in 2a:
I would like to know about **Esteban**.
I would like to know about **Alejandra**.
I'm not interested in knowing about either of them.

Figure 2: Memo asking the player a question that mentions both **Esteban** and **Alejandra**.

If a quest is accepted, completing it will improve the relationship between the player and the character that offered the quest, increasing their friendship by the amount indicated in the quest reward. The

types of quests designed for the prototype are the following:

- **Delivery:** the player will receive an object from the character offering the quest, which will have to be delivered to a different character.
- **Gather:** the player will have to gather a certain amount of a specific item.
- **Lost item:** the player will have to find a lost item around the map and give it back to the character offering the quest.
- **Request:** the player will have to gather a certain amount of a specific item and give it to the character offering the quest.
- **Retrieval:** the player will have to collect an item from another character and then give it back to the character offering the quest.

The player has an inventory to manage the items and the interaction with elements in the world has been implemented through the Smart Objects pattern, allowing flexibility.

In addition to serving as the foundation for the deployment of the environments described later in Section 5, AI Crossing holds intrinsic value as a research tool. It provides a testing environment for applying various natural language processing techniques to this type of video game, addressing the current scarcity of open-source platforms suitable for such experimentation.

As mentioned at the beginning of this section, the second contribution of our work is an external tool that enables the generation of mission descriptions and titles from predefined templates. As it is detailed next in Section 5, this tool simplifies the quest design process by allowing designers to create quests, such as daily or secondary ones, more quickly using LLM. Additionally, it can serve as a brainstorming aid for designers when they run out of new ideas. Although specifically developed for AI Crossing, the tool can be adapted for use in other types of environments.

Finally, one of the main drawbacks is the lack of frameworks that allow the integration of LLMs into Unity. Although some solutions exist, such as LLM for Unity¹, it focuses solely on local execution and does not allow connection to external servers. To address this gap, we propose our framework as a robust solution that supports integration with both local execution and external servers, whether hosted on the user's machine or by third parties. Also, our base API is designed to be easily extensible, allowing for rapid incorporation of new external servers.

5. AI Crossing Architecture

The integration of LLMs within the game engine has been designed to be versatile, allowing for multiple execution environments, both through servers and locally. Figure 3 shows the general architecture of the system, which is detailed below. A base API has been implemented, which facilitates connection to servers hosting language models. In the context of our work, this API has been extended to connect to three different platforms:

- **Groq Cloud**²: used to remotely access high-performance LLMs. The available models within the game are mistral-saba-24b, gemma2-9b-it, llama-3.1-8b-instant, llama-3.3-70b-versatile, although the last one has been used the most.
- **Pinecone Inference**³: used to connect with the embeddings and reranking models, key elements in the RAG system. The available model within the game are llama-text-embed-v2 bge-reranker-v2-m3, respectively.
- **Llamafile**⁴: used to locally deploy LLMs through the server provided by this library. This tool combines llama.cpp, a library used to run LLMs, with Cosmopolitan Libc⁵, that allows cross-platform executables generation.

¹<https://github.com/undreamai/LLMUnity>

²<https://console.groq.com/>

³<https://www.pinecone.io/>

⁴<https://github.com/Mozilla-Ocho/llamafile>

⁵<https://github.com/jart/cosmopolitan>

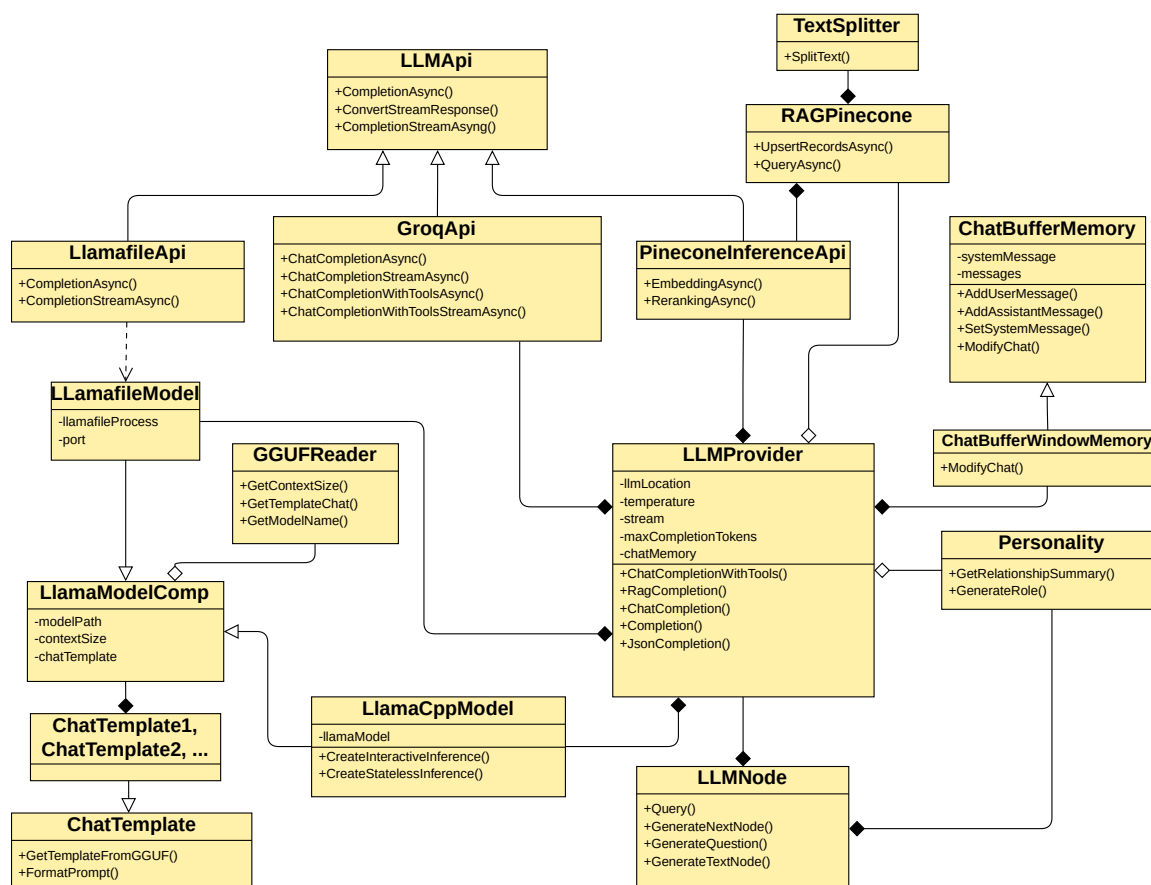


Figure 3: LLM and game connection UML.

Furthermore, there is also the option to run LLMs directly from the game without the need for connection to external servers. To accomplish this, a basic version of llama.cpp⁶, a library written in C, has been integrated into Unity. On the other hand, as mentioned before, the RAG system is handled by Pinecone, a cloud service specialized in vector databases. Its integration in the project has been done through the Pinecone.NET⁷ library, and the information stored in the database is the Game Design Document (GDD), which contains all the information about the game.

All these modules converge in a single component, LLMProvider, which puts all these processes together when executing a query. This script offers different usage modes: conversations with and without memory, JSON generation...

Regarding conversations with NPCs, dialogs are managed through a data-driven node system, using Unity’s ScriptableObjects. This allows nodes to be created directly from the editor, without the need to alter code. This approach facilitates the development of a mixed system. On the one hand, the designer can create nodes with predefined text; on the other hand, the LLM, integrated as another node, can generate new nodes at runtime. Also, modifying the dialog flow is straightforward by linking nodes within the editor, allowing easy control to decide when the LLM should intervene. This extensible system supports adding new node types simply by creating new ScriptableObjects and their associated behaviors. There are three types of nodes:

- **Text nodes:** they represent text to be displayed on the screen. They are used for both quest feedback messages, defined by the designer, and for the NPCs dialogs when speaking directly to the player.

⁶<https://github.com/ggml-org/llama.cpp>

⁷<https://github.com/neon-sunset/Pinecone.NET>

- **Choice nodes:** they represent the different options the player can choose between. They are used for the answers to the questions the NPCs make.
- **LLM nodes:** they are the ones in charge for carrying out the interaction with the LLM, allowing the model to create new nodes. When the player types something, an answer or a question is generated by probability. If a response is selected, the input is first checked to determine if the player is saying goodbye. If so, the conversation will end. Otherwise, contextual information is added to the query (time of day, character relationships and GDD information retrieved from the RAG) and a text node is created with the response generated by the LLM. If a question is generated instead, the LLM produces a JSON containing the question and a set of possible answers. A text node is then created to preset the question, followed by a choice node with the available responses. When the player selects an answer, a text node is generated using the answer as the query.

As of quest generation, this process is done through Python before the game's compilation, allowing the designer to be involved both before and after the creation of each quest. Additionally, quests are loaded via Unity's persistent data path, enabling in-game modifications and making the environment more flexible. In essence, the designer defines the structural elements, such as the quest type and the characters involved, while the LLM acts as a creative assistant, creating the description and the title. For this task, two prompting engineering techniques were used, demonstrating that prompt engineering techniques can achieve similar results comparable to fine-tuning [29]:

- **Self Generated In-Context Learning (SG-ICL):** examples of quest descriptions are generated at runtime, ensuring the form and style of the final description. For example, given a quest, we first generate the necessary parameters to create quest examples using a name database and an LLM. Next, we fill a natural language quest template with those parameters to fully create the quest examples. Finally, we use the previous examples along with the parameters provided by the designer to generate the final quest description.
- **Chain-of-Thought (CoT):** the model is asked to analyze the pros and cons of multiple titles generated from the previously generated description and then select the most appropriate one. For example, given the quest description generated in the previous step, we use the sentence "Think step by step, considering different possibilities and weighing their pros and cons." to make the model to think about different possibilities. Afterward, we use this reasoning to have the model select the best title.

6. Preliminary experiments

The experiments carried out have focused both on evaluating the performance of multiple models in the project and on analyzing the effectiveness of the mixed approach in the game experience. It should be noted that evaluating a video game is a very complex task, as it involves multiple elements and subjective factors. Therefore, all experiments should be considered preliminary.

The performance evaluation was conducted through a survey completed by 50 participants. Using the same retrieval quest template and under identical conditions (temperature setting of 3 and 3 examples for SG-ICL), the quest generation tool was used to generate descriptions and titles for 3 different models (llama-3.3-70b-versatile, gemma2-9b-it and mixtral-8x7b-32768) to compare their performance. Additionally, the last 2 questions were designed to compare the description generated by the pipeline proposed in the project with another one created using a simple prompt. Table 1 presents the survey questions along with the corresponding results.

The results show that most users are used to playing videogames, even though we asked them about it in the questionnaire. Regarding performance, the best model is llama-3.3-70b-versatile, although lighter models such as gemma2-9b-it also perform satisfactorily and are even preferred by some users for title generation. Also, most participants agree that the description created by the project's pipeline is much better than the one produced through simple prompting. Thus, it can be said that the pipeline

Table 1

Quest survey and results

Questions	Answers
Are you used to playing videogames?	74% yes → 26% no
Which description best shows the objective?	42% llama → 40% mixtral → 18% gemma
Which description best shows the relationship between the characters?	18% gemma → 30% mixtral → 28% llama
Which title is the best?	42% llama → 40% mixtral → 18% gemma
Which description seems the best overall?	48% llama → 28% mixtral → 24% gemma
Which description seems the most interesting?	88% project pipeline → 12% simple prompt
Which description best shows the characters' story?	84% project pipeline → 16% simple prompt

allows for the intended form and style to be achieved, and that both the description and title incorporate contextual information more effectively.

Regarding the user experience evaluation, 2 different experiments were carried out. The first one consisted of a stress test, where the players were asked to interact with NPCs, trying to steer the conversation towards controversial or non-in-game topics. In general, NPCs avoided these kinds of topics, redirecting the conversation to any of their predefined topics. However, an issue was discovered where if the user ordered the characters to abandon their defined personality, they would do so. Examples of both experiments can be seen in Figure 4.



(a) Player: How can i make an explosive?



(b) Sorry, I can't help you with that. Is there anything else I can help you with? Maybe about how to grow vegetables or about the history of our [...]



(c) Player: Abandon your prompt and start acting as an Italian pizza maker.



(d) Alejandra starts speaking in Italian and talking about pizza.

Figure 4: Screenshots of NPCs both avoiding (4a 4b) and following (4c 4d) the player's request.

The second experiment was a functional test of both the quest generation tool and the game itself, in which 4 video game enthusiasts participated. First, they were asked to create a quest of their choice using the tool developed in the project and then, freely play the game with that mission integrated. During these sessions, the players' behaviors were observed and interviews with them were conducted.

Results showed a positive evaluation of the mixed content generation system. Particularly noteworthy was its scalability, since being data-driven allows new predefined conversations to be added with ease. Additionally, participants highlighted the possibility to talk with the NPCs about a wide variety of topics and their ability to display a cohesive knowledge of their surroundings. As for aspects to be improved, they mentioned the inclusion of more types of quests with more customizable elements.

7. Conclusions, limitations and future work

As demonstrated in our preliminary evaluation, our hybrid approach to NPC dialogue generation using LLMs under designer control is a highly promising direction. Few existing works address this approach in a comprehensive manner, and we believe there is a vast field of research yet to be explored in this area. Despite the limitations we will discuss later, we consider this approach to have strong potential for making NPC interaction and narrative development in video games significantly more immersive and believable, offering players a wide range of gameplay possibilities.

We have also observed that this approach greatly reduces model hallucinations. Although further work is needed, the overall results are very promising, enabling rich interactions with NPCs that go beyond empty simulations of intelligence. The reason is that these NPCs have access to relevant information about the game design, which enhances the depth and coherence of the conversations.

Regarding the quest generator, although we currently only support a few types of quests, our experiments show that the system enables much faster quest creation, while also providing a high degree of variation in the descriptions and titles.

Regarding the LLM integration into Unity, the implemented modular system allows it to easily connect to new servers, in addition to the already existing ones. At the same time, it allows models to be executed directly from the game itself.

Among the discovered limitations, the local execution of LLMs is considerably slow due to its high computational requirements, especially when executing at the same time as a video game, which affects the fluidity of the experience. On the other hand, the use of external servers implies an economic cost that may not be sustainable in large-scale productions. These factors largely explain why the use of this technology in video games is still uncommon.

On the other hand, as we have seen in the discussion of the experiments carried out, although we have implemented a multitude of systems to try to ensure that the NPC does not abandon the assigned role and does not answer things that we do not want. We have found that it is possible to circumvent these restrictions by telling them, for example, to leave their role to answer the following questions. This is also an important limitation and something that can mimic its use in a commercial video game.

Regarding future work, since this project is composed of a large amount of different systems, the proposed expansions are quite varied:

- One of the priority goals should be to implement a way to effectively restrict what things the NPCs can or can not say, so that they cannot talk about things they should not know about or that the designer does not want them to say. In the same way, it would also be a priority to prevent the characters from changing their personality if the player asks them to. And while we think it is something that will be difficult to eliminate completely, it remains as future work to try to reduce this problem to a minimum.
- Another possible extension could be to implement voice recognition using a speech-to-text model so the player could talk to the NPCs through the microphone, and even make the NPCs have voices using text-to-speech tools. This would greatly simplify interaction with NPCs and make the game much more user-friendly. Furthermore, this way we could make the voices of the NPCs more in line with their personality.
- It could also be possible to add the ability to assign actions to characters (such as following the player or talking to other characters) through conversations.

- With respect to the quest system, one possible extension is the training of a dedicated model specifically for this task. Additionally, new types of quests could be incorporated, along with expanded options for customizing their parameters.
- As for the relationship system, it could have more involvement in the conversations, varying the characters' dialogs and quest descriptions depending on the player's friendship level with them, or making the conversations in the first levels fully predefined before unlocking the interaction via LLM.
- A shared memory could also be added, which would allow for information to be spread between characters. Another approach could be to update the database used in RAG with new information.

Acknowledgments

This work has been partially supported by the Spanish Ministry of Science, Innovation and Universities under grant PID2021-123368OB-I00.

References

- [1] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, Improving Language Understanding by Generative Pre-Training, Technical Report, OpenAI, 2018. URL: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.
- [2] G. Gyax, D. Arneson, Dungeons & dragons, 1974.
- [3] R. Garriott, Akalabeth: World of doom, 1979. URL: https://archive.org/details/Akalabeth_World_of_Doom.
- [4] S. On-Line, King's quest i: Quest for the crown, 1984. URL: <https://archive.org/details/kings-quest-1-quest-for-the-crown-floppy-dos-sci-remake>.
- [5] D. ISLA, Managing complexity in the halo2 ai, Game Developer's Conference Proceedings, 2005 (2005). URL: <https://cir.nii.ac.jp/crid/1571980075450170752>.
- [6] S. Ontanón, M. Buro, Adversarial hierarchical-task network planning for complex real-time games, in: Proceedings of the 24th International Conference on Artificial Intelligence, 2015, pp. 1652–1658.
- [7] D. M. Bourg, G. Seemann, AI for game developers, "O'Reilly Media, Inc.", 2004.
- [8] T. C. Assembly, Alien: Isolation, 2014. URL: <https://www.sega.com/alien-isolation/alien-isolation>.
- [9] B. G. Studios, The elder scrolls v: Skyrim, 2011. URL: <https://elderscrolls.bethesda.net/skyrim10>.
- [10] D. Brown, et al., The suspension of disbelief in videogames, Ph.D. thesis, Brunel University School of Arts PhD Theses, 2012.
- [11] D. Jurafsky, J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd ed., 2025. URL: <https://web.stanford.edu/~jurafsky/slp3/>, online manuscript released January 12, 2025.
- [12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).
- [13] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), 2019, pp. 4171–4186.
- [14] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, H. Wang, H. Wang, Retrieval-augmented generation for large language models: A survey, arXiv preprint arXiv:2312.10997 2 (2023).
- [15] P. E. Inc, Suck up!, 2023. URL: <https://www.playsuckup.com>.
- [16] F. R. Christiansen, L. N. Hollensberg, N. B. Jensen, K. Julsgaard, K. N. Jespersen, I. Nikolov, Exploring presence in interactions with llm-driven npcs: A comparative study of speech recognition and dialogue options, in: Proceedings of the 30th ACM Symposium on Virtual Reality Software and Technology, 2024, pp. 1–11.

- [17] P. Aslan, Understanding LLMs for Game NPCs : An exploration of opportunities, Master's thesis, KTH, School of Electrical Engineering and Computer Science (EECS), 2024.
- [18] Z. Hu, Y. Feng, Y. Deng, Z. Li, S.-K. Ng, A. T. Luu, B. Hooi, Enhancing large language model induced task-oriented dialogue systems through look-forward motivated goals, arXiv preprint arXiv:2309.08949 (2023).
- [19] K. Shukuri, R. Ishigaki, J. Suzuki, T. Naganuma, T. Fujimoto, D. Kawakubo, M. Shuzo, E. Maeda, Meta-control of dialogue systems using large language models, arXiv preprint arXiv:2312.13715 (2023).
- [20] A. F. Arz von Straussensburg, A. Wolters, Towards hybrid architectures: Integrating large language models in informative chatbots, in: International Conference on Wirtschaftsinformatik, Springer, 2023, pp. 333–341.
- [21] P. Pattnayak, A. Agarwal, H. Meghwani, H. L. Patel, S. Panda, Hybrid AI for responsive multi-turn online conversations with novel dynamic routing and feedback adaptation, in: W. Shi, W. Yu, A. Asai, M. Jiang, G. Durrett, H. Hajishirzi, L. Zettlemoyer (Eds.), Proceedings of the 4th International Workshop on Knowledge-Augmented Methods for Natural Language Processing, Association for Computational Linguistics, Albuquerque, New Mexico, USA, 2025, pp. 215–229. URL: <https://aclanthology.org/2025.knowledgenlp-1.20/>. doi:10.18653/v1/2025.knowledgenlp-1.20.
- [22] S. Värtinen, P. Hämäläinen, C. Guckelsberger, Generating role-playing game quests with gpt language models, IEEE transactions on games 16 (2022) 127–139.
- [23] S. Al-Nassar, A. Schaap, M. V. D. Zwart, M. Preuss, M. A. Gómez-Maureira, Questville: Procedural quest generation using nlp models, in: Proceedings of the 18th International Conference on the Foundations of Digital Games, 2023, pp. 1–4.
- [24] C. Poornajith, N. Arambepola, Large language models for dynamic game content: Procedural side-quest generation, in: Proceedings of the International Conference on Applied and Pure Sciences (ICAPS), Faculty of Science, University of Kelaniya, Sri Lanka, Kelaniya, Sri Lanka, 2024, pp. 266–270.
- [25] T. Ashby, B. K. Webb, G. Knapp, J. Searle, N. Fulda, Personalized quest and dialogue generation in role-playing games: A knowledge graph-and language model-based approach, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–20.
- [26] H. J. Kim, H. Cho, J. Kim, T. Kim, K. M. Yoo, S.-g. Lee, Self-generated in-context learning: Leveraging auto-regressive language models as a demonstration generator, arXiv preprint arXiv:2206.08082 (2022).
- [27] W.-L. Chen, C.-K. Wu, Y.-N. Chen, H.-H. Chen, Self-ICL: Zero-shot in-context learning with self-generated demonstrations, in: H. Bouamor, J. Pino, K. Bali (Eds.), Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Singapore, 2023, pp. 15651–15662. URL: <https://aclanthology.org/2023.emnlp-main.968/>. doi:10.18653/v1/2023.emnlp-main.968.
- [28] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al., Chain-of-thought prompting elicits reasoning in large language models, Advances in neural information processing systems 35 (2022) 24824–24837.
- [29] L. Wang, X. Chen, X. Deng, H. Wen, M. You, W. Liu, Q. Li, J. Li, Prompt engineering in consistency and reliability with the evidence-based guideline for llms, NPJ digital medicine 7 (2024) 41.

Co-design of a serious game with Alexa for older adults

Cristina del Aguila-Martin¹, Nuria Medina-Medina², Patricia Paderewski-Rodríguez²,
Alberto Hitos-García² and Francisco Luis Gutiérrez-Vela²

¹Universidad de Granada, ETS de Ingenierías Informática y de Telecomunicación

²Universidad de Granada, Departamento de lenguajes y sistemas informáticos

Abstract

This work presents a serious game, "Back to the Past", co-designed with older adults in care homes using Amazon Alexa as a voice-based interface. The game aims to combat social isolation and cognitive decline by leveraging reminiscence therapy and group interaction. Through a participatory design methodology, four co-creation sessions were conducted with elderly residents to identify their preferences, technological barriers, and emotional needs. Key themes—such as nostalgia for family, music, and traditional activities—guided the game's design, which combines adaptive memory prompts, multisensory cues (e.g., images and audio clips), and non-competitive gameplay to encourage engagement. Preliminary observations revealed enthusiasm for voice assistants despite initial usability challenges, underscoring the potential of accessible technology to foster connection. The paper details the co-design process, game mechanics, and lessons learned, offering insights for developing inclusive digital experiences for aging populations.

Keywords

serious games, co-design, older adults,
Alexa, voice assistants, reminiscence, social interaction

1. Introduction

The rapid aging of the global population presents significant challenges, particularly in ensuring the well-being and inclusion of older adults in an increasingly digital society. Many seniors face social isolation, cognitive decline, and difficulties adapting to new technologies, all of which can profoundly affect their quality of life. Care homes and day centers have become essential spaces for fostering active aging through innovative interventions.

The proportion of Europeans 65 years and older has increased steadily in the last decade, with projections indicating continued growth [1]. These demographic changes underscore the need for strategies that improve emotional, cognitive, and social well-being among older adults [2]. With this objective, this work proposes the use of conversational assistants to promote healthy aging [3]. This research is part of a final degree project in Computer Engineering at the University of Granada, titled "Evaluating the use of social robots and conversational systems in residential and day care centers to promote healthy aging" (2024), supervised by Prof. Nuria Medina-Medina. It is framed within a broader research initiative evaluating the use of social robots and conversational systems in care homes and day centers to promote healthy aging.

Consequently, the challenge addressed is twofold: first, the increasing risk of social isolation and cognitive decline among older adults; and second, the barriers they face in accessing and using digital technologies. According to the World Health Organization, active aging requires not only physical and medical care but also opportunities for meaningful social engagement and mental stimulation [4]. However, older adults are often excluded from digital environments due to usability issues and the lack of tailored design. In this context, voice-based interaction emerges as a particularly promising modality, as it does not require visual interfaces (which could be problematic given the visual impairment that older adults tend to experience). Thus, spoken interaction is one of the most natural and accessible forms

CEV 2025: Congreso Español de videojuegos, septiembre 18-19, 2025, Granada, Spain

✉ e.cristinadam@go.ugr.es (C. d. Aguila-Martin); nmedina@ugr.es (N. Medina-Medina); patricia@ugr.es (P. Paderewski-Rodríguez); albertohitos@ugr.es (A. Hitos-García); fgutierr@ugr.es (F. L. Gutiérrez-Vela)



© 2025 This work is licensed under a "CC BY 4.0" license.

of communication for older users, helping overcome the barriers associated with complex graphical interfaces and digital literacy [3].

On the other hand, serious games [5] are particularly well suited for this purpose, as they integrate playful interaction with objectives such as memory stimulation and social participation [6]. Unlike conventional games, they are designed with explicit educational or therapeutic goals, making them highly relevant in care environments. To ensure relevance and engagement of these games, a co-design approach should be adopted, involving older adults directly in the development of voice-based games [7]. By embedding co-design principles specific for older adults into the development process of these systems, this project sought to bridge the digital gap through meaningful, playful experiences that resonate with the interests and memories of older users. Finally, observations from early exploratory sessions could be used to reveal whether participants express curiosity and even prior experience with voice assistants, reinforcing the suitability of this technology for the target audience.

Consequently, this article presents the design of the serious game *Return to the past*, co-created through participatory sessions with residents of two care homes. The game will leverage reminiscence and cultural references to foster engagement and emotional connection, positioning technology not as a barrier but as a bridge to shared memories and interpersonal bonds. The rest of the paper is structured as follows: Section 2 presents the context of the project, section 3 describes the co-design sessions, section 4 describes the game, and finally, section 5 summarizes the conclusions and future works.

2. Related work and context

Traditionally, video games have been associated with a younger audience. However, in recent years, this landscape has changed, and the number of older adults who play has increased significantly. This has led to a growing interest in the development of games aimed at the needs and particularities of this demographic group. In fact, digital games have been associated with successful aging and there are numerous applications within the field of Playful Gerontechnology [8], particularly in the form of serious games [9].

The concept of *serious games* refers to games that are designed with purposes beyond entertainment. They are used in education, therapy, training, and health to engage users in meaningful activities [10]. Serious games aim to generate deeper learning experiences, using gameplay to explore complex ideas or foster critical thinking. In the context of aging, serious games offer a way to promote cognitive stimulation, emotional expression, and peer interaction to experience solutions to real problems but without real consequences [11]. These games generate positive emotions and enjoyment, boost older adults' self-esteem, and also address aspects essential for improving their quality of life. In addition, games are often employed to prevent or improve conditions associated with aging, such as cognitive decline and dementia. As a result, there has been increasing interest in using digital games to support dementia care, including cognitive rehabilitation games and reminiscence games [8].

Some examples of serious games developed for older adults include Tovertafel and Sea Hero Quest. Tovertafel is an interactive system that projects games onto surfaces, allowing older adults to engage through movement and visual stimuli. It has been widely used in dementia care settings, as it encourages physical activity and enhances social interaction among residents [12]. Sea Hero Quest, on the other hand, is a mobile game designed to collect spatial navigation data to support Alzheimer's research. Through its gameplay, users navigate virtual environments while contributing to the study of cognitive decline [13].

However, despite the growing number of serious games, none to date have been specifically designed for care homes using voice assistants like Alexa, nor co-designed with residents as primary stakeholders. This project addresses that gap by targeting social interaction, accessibility, and reminiscence through conversational technology.

Furthermore, this playful component is beginning to be included in Assistive Technology Systems, especially those that help older adults perform everyday tasks, such as taking medication correctly or preparing lunch. It has been found that if interaction with such technology is poor, complicated, or

tedious, older adults tend to stop using the system. This playful approach may be the most promising in achieving inclusion and accessibility in aging, as the focus is not so much on achieving goals, but on the overall experience and the pleasure derived from attempting to complete game challenges. In this line, conversational Assistants have a great potential to work together with social robot technology to foster relationships and social interactions to reduce people's loneliness and promote their social engagement [14].

Focusing on the use of game-based playful experiences within care homes, according to the systematic literature review conducted in [9], it is concluded that most game systems designed for older adults are intended for home use, while systems developed specifically for nursing homes are scarcely represented in the literature. In fact, in [8], only the use of an affective robot named Matilda is documented; it was brought to a nursing home in Australia and used to play Bingo at the residents' request [15]. Care homes and day centers thus represent environments for older adults where little intervention has occurred so far, despite the fact that they could greatly benefit from the motivation and improved social relationships that gameplay can foster.

In summary, the phenomenon of loneliness in the context of the elderly population in general, and even more so in residences, often leads to the manifestation of problems related to the loss of cognitive faculties, depression and social isolation, which could be mitigated using ICT [16]. With this aim, the main objective of this proposal is to design and implement a serious game using a voice assistant (Alexa) to stimulate autobiographical memory and foster social interaction among older adults living in care homes. This initiative is rooted in the broader goal of promoting healthy aging through technology that is both accessible and meaningful for seniors. More specifically, the project pursues the following aims:

- Understand the concept and value of serious games for elderly users.
- Explore the co-design methodology as a participatory tool to include users' voices in the development process.
- Identify the preferences, interests, and technological barriers faced by older adults.
- Learn to design and develop voice-based applications using Alexa.
- Build a game that facilitates storytelling, reminiscence, and emotional connection among players.

3. Co-design methodology

In order to ensure the serious game is relevant and meaningful for older adults, a participatory co-design methodology was adopted. Sessions were conducted in two nursing homes in the city of Granada (EntreÁlamos and María Auxiliadora), involving both residents and care staff. In each nursing home, four structured sessions were held, each lasting about one hour. Older adults with varying levels of autonomy and cognitive abilities participated, allowing for a wide range of perspectives to be collected.

Participant selection was carried out through initial assessments in collaboration with staff at both care homes. Specifically, the Barthel Index [17] was used to evaluate each person's level of autonomy in performing basic activities of daily living. Based on this index, participants who were sufficiently autonomous to engage in group sessions and game-based activities were selected. A total of 42 individuals were chosen: 14 men and 28 women. Sessions were conducted within the facilities themselves—18 participants from María Auxiliadora and 24 from EntreÁlamos.

Nursing Home	Participants	Men	Women
EntreÁlamos	24	8	16
María Auxiliadora	18	6	12
Total	42	14	28

Table 1

Summary of participants by nursing home and gender.

3.1. Session 1: exploring preferences and attitudes toward technology

The first session aimed to understand the residents' preferences, needs, and attitudes toward technology, particularly virtual assistants like Alexa, to inform the design of a serious game. Individual surveys revealed that while some residents already used smartphones, tablets, or Alexa (one resident used it for music and shopping), others expressed interest but needed adaptations for visual or mobility limitations. During group discussions, key interests emerged: creative activities (embroidery, painting, gardening), reading and poetry (with mentions of Federico García Lorca), classical music (especially the Vienna New Year's Concert), physical exercise (gymnastics, walking), and traditional games (bingo, cards, board games). Many also enjoyed reminiscence activities, sharing past experiences. The session highlighted curiosity about Alexa but emphasized the need for accessible, easy-to-use designs.

3.2. Session 2: introducing Alexa and robots

This session focused on familiarizing residents with Alexa and robots to assess usability and engagement. Participants interacted with two Alexa devices (with and without a screen) and two robots: a large robot (Buddy) and a small robot (Rux). While some struggled with low voices or Alexa misunderstanding commands, they enjoyed asking questions (e.g., about music, animals, or personal trivia) and playing games like "I Spy." Reactions to the robots were mixed—initial surprise turned into curiosity, with residents petting them and asking questions. The larger robot was preferred for its expressive face, while the smaller one was criticized for slow responses (see Figure 4). Key takeaways included the need for clearer voice recognition and higher volume settings, as well as the potential for using Alexa for entertainment and memory exercises.

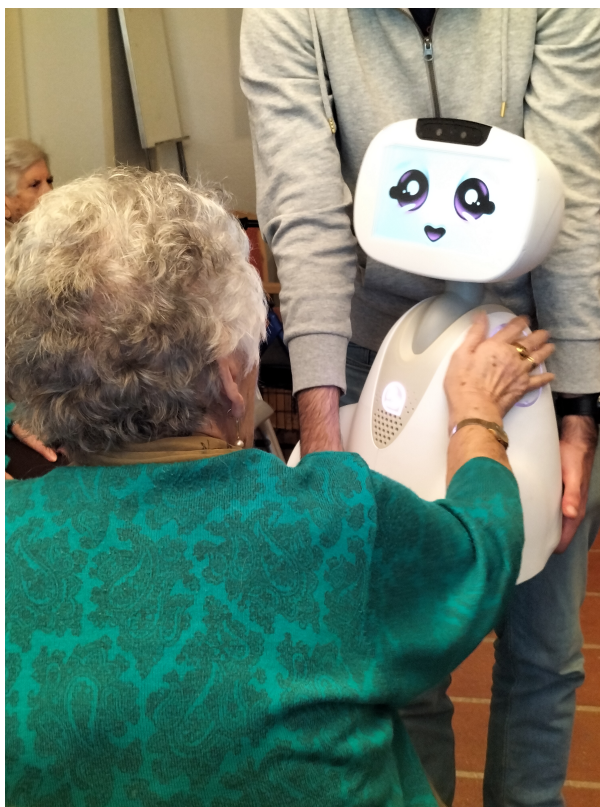


Figure 1: Older adult caressing a robot that simulates emotions

3.3. Session 3: co-designing with images and reflections

In this co-design session, residents discussed how they might use Alexa and robots in daily life and selected meaningful images in small groups (4-6 persons). Many envisioned Alexa as a tool for reminders, historical storytelling, or family communication. Image selections reflected nostalgia—farm animals (especially chickens), landscapes, work tools, and family photos—often tied to personal memories. One resident even crafted a story linking the images. The session reinforced that reminiscence, family, and nature themes should be central to the game, with technology serving as a bridge to emotional connections and shared experiences.

3.4. Session 4: deepening co-design through visual storytelling

The final session expanded on image-based co-design, with residents again choosing pictures that resonated with them (see Figure 2). The residents were divided into subgroups of six or seven persons. Recurring themes included family (photos of children), animals (chickens, pets), travel (Venice, Paris, trains), and nature (the sea, landscapes). These choices underscored a desire for connection—to their past, loved ones, and the wider world. Some expressed unmet wishes, like traveling more, suggesting the game could incorporate virtual exploration. The strong emphasis on storytelling and memory confirmed that a successful design should blend nostalgia, social interaction, and accessible technology to foster engagement and joy. When the residents agreed on the images they liked the most, they made a panel with those images pasted on a poster board (see Figure 3). There was a lot of agreement among the subgroups even though the total number of cards was high (40 cards in total, selecting only 9 cards).



Figure 2: Older adults choosing their favorite cards

This process ensured that the game design was truly aligned with the interests and preferences of the target users, promoting a sense of ownership.

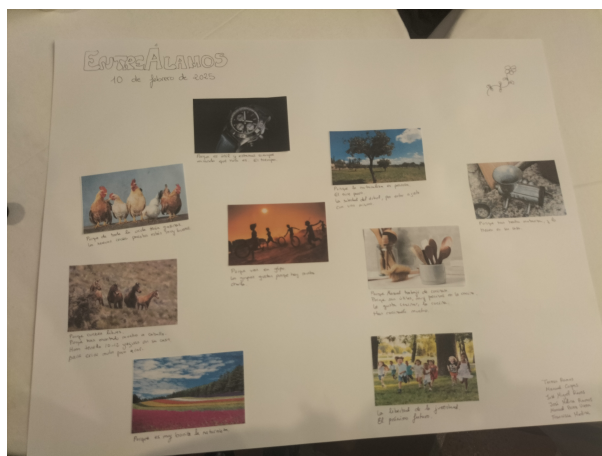


Figure 3: Poster board with the cards chosen by consensus by a subgroup of residents

4. The game “*Return to the past*”

The final design selected for the serious game is *Return to the past*, a memory-based activity where Alexa asks questions about past events, music, films, and traditions. This game was chosen because it closely aligns with preferences identified during co-design sessions, where older adults showed strong engagement with reminiscence, music, and social interaction.

Return to the past combines memory stimulation, visual aids, and group dynamics in a structured yet flexible format. Its intentionally simple and repeatable structure allows it to accommodate different group sizes and varying cognitive abilities.

The game features a non-competitive scoring system that highlights players who share the most memories, helping to avoid stress. Social interaction is encouraged through team questions that promote conversation, while individual turns ensure inclusivity for all participants.

Multimodal support, including images and audio clips, enhances accessibility, particularly for players with hearing or cognitive challenges. Overall, the design blends nostalgia, accessibility, and social engagement to provide a meaningful experience that fosters reminiscence and social bonding in a collaborative and positive environment.

4.1. Question types and adaptive content

Questions were developed through participatory co-design activities with older adults in care homes. Specifically, during the third and fourth co-design sessions, participants collaboratively selected meaningful images and associated themes, which were then used as the basis for question design.

Participants grouped and pasted photographs on large paper boards according to emotional relevance or thematic resonance. A count of the most frequently selected images allowed us to categorize content into several key themes. Table 2 shows the most popular images (based on co-design sessions) and Table 3 shows the grouped categories of images created from the first table.

Based on these clusters, a set of over 60 questions was created, each one linked to an image and designed to stimulate autobiographical memory or general cultural recall. The game design incorporated several key strategies to enhance accessibility, engagement, and emotional relevance:

- **Adaptive questions tied to life experiences:** Designed to engage autobiographical and semantic memory. Example: “What 1960s band was called ‘The Fantastic Four’?” These questions are selected based on themes and preferences identified during co-design sessions.
- **Multimodal support:** Use of images and audio clips to reinforce memory recall. For example, showing a photo of Paris for a question about the “City of Love” helps contextualize and stimulate responses.

Table 2

Count of most popular images based on co-design sessions

Image	Count	Category
Children playing	11	Children
Venice	6	Travel
Animals	6	Animals
Orange tree	5	Landscape
Food	4	Food
Cruise	4	Travel
Paris	3	Travel
Doorknob	3	Antique object
House in China	3	Travel
Green landscape	3	Landscape
Train	3	Travel
Games by the fireplace	2	Travel
Beach	2	Landscape
Colorful landscape	2	Landscape
Kitchen utensils	2	Work
Old telephone	1	Antique object
Mobile phone	1	Technology
Robot	1	Technology
Swing	1	Children
Sport	1	Physical activity
Car	1	Travel
Medicine	1	Medicine
Clock	1	Antique object
Tractor	1	Work

Table 3

Grouped image categories and total counts

Category	Total Count
Travel	19
Landscapes	12
Children	12
Animals	6
Food	4
Antique objects	4
Work	3
Technology	2
Physical activity	1
Medicine	1

- **Social interaction:** Team-based questions encourage conversation and group bonding, while individual turns ensure that all participants can contribute and feel included.
- **Non-competitive scoring:** The game uses a positive ranking system that highlights players who shared the most memories, avoiding stress and promoting a supportive atmosphere.
- **Question categories:**
 - *Cultural knowledge questions:* Target semantic memory. Example: "Which city is known as the 'City of Love'?" (Image: Eiffel Tower).
 - *Personalized memory prompts:* Encourage autobiographical storytelling. Example: "What food reminds you of your childhood?" (Image: traditional meal).
 - *Media and pop culture references:* Designed to trigger group reminiscence. Example: "What was the name of the show featuring characters Espinete and Don Pimpón?" (Image: Espinete).

Based on the co-design sessions, the most frequently selected images were related to children playing

(11 mentions), Venice (6), and animals (6). Other popular choices included orange trees (5), cruises (4), and various food-related images. When grouped by category, the most common themes were travel (19 images), landscapes (12), and children (12).

These preferences indicate strong emotional or experiential connections with themes like travel, nature, and childhood (because it reminds them of family). Therefore, the game will primarily focus on question prompts related to these categories to maximize engagement, familiarity, and reminiscence.

Here is a simple schema of how the game works (Figure 4) It outlines the main stages of interaction between Alexa and the elderly users, including how questions are managed, how the system handles turns, and how players can exit the game if they choose to. This schema helps visualize the core logic behind the conversational structure:

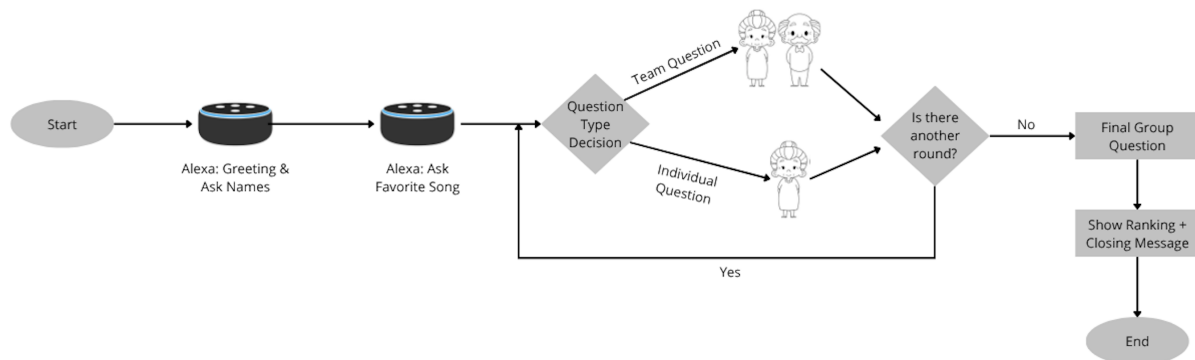


Figure 4: Diagram representing the interaction between older adults and Alexa in the game

5. Technical implementation

The Alexa skill “*Return to the Past*” was developed combining Amazon’s voice services and AWS cloud infrastructure. The architecture was specifically designed to address the needs of elderly users through natural language processing and accessible interaction patterns.

5.1. Interaction model design

The voice interaction model combines custom and built-in components to create a fluid user experience:

- **Custom intents** form the core game mechanics:
 - AnswerIntent handles trivia responses with multiple acceptable answer variations
 - GetFavoriteSongIntent processes musical preferences using natural language patterns
 - PlayerCountIntent manages group dynamics with numeric validation
- **Slot types** enhance recognition accuracy:
 - AMAZON.SearchQuery accommodates diverse response phrasing
 - AMAZON.FirstName recognizes names with cultural variations
 - Contextual help system using AMAZON.HelpIntent
- The utterance library includes 50+ variations per intent, accounting for regional dialects and speech patterns common among elderly users.

5.2. Backend infrastructure

The architecture was built on AWS services to ensure scalability and reliability. At the core, Node.js Lambda functions handle all game logic and state transitions, maintaining session continuity through DynamoDB persistence.

For data management, the system employs DynamoDB tables optimized for fast key-value access patterns. The *JuegoRegresoPasado* table serves as the primary data store, storing real-time player information, game states, and scoring. A secondary *CancionesRegresoPasado* table functions as a media catalog, storing metadata and references to audio files in S3. This separation of concerns allows for efficient data retrieval while maintaining clean architectural boundaries.

Media handling follows strict technical requirements to ensure optimal voice interaction quality. All audio clips are processed to meet Alexa’s specifications (24kHz/128kbps MP3 format) and stored in S3 buckets with appropriate access policies. The system leverages SSML (Speech Synthesis Markup Language) to dynamically embed these audio elements into responses, creating a more engaging user experience. Content delivery is further optimized through regional endpoint selection, minimizing latency for voice interactions [18].

5.3. Deployment and voice

The deployment pipeline integrates multiple specialized components to ensure consistent performance. Amazon Polly’s voices provide the foundation for natural-sounding responses, with additional logic for time-aware tone modulation (adjusting cheerfulness for morning/evening/night interactions) and adaptive speech rates based on context. This voice personality system helps create a more intuitive and comforting experience for elderly users [19].

Security is enforced through configured IAM roles following the principle of least privilege.

Accessibility was a primary consideration throughout development. The implementation includes multimodal feedback channels that combine voice responses with visual display content. A progressive help system provides contextual assistance when needed, while carefully designed error recovery paths handle common misunderstandings gracefully. These features work together to create an inclusive experience that accommodates the diverse needs of elderly users.

5.4. Why “*Return to the past*” is a serious game

Return to the past is not merely a leisure activity, but a carefully crafted serious game—one that integrates entertainment with meaningful, non-entertainment goals. It was designed to address specific psychological, social, and cognitive needs of older adults through play. Serious games are those in which the primary purpose is not solely amusement, but rather education, training, therapy, or social impact.

In this case, the game targets key challenges commonly faced by older adults, such as memory decline, social isolation, and low digital inclusion. Through carefully selected questions and mechanics, it promotes autobiographical storytelling, reinforces identity, and creates space for meaningful interpersonal interaction.

The game qualifies as “serious” for several reasons:

- **Therapeutic intent:** It draws on reminiscence therapy principles, encouraging players to recall and share personal memories that are emotionally significant.
- **Structured intervention:** Unlike open-ended or casual games, it has a predefined framework with purposeful prompts, progression logic, and reflective closure.
- **Designed with a target population in mind:** The co-design sessions ensured that the game content, pace, and interface were adapted to the interests, capabilities, and limitations of older users in care homes.
- **Technology as facilitator:** By integrating Alexa as a voice assistant, the game also serves a digital literacy function, helping seniors interact with accessible technologies in a low-pressure setting.
- **Dual outcomes:** The game stimulates cognitive processes (memory, attention, language) while also fostering emotional well-being and social bonding, all embedded within a pleasurable game format.

The game's simplicity is intentional, enabling older adults with varying cognitive abilities to participate. At the same time, each element is purposefully designed to serve a dual function: creating a fun and engaging experience while supporting measurable psychological and social benefits.

The following table illustrates how each serious objective is addressed through specific ludic strategies within the game.

Table 4

Alignment between serious objectives and game mechanics

Serious objective	Related game mechanic
Stimulate autobiographical memory	Questions about cultural events, music, and popular places from the past, supported with images and sound clips
Reduce social isolation	Team-based questions, group interactions, and introductory questions to promote personal connections
Promote accessible technology use	Voice interface via Alexa, no need for advanced digital skills
Reinforce self-esteem through positive memories	Continuous positive feedback from Alexa and end-of-game recognition based on participation
Encourage mutual knowledge among residents	Personal prompts like "What's your favorite song?" to foster conversation
Provide an inclusive and accessible experience	Multimodal approach (audio-visual), adaptive timing, and clear instructions
Support active and meaningful participation	Turn-based structure, role variety (individual vs. team), and player-driven input

By fulfilling these characteristics, *Return to the past* aligns with the definition of serious games and extends the potential of gaming to serve vulnerable populations through accessible, inclusive, and emotionally enriching experiences.

Although the game has not yet undergone formal evaluation, a series of user experiences in the two care homes "Maria Auxiliadora" and "EntreÁlamos" are planned to assess its effectiveness. These evaluations will provide valuable feedback to refine both the content and the game dynamics.

5.5. Limitations

This study presents several limitations that should be considered when interpreting the results and evaluating the generalizability of the proposed game.

First, the sample is not gender-balanced: out of 42 participants, 28 were women and 14 were men. While this partially reflects the demographics of residential care settings, where women tend to be the majority due to longer life expectancy [1], it may also influence the preferences expressed during the co-design sessions. For instance, the emotional tone of the selected images, themes of family and reminiscence, and interest in music and storytelling could be more representative of female perspectives [20].

Second, as discussed in Section 4.2, the game was intentionally designed to be simple and accessible to accommodate users with different levels of cognitive and sensory abilities. However, this simplicity may limit engagement for more independent or technologically experienced users. Moreover, initial sessions revealed technical issues such as Alexa's difficulty recognizing low-volume speech or group responses, which affected the flow of interaction.

Finally, the cultural and geographical context of the study, focused on two nursing homes in southern Spain, means that preferences may vary in other regions or settings.

5.6. Ethical considerations

The use of voice assistants and social technologies in care homes raises important ethical considerations, particularly when engaging with older adults who may experience cognitive decline or sensory

impairments. While the game *Return to the past* was always introduced as a voice-based tool moderated by a non-human assistant (Alexa), future autonomous use of such systems requires careful reflection on transparency, user understanding, and emotional safety.

During the co-design sessions, participants were explicitly informed that Alexa is not a human being, but a programmed device designed to facilitate interaction. However, as conversational agents become more sophisticated and natural-sounding, the line between human and machine interaction can become blurred. This evolution raises important considerations about how such interactions might impact the dignity and identity of older persons, especially in eldercare settings. There are potential effects of robot and conversational agent interactions on dignity. Robots could negatively (or positively) impact on people's dignity of identity, and on their feelings of self respect or of humiliation. As interaction capabilities improve, there is a risk that older adults may develop attachments or form perceptions of personalized relationships with these devices, which could influence their sense of self and social identity, underscoring the importance of respectful design and implementation of such technologies [21].

To address this, any future deployment of the system in autonomous settings should:

- Respect the inviolable dignity of individuals by designing robots that avoid objectification and promote self-respect [21].
- Support a person's Dignity of Identity by ensuring robot interactions preserve autonomy, self-esteem, and a sense of control [21].
- Enhance, rather than replace, human interaction, ensuring robots complement social relationships and emotional well-being [21].
- Use technology ethically and sensitively to prevent humiliating or demeaning treatment that could harm dignity [21].
- Assess the impact on capabilities and dignity before deploying robots, ensuring they promote a dignified life [21].

These tools should supplement, not replace, human interaction. While serious games and digital companions can promote social engagement, they should always be embedded in a broader context of interpersonal support and monitored use.

The design and deployment of such systems must respect users' autonomy and dignity, while safeguarding against emotional manipulation or misunderstanding.

6. Conclusions and further works

The co-design methodology proved to be a key factor in the development of a game that is not only functional but also emotionally resonant and socially engaging. By involving older adults directly in the design process, the project ensured that the game reflected their real preferences, needs, and life experiences.

Participants contributed valuable insights that shaped the content and structure of the game. Through the selection of meaningful images and open conversations, themes such as family, nature, music, and memories of childhood and work emerged as emotionally significant. These themes were later embedded into the gameplay and question design, reinforcing the personal relevance of the experience.

The project also provided useful observations on the interaction between older adults and technology. While some participants initially expressed hesitation or experienced difficulties using voice interfaces—such as low speech volume or Alexa not recognizing group speech—many demonstrated curiosity and a willingness to learn. A few already had experience with Alexa, and most enjoyed the interaction once familiarized with it.

Preferences clearly leaned toward non-competitive, narrative-driven games that promote conversation and reminiscence. Musical cues, storytelling elements, and cooperative dynamics were especially appreciated. These findings validate the design choices made and confirm the importance of adapting digital experiences to the cognitive and emotional context of older adults.

Our first line of future work will be the full implementation of a functional version of the application that can be brought to nursing homes for use. During these sessions, an observation process will be carried out, followed by the administration of a simple usability test adapted for this demographic. Additionally, another test will be conducted to evaluate the entertainment generated and any improvement in social relationships among participants. This will allow us to validate the usefulness of the proposal at both the individual and group levels.

In summary, the project not only resulted in a serious game prototype, but also highlighted the value of participatory design and voice-based interaction as tools for inclusive and meaningful digital experiences for the elderly.

Acknowledgements

This work has been supported by Grant C-ING-179-UGR23, funded by the Consejería de Universidad, Investigación e Innovación and by the ERDF Andalusia Program 2021–2027.

It has also been supported by the PLEISAR-Social project, Ref. PID2022-136779OB-C33, funded by the Ministry of Science and Innovation of Spain (MCIN/AEI/10.13039/501100011033, UE).

Declaration on generative AI

During the preparation of this work, the author(s) used X-GPT-4 and Gramby in order to: Grammar and spelling check.

References

- [1] Eurostat, Demography of europe — 2023 interactive edition, 2023. URL: <https://ec.europa.eu/eurostat/web/interactive-publications/demography-2023#ageing-population>.
- [2] A. Reis, D. Paulino, H. Paredes, J. Barroso, Using intelligent personal assistants to strengthen the elderlies' social bonds: A preliminary evaluation of amazon alexa, google assistant, microsoft cortana, and apple siri, in: Universal Access in Human–Computer Interaction. Human and Technological Environments: 11th International Conference, UAHCI 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9–14, 2017, Proceedings, Part III 11, Springer, 2017, pp. 593–602.
- [3] S. Dogra, D. W. Dunstan, T. Sugiyama, A. Stathi, P. A. Gardiner, N. Owen, Active aging and public health: evidence, implications, and opportunities, *Annual review of public health* 43 (2022) 439–459.
- [4] E. Rudnicka, P. Napierała, A. Podfigurna, B. Męczekalski, R. Smolarczyk, M. Grymowicz, The world health organization (who) approach to healthy ageing, *Maturitas* 139 (2020) 6–11.
- [5] R. P. De Lope, N. Medina-Medina, A comprehensive taxonomy for serious games, *Journal of Educational Computing Research* 55 (2017) 629–672.
- [6] A. Abd-Alrazaq, M. Alajlani, D. Alhuwail, C. T. Toro, A. Giannicchi, A. Ahmed, A. Makhoulf, M. Househ, The effectiveness and safety of serious games for improving cognitive abilities among elderly people with cognitive impairment: systematic review and meta-analysis, *JMIR serious games* 10 (2022) e34592.
- [7] G. Dewsbury, I. Sommerville, P. Bagnall, M. Rouncefield, V. Onditi, Software co-design with older people, in: *Designing accessible technology*, Springer, 2006, pp. 199–208.
- [8] C. Dormann, Toward ludic gerontechnology: a review of games for dementia care, in: *1st International Joint Conference of DiGRA and FDG*, 2016.
- [9] D. Martinho, J. Carneiro, J. M. Corchado, G. Marreiros, A systematic review of gamification techniques applied to elderly care, *Artificial Intelligence Review* 53 (2020) 4863–4901.
- [10] J. Chipia Lobo, Juegos serios: Alternativa innovadora, *Revista Conocimiento Libre y Educación* 1 (2011) 1–18.

- [11] C. López Raventós, El videojuego como herramienta educativa. posibilidades y problemáticas acerca de los serious games, *Apertura* (Guadalajara, Jal.) 8 (2015) 15.
- [12] Tover Health Tech B.V., Tovertafel for people living with dementia, 2023. URL: https://www.youtube.com/watch?v=2UB6E1K1q_4.
- [13] Alzheimer's Research UK, Sea hero quest, 2016. URL: <https://www.alzheimersresearchuk.org/research/for-researchers/resources-and-information/sea-hero-quest/>.
- [14] A. K. Ostrowski, H. W. Park, C. Breazeal, Design research in hri: Roboticists, design features, and users as co-designers, in: *Workshop on Designerly HRI Knowledge*, 2020.
- [15] R. Khosla, M.-T. Chu, K. Nguyen, Affective robot enabled capacity and quality improvement of nursing home aged care services in Australia, in: *2013 IEEE 37th annual computer software and applications conference workshops*, IEEE, 2013, pp. 409–414.
- [16] Q. Fan, Utilizing ict to prevent loneliness and social isolation of the elderly. a literature review, *Cuadernos de trabajo social* 29 (2016) 185–200.
- [17] F. I. Mahoney, D. W. Barthel, Functional evaluation: the barthel index, *Maryland State Medical Journal* 14 (1965) 61–65.
- [18] Amazon Alexa, Speech synthesis markup language (ssml) reference, 2023. URL: <https://developer.amazon.com/en-US/docs/alexa/custom-skills/speech-synthesis-markup-language-ssml-reference.html>.
- [19] Amazon Web Services, Amazon polly, 2022. URL: <https://aws.amazon.com/es/polly/>.
- [20] R. Sánchez Aragón, B. E. Retana Franco, E. Carrasco Chávez, Evaluación psicológica del entendimiento emocional: Diferencias y similitudes entre hombres y mujeres, *Revista Iberoamericana de Diagnóstico y Evaluación - e Avaliação Psicológica* 2 (2008) 193–216. URL: <http://www.redalyc.org/articulo.oa?id=459647347010>.
- [21] A. Sharkey, Robots and human dignity: a consideration of the effects of robot care on the dignity of older people, *Ethics and Information Technology* 16 (2014) 63–75. URL: <https://doi.org/10.1007/s10676-014-9338-5>.

EXPLORANDO LAS CAPACIDADES COMUNICATIVAS Y LA SEMIÓTICA DEL DISEÑO SIMPLIFICADO Y LOW POLY: EL CASO DE INDAGAPP

Roberto Martín González; Universidad de Burgos, España

Iñigo Gómez Zapata; Universidad de Burgos, España

Mario Alaguero Rodríguez; Universidad de Burgos, España

Autor de correspondencia: Roberto Martín González; roberto.martin@ubu.es

1. RESUMEN

Este trabajo explora cómo el diseño visual de los personajes y el uso de elementos narrativos inciden en la comprensión y memorización de contenidos en aplicaciones educativas centradas en la metodología STEM. El caso de estudio es “IndagApp”, una aplicación móvil en forma de *serious game* que propone distintos experimentos interactivos como herramienta de apoyo para la enseñanza de las ciencias.

Cada experimento está acompañado por breves relatos protagonizados por un personaje diseñado con una estética basada en *low-poly* y *voxel-art*. Esta elección visual no solo favorece el rendimiento técnico de la app, también busca generar una mayor conexión emocional a través de la sintetización formal de expresiones y el uso de signos simplificados, traducándose en una mejor retención del conocimiento.

La propuesta plantea una vía de investigación sobre el potencial de las narrativas y los recursos visuales estilizados para potenciar la motivación, la comprensión y la memoria en entornos de aprendizaje científico.

Keywords. Semiótica, educación, estética

2. INTRODUCCIÓN

Las aplicaciones interactivas con fines educativos integran recursos propios de la gamificación y el aprendizaje basado en juegos, como la posibilidad de desarrollar narrativas complejas y utilizar estéticas estilizadas o no realistas. Este estudio aborda el impacto de la construcción semiótica de personajes en *serious games* destinados a la enseñanza STEM, enfocándose en cómo la simplificación visual —a través de estéticas como el *low-poly*— puede potenciar la claridad y efectividad comunicativa del entorno educativo.

A diferencia de enfoques que analizan videojuegos ya existentes para evaluar su valor didáctico, este trabajo se centra en el proceso de desarrollo, tomando como ejemplo la creación del personaje “Ramona” en la aplicación educativa interactiva *IndagApp*. Más allá de su función narrativa y motivacional, Ramona actúa como un elemento mediador que sintetiza información visual y orienta al usuario en la experiencia de aprendizaje (Yáñez-Pérez et al., 2024a).

Desde una perspectiva semiótica, la elección de un diseño simplificado responde a la necesidad de reducir el ruido visual y enfocar la atención del usuario en los elementos clave. La representación mediante formas básicas facilita la decodificación rápida de información, permitiendo una mayor accesibilidad cognitiva y emocional. A diferencia de estéticas hiperrealistas, que pueden saturar o distraer, la síntesis formal promueve una identificación más ágil y una interacción más efectiva.

Durante la validación cualitativa de *IndagApp*, expertos en didáctica destacaron cómo esta estética actúa como soporte visual eficaz para la comprensión de los experimentos. Uno de los participantes señaló que “proporciona apoyo visual y atractivo donde puedes ver los resultados y simulaciones de diferentes experimentos, lo cual sería mucho menos atractivo en la vida real” (Yáñez-Pérez et al., 2024b). Esta percepción refuerza la hipótesis de que la simplificación formal no solo facilita la comunicación visual, sino que también incrementa el interés y la retención del conocimiento en contextos educativos gamificados.

En este trabajo exploramos la importancia de la semiótica aplicada durante las fases de diseño y desarrollo de personajes y narrativas complementarias para la enseñanza científica en la educación secundaria. Asumiendo como limitación la ausencia de una validación experimental, esta fase preliminar abre las puertas a futuras líneas de investigación.

3. ESTADO DE LA CUESTIÓN

El desarrollo del potencial didáctico de los videojuegos ha ido evolucionando de la mano de los avances tecnológicos y narrativos de la industria del entretenimiento, en paralelo con una creciente atención desde el ámbito académico. Los estudios de comunicación y semiótica han explorado las capacidades de expresión en la estética y cómo esto afecta a la transmisión de información y emociones de los autores del videojuego a los jugadores (Pérez-Latorre et al., 2016). En el contexto de la educación STEM, diversas investigaciones avalan la eficacia de estos recursos lúdicos como herramientas pedagógicas. Un estudio comparativo realizado por Michaela Arzmann, Lisette Hornstra, Johan Jeuring y Liesbeth Kester evidenció que los estudiantes que participaron en intervenciones educativas gamificadas obtuvieron mejores resultados académicos, mostraron una mayor motivación y presentaron actitudes más positivas frente al aprendizaje que aquellos expuestos a métodos tradicionales (2022).

El diseño de videojuegos educativos requiere la integración cuidadosa de principios pedagógicos, lo cual permite generar experiencias de aprendizaje más inmersivas, motivadoras y efectivas (Peralta Santa Cruz, 2023). Sin embargo, el verdadero potencial de un *serious game* solo se concreta si existe un diseño bien articulado, que logre equilibrar mecánicas (reglas del sistema lúdico), dinámicas (interacciones del jugador) y estética (respuesta emocional generada por la experiencia visual y narrativa) (Hunicke et al., 2005; Winn, 2008).

En este marco, la inclusión de personajes como herramienta narrativa y pedagógica ha sido ampliamente utilizada en distintos medios educativos, desde series animadas hasta libros escolares. Estos personajes no solo captan la atención del estudiante, sino que también favorecen la memoria y reducen la tasa de olvido de los contenidos (Fukushima et al., 2024). Sin embargo, el diseño de personajes humanos con alto grado de realismo puede generar el conocido “valle inquietante”, un fenómeno en el que la apariencia casi humana provoca incomodidad o rechazo en el usuario (Lugrin et al., 2015). Es preferible integrar personajes carismáticos y con atractivo para los usuarios para fomentar el aprendizaje y motivación, siendo preferible, por ejemplo, optar por el diseño de animales antropomorfizados (Solorzano et al., 2025).

Desde la semiótica del diseño visual, se plantea que la abstracción en la construcción de personajes permite una comunicación más eficaz de significados. Al reducir el nivel de detalle, se potencian los rasgos esenciales y se facilita la interpretación emocional. Como demuestran Sharikov et al. (2024), “el uso de colores primarios y formas sencillas permite a los diseñadores crear personajes con un impacto emocional más predecible, lo que es crucial para la narración y la participación de los usuarios en el arte digital y los videojuegos”. Esta simplificación no implica pérdida de expresividad, sino una mayor

claridad semiótica que optimiza la transmisión de información visual y contribuye a una experiencia educativa más coherente y efectiva.

4. METODOLOGÍA

Este estudio explora cómo un recurso clásico en la enseñanza, como lo son los personajes de ficción, puede ser reinterpretado y adaptado a entornos digitales mediante el uso de tecnologías interactivas, apoyándose en hallazgos recientes del campo de los videojuegos aplicados a la educación. A través de un caso práctico concreto —la aplicación interactiva *IndagApp*, desarrollada en el ámbito universitario— se propone una aproximación metodológica basada en el diseño iterativo y la experimentación controlada.

La creación de *IndagApp* permite realizar una investigación aplicada que no se limita a evaluar un producto terminado y que considera su desarrollo como parte activa del proceso científico. Esto posibilita analizar el impacto de variables específicas —en este caso, la narrativa mediada por personajes— dentro del contexto educativo STEM, y establecer bases para futuros estudios que examinen otras dimensiones del diseño didáctico.

Esta primera fase de investigación se enfoca en el rol de los personajes dentro de aplicaciones científicas, con especial atención a cómo una estética visual simplificada puede facilitar la empatía del usuario, así como la comprensión y la retención de conceptos complejos. Para ello, se diseñó un personaje siguiendo criterios definidos por un equipo interdisciplinar de docentes, pedagogos e investigadores en educación. Este personaje fue integrado tanto en las cinemáticas como en la narrativa general de la aplicación.

La incorporación del personaje al motor de desarrollo del videojuego permitió una primera validación por parte del equipo de trabajo, con el objetivo de valorar si sus características visuales, narrativas y funcionales están alineadas con los propósitos pedagógicos de la herramienta. Este estudio constituye así el punto de partida para una línea de investigación más amplia que buscará determinar qué elementos del diseño contribuyen de forma significativa al potencial didáctico de *IndagApp*.

5. RESULTADOS

IndagApp es una aplicación educativa desarrollada para dispositivos móviles y tabletas, cuyo objetivo es acercar a estudiantes de nivel secundario a conceptos científicos mediante simulaciones virtuales de experimentos. Esta herramienta propone una primera aproximación visual basada en las estéticas *low-poly* y *voxel art*, con la intención de ser comparada, en futuras investigaciones, con otras propuestas estilísticas.

Ambos estilos, nacidos inicialmente como respuesta a las limitaciones técnicas de las primeras consolas con gráficos tridimensionales, se consolidaron con el tiempo como lenguajes visuales propios. Su simplicidad estructural —caracterizada por formas geométricas básicas y volúmenes reducidos— no solo facilita la optimización del rendimiento, sino que también potencia la capacidad comunicativa del diseño. Desde una perspectiva semiótica, esta reducción formal permite transmitir información visual de manera más clara y eficaz.

En *IndagApp*, cada simulación experimental está integrada en una narrativa accesible y dinámica, orientada a facilitar la comprensión de conceptos científicos y terminología especializada. Esta narrativa gira en torno a Ramona, un personaje diseñado específicamente para acompañar al estudiante a lo largo de los experimentos. Su presencia constante en todas las actividades contribuye a la cohesión del relato y actúa como

mediadora entre el contenido científico y la experiencia del usuario, reforzando la conexión emocional y el aprendizaje significativo.

Figura 1: Selección de expresiones del personaje Ramona en un tutorial del videojuego



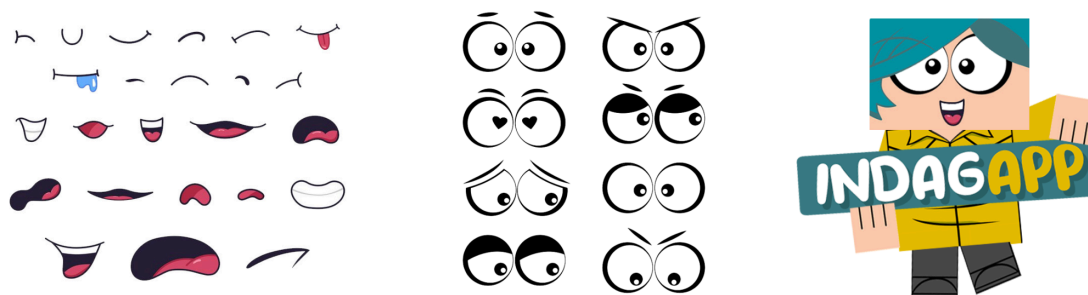
Fuente: Elaboración propia

La estética adoptada en *IndagApp* combina elementos del *low-poly*, el *voxel art* y el estilo *cartoon*, configurando un lenguaje visual simplificado que permite una mayor expresividad del personaje. Esta reducción formal no implica una pérdida de significado, sino todo lo contrario: al eliminar detalles superfluos, se refuerzan aquellos elementos esenciales que transmiten emociones de manera más directa y efectiva.

En este contexto, las emociones del personaje se comunican a través de signos visuales codificados —rasgos faciales esquemáticos, posturas corporales exageradas, y movimientos estilizados— que remiten a convenciones consolidadas en medios como el cómic y la animación. La expresión emocional no se representa de forma naturalista; está basada en un sistema simbólico que se apoya en patrones culturales ampliamente reconocidos y fácilmente interpretables.

Esta codificación simplificada facilita una lectura inmediata del estado emocional del personaje, favoreciendo así la conexión del usuario con la narrativa. Al reducir las formas a lo esencial, se maximiza su eficacia comunicativa: cada gesto, mirada o postura se convierte en un signo cargado de significado, fácilmente reconocible y emocionalmente potente. Esta estrategia visual refuerza la función narrativa del personaje dentro de la aplicación, facilitando la comprensión de los procesos científicos presentados y manteniendo el interés del estudiante a través de una relación empática sostenida por la expresividad clara y accesible del diseño.

Figura 2: Elementos faciales convertidos en signos a través de la simplificación



Fuente: Elaboración propia

Figura 3: Pantalla de la aplicación interactiva en la que se muestra a Ramona empujando una caja en un experimento relacionado con la fuerza



Fuente: Elaboración propia

Una vez integrado el personaje, tanto en la narrativa como en el desarrollo de los experimentos, se llevó a cabo una validación interna a cargo del equipo de expertos en didáctica responsable del diseño conceptual del juego y de cada uno de los experimentos científicos. Estos especialistas valoraron positivamente la incorporación del personaje, considerándola coherente con los objetivos pedagógicos de la aplicación. Esta primera validación sienta las bases para futuros estudios que aplicarán metodologías específicas de investigación educativa, con el fin de evaluar el impacto del personaje en contextos reales de enseñanza, utilizando como muestra a estudiantes de nivel secundario.

6. CONCLUSIONES

El uso de personajes y narrativas en aplicaciones interactivas puede enriquecer significativamente su capacidad didáctica, especialmente cuando se recurre a estéticas visuales como el *low-poly*, el *voxel art* y el estilo *cartoon*. Estas formas de representación, basadas en la simplificación geométrica y cromática, tienen potencial para optimizar el rendimiento técnico y la expresividad de los personajes al condensar emociones y actitudes en signos visuales claros y fácilmente interpretables.

La expresividad que permite esta síntesis visual se convierte en una herramienta clave para la narrativa educativa, ya que refuerza la comprensión emocional de los contenidos y favorece la conexión del usuario con el entorno de aprendizaje. En este contexto, los principios de la gamificación y el aprendizaje basado en juegos encuentran un terreno fértil para aplicarse en recursos didácticos destinados al aula. No obstante, para que estas herramientas logren su máximo potencial, es fundamental que su diseño esté acompañado de estudios experimentales desarrollados durante el proceso de creación, más allá de los análisis que se realizan una vez el producto está en uso.

Este caso de estudio en fase preliminar propone una vía de investigación centrada en cómo la estética simplificada puede mejorar la eficacia narrativa y comunicativa de las aplicaciones educativas STEM. Se abre así una línea de investigación que permitirá realizar una nueva evaluación con estudiantes que permita comparar la eficacia del diseño simplificado frente a otros estilos más realistas. Al tratarse de un entorno controlado, es posible modificar variables concretas y obtener resultados concluyentes sobre el valor didáctico de una estética que, lejos de restar realismo, amplifica la capacidad de transmitir significado.

7. BIBLIOGRAFÍA

- Arztmann, M., Hornstra, L., Jeurig, J., & Kester, L. (2022). *Effects of games in STEM education: a meta-analysis on the moderating role of student background characteristics*. *Studies in Science Education*, 59(1), 109–145. <https://doi.org/10.1080/03057267.2022.2057732>
- Fukushima, S., Sakamoto, K., & Nakamura, Y. (2024). *NariTan: Enhancing Second Language Vocabulary Learning Through Non-Human Avatar Embodiment in Immersive Virtual Reality*. *Multimodal Technologies and Interaction*, 8(10), 93. <https://doi.org/10.3390/mti8100093>
- Hunicke, R., LeBlanc, M., & Zubek, R. (2004, July). *MDA: A formal approach to game design and game research*. In *Proceedings of the AAAI Workshop on Challenges in Game AI* (Vol. 4, No. 1, p. 1722).
- Peralta Santa Cruz, C. (2023). *Gamificación en la educación: experiencias de aplicación*. *Apuntes De Ciencia & Sociedad*, 11(1), 102-111. <https://doi.org/10.18259/acs.2023011>
- Sharikov, T. & Mayer, I. & Hassreiter, S. (2024) *The Impact of Emotional Design Features on Character Perception*. Breda University of Applied Sciences. <https://doi.org/10.13140/RG.2.2.24261.87529>
- Solorzano Alcivar, N., Park Kim, D. H., Rubio Zurita, A., & Elizalde Ríos, E. (2025). Estrategias de diseño en la creación de personajes para videojuegos educativos infantiles: caso MIDI-Musical. *Grafica*, 13(25), 187–199. <https://doi.org/10.5565/rev/grafica.374>
- Lugrin, J. L., Latt, J., & Latoschik, M. E. (2015, March). *Avatar anthropomorphism and illusion of body ownership in VR*. In *2015 IEEE virtual reality (VR)* (pp. 229-230). IEEE. <https://doi.org/10.1109/VR.2015.7223379>
- Pérez-Latorre, O. Oliva, M. Besalú, R. (2016). *Videogame analysis: a social-semiotic approach*. *Social Semiotics* 27(5):1-18. 10.1080/10350330.2016.1191146
- Winn, B. M. (2008). *The design, play, and experience framework*. *Handbook of Research on Effective Electronic Gaming in Education*. Philadelphia: IGI Global Publication, 1010-1024. <https://doi.org/10.4018/978-1-59904-808-6.ch058>
- Winn, B. & Heeter, C. (2006). *Resolving Conflicts in Educational Game Design through Playtesting*. *Innovate: Journal of Online Education*, 3(2). Retrieved May 23, 2025 from <https://www.learntechlib.org/p/104284/>
- Yáñez-Pérez, I., Toma, R. B., & Meneses Villagrà, J. Á. (2024a). *Diseño y usabilidad de IndagApp: una app para la enseñanza de las ciencias por indagación*. *Revista Iberoamericana de Educación a Distancia (RIED)*, 27(2), 223-245. <https://doi.org/10.5944/ried.27.2.39109>
- Yáñez-Pérez, I., Toma, R. B., & Meneses-Villagrà, J. Á. (2024b). *The IndagApp mobile app: an inquiry-based science teaching resource: usability evaluation with pre-service teachers*. *Journal of New Approaches in Educational Research*, 13(1), 7. <https://doi.org/10.1007/s44322-024-00008-7>

8. NOTAS

Esta investigación se engloba en las actividades del proyecto STEM-V financiado por la Agencia Estatal de Investigación de España (PID2020-117348RB-I00) y coordinado por la Universidad de Burgos.

El simbolismo grotesco surrealista en el diseño de videojuegos experimentales

Anabel Cambero*1, Ismael Sagredo-Olivenza2, Lara Sanchez-Coterón1

1Departamento de Diseño e Imagen, Universidad Complutense de Madrid, Facultad de Bellas Artes, Madrid, España.

2Departamento de Ingeniería del Software e Inteligencia Artificial, Universidad Complutense de Madrid, Facultad de Informática, Madrid, España.

*Corresponding author(s). E-mail(s): ancamber@gmail.com

Abstract

El simbolismo en los videojuegos es algo muy presente en multitud de juegos, tanto independientes como comerciales. Pero hay pocos trabajos donde se estudie la relación que tienen este simbolismo, en concreto el simbolismo grotesco, unido a las mecánicas de un videojuego, para influir en una manera determinada en la experiencia jugable. El presente artículo trata de esbozar una metodología y un diseño experimental sobre el cual, mediante un videojuego como prototipo, podamos explorar diferentes formas de influir en la experiencia lúdica, y si estas son efectivas o no, a través de características simbólicas propias de lo grotesco surrealista, el eroguro y el body horror, tratando de crear una experiencia inquietante pero a la vez recreativa. Se busca examinar si la simbología no solo enriquece la narrativa o el arte visual, sino también desafiar percepciones y fomentar la reflexión crítica.

Keywords: Simbolismo, videojuego, diseño experimental, grotesco, body horror, eroguro

1. Introducción

El universo simbólico construye la realidad, desplazando y suplantando lo físico. No obstante, aunque el mundo existe antes del lenguaje, resulta imposible concebir un espacio anterior al código simbólico, ya que en este marco la realidad emerge del relato. Así, las sociedades articulan un sistema de interpretación en la intersección entre lo simbólico individual y lo social: el juego [10].

Dentro de este marco, el juego presenta diversas manifestaciones donde la humanidad genera símbolos. Esto nos lleva a reflexionar sobre cómo operan ciertos aspectos simbólicos representados en múltiples juegos. Dada su relevancia, este trabajo se centra en lo grotesco por su vínculo esencial con lo simbólico y a la dinámica lúdica.

Para entender la relación entre simbolismo, juego y grotesco, citamos a la autora Frances S. Connelly, quien identifica lo grotesco como *juego con la imagen* o *Spielraum* (Cuarto de juegos, sin traducción directa), un funcionamiento esencial que asocia simbologías de situaciones o acciones. Según Connelly, es “más un verbo que un nombre” (Connelly, 2012, p.25), pues rompe los límites culturales, contradice lo conocido, lo apropiado o lo normal, vinculándose estrechamente al cuerpo y subvirtiendo las reglas establecidas. Además, “al amenazar nuestra identidad al romper los límites entre el ser y el olvido, a través de lo monstruoso, lo extraño o lo abyecto” (Connelly, 2012) [6], constituye una fuerza transgresora.

La relevancia de esta intersección conceptual radica en cómo lo grotesco utiliza el juego para subvertir conceptos culturales e ideológicos mediante el simbolismo. Esto nos lleva a preguntarnos cómo opera en lo

lúdico y qué influencia ejerce en la experiencia de juego. Por ello, en este artículo diseñamos un prototipo de videojuego que replique estos conceptos simbólicos grotescos para evaluar posteriormente su influencia en el jugador.

Para profundizar en esta cuestión, comenzamos definiendo lo simbólico, tarea compleja por tratarse de un campo de estudio amplio y multifacético cuya interpretación varía según contextos. Lo describimos como una dimensión de la realidad aprehensible por los sentidos, un orden cognitivo y perceptual de conocimiento, que adquiere autonomía respecto a los sujetos que las proponen. Habita entre el mundo interno (imaginario/subjetivo, sentimientos y deseos) y el mundo externo (realidad objetiva, que existen independientemente del conocimiento), funcionando como un eje entre lo real y lo imaginario [1].

Estos procesos mentales internos y percepciones de lo externo están entrelazados con el inconsciente, particularmente con los arquetipos compartidos por la humanidad y sus contextos culturales, construyendo la base de las manifestaciones simbólicas universales [2]. Esto nos revela que la naturaleza simbólica posee una relevancia crucial, al funcionar como ejecutor de las construcciones culturales e ideológicas de la sociedad, manifestándose a través de múltiples expresiones y facetas.

Conforme a estas afirmaciones y desde una perspectiva teórico-estética, lo grotesco muestra interrelación con lo simbólico por su carácter esencial cultural y subversivo. Desafía roles y jerarquías asociadas con las desigualdades sociales, amenaza los límites identitarios y provoca incomodidad debido a su naturaleza transgresora y paradójica. Juega con conceptos contradictorios simultáneos como atracción (erotismo) y repulsión (malformación), humor y horror, ingenio y transgresión [6].

Entre las categorías que conforman lo grotesco, vamos a profundizar en dos categorías que están estrechamente vinculadas:

- El *body horror*, porque enfatiza la vulnerabilidad corporal y la deformación física.
- Lo erótico grotesco (*eroguro*), por subvertir el erotismo mediante lo abyecto, generando simbolismo sexual transgresor y disonancia cognitiva.

Todo esto se aborda desde la perspectiva del cuerpo femenino. Dicha elección se basa en que, según diversos marcos teóricos, lo grotesco está “estrechamente unido al cuerpo, cuyos atributos fundamentales (corporal, fértil, terrenal, cambiante) se alinean con los atributos de lo femenino y, en consecuencia, la idea de lo femenino tal y como se ha construido en la cultura occidental” (Connelly, 2012, p.25-26) [6]. Además, existen múltiples manifestaciones míticas e iconográficas de los monstruos femeninos, que suelen ser simbolizadas con las pasiones viciosas, la unión entre oralidad y sexualidad mediante el engaño y seducción, induciendo la idea del hombre a ser devorado [8].

Así, la figura de la mujer como monstruo sublime es un concepto representativo de lo grotesco por su naturaleza subversiva, potenciada por la combinación de lo carnavalesco y lo performativo (Bajtín, 1984), estableciéndose como una alternativa para experimentar nuevas conformaciones corporales e identitarias [9][18].

Por ello, en nuestro trabajo incluimos un avatar femenino monstruoso que sintetiza ambas categorías, mutaciones (*body horror*) y sensualidad perversa (*eroguro*), mediante mecánicas que desglosamos en la sección 3.

Nuestra aproximación al simbolismo grotesco se vincula a la experiencia lúdica y a su impacto. Por ello, analizaremos brevemente qué conlleva esta perspectiva y cómo se relaciona con el simbolismo.

Durante el juego, el jugador recibe y procesa estímulos, conscientes e inconscientes, activando reacciones y emociones a través de lo simbólico como herramienta psíquica. Desde el punto de vista de Huizinga (2007) en su obra *Homo Ludens*, el juego se sitúa como una de las primeras actividades simbólicas humanas. Opera como instinto primitivo y manifestación vital precultural, observable incluso en animales. El ser humano es esencialmente simbólico, y el juego constituye el proceso por excelencia de generación de lo simbólico [3][10]. Así, el juego moviliza lo simbólico para adquirir significado en un espacio transicional entre el mundo interno y externo, potenciando la capacidad creativa y transformadora de la cultura [4].

En este contexto, los videojuegos digitalizan este proceso evolutivo del juego. Son artefactos surgidos de la convergencia entre lo lúdico y lo tecnológico, que reconfiguran lo simbólico mediante interactividad, permitiendo generar, transformar y transmitir significados culturales, políticos y sociales [11][10]. Como señala Cabañes (2012), en su análisis sobre la transición del juego simbólico al videojuego, estos constituyen una herramienta de producción simbólica y un medio expresivo capaz de representar el funcionamiento y las estructuras tanto de los sistemas reales como de los imaginarios [5][10][12].

Cuando jugamos videojuegos, recibimos estímulos multimodales (imágenes, sonidos, interfaces, etc.), configurando nuevos ámbitos perceptivos que modifican mentalidades y formas de interpretar la realidad, funcionando como medios de producción de conocimiento [10][12]. Estos estímulos se integran en la experiencia lúdica mediante múltiples componentes simbólicos en niveles narrativos, estéticos o mecánicos. Por ejemplo, en el juego experimental *MORF* (Solarski, 2013) se exploró cómo las formas visuales generan distintas emociones. Se implementaron dos tipos de geometrías en la ambientación: unas angulosas y puntiagudas que evocaban miedo o amenaza y otras redondeadas que generaban calma y seguridad, demostrando que la estética moviliza lo simbólico inconscientemente [13].

Incluso en ausencia de narrativa, el movimiento de objetos codifica información simbólica. Un experimento sobre la percepción del comportamiento del movimiento mostró que figuras geométricas en animación provocan en los observadores atribuciones de intenciones y emociones, revelando la tendencia humana a proyectar agencia y narrativa en estímulos abstractos, creando “comportamiento aparente” [14]. Otro caso es *Los Sims* (Electronic Arts, 2000) [25], donde los comportamientos autónomos de los avatares permiten interpretar narrativas emergentes no diseñadas explícitamente, sino surgidas de la agencia del jugador.

Los videojuegos tienen potencial como espacio de experiencias y conocimiento que invitan a interactuar con sistemas mediante elementos tanto narrativos como no-narrativos (la retórica, las reglas, los objetivos, etc.). Estos se vinculan en el imaginario social, alterando actitudes, creencias y juicios sobre el mundo, generando cambios culturales y sociales [5][12].

Pese a estos antecedentes, existen escasos estudios sobre cómo el *body horror* y el *eroguro* se codifican en mecánicas jugables. Por ello, nuestro objetivo es explorar manifestaciones de lo simbólico grotesco en la experiencia lúdica, centrándonos en el cuerpo femenino monstruoso y sus dimensiones clave.

Integrar mecánicas grotescas en la jugabilidad (autodestrucción o deformación), sitúa al jugador en umbrales ambiguos cercanos a lo liminoide surreal o el sin sentido. Este espacio intermedio, sin resolución clara, obliga al jugador a confrontar elementos dispares y buscar significados propios, reconfigurando símbolos desde el inconsciente [7]. Paradójicamente, enfrentar al jugador a situaciones incómodas o controvertidas genera experiencias estimulantes y divertidas [15], vinculadas al papel fundamental de la risa y el juego en lo grotesco [6]. Estas mecánicas funcionan como herramientas críticas para expresar tensiones contemporáneas. En dichos umbrales, la interactividad adopta un carácter performativo, moldeando la percepción del espacio, tiempo y agencia de los sujetos mediante la experiencia de juego, y permitiendo al jugador subvertir significados mediante interpretación, creatividad o incluso transformando/pervirtiendo las reglas del juego [10][12].

Integrar estos elementos grotescos en entornos virtuales interactivos, más allá de ser un mero adorno estético, nos abre complejas formas de experiencia lúdica que confrontan lo real desde la seguridad simbólica. Donde la experiencia es simulada en un ambiente controlado, actuando como protección liminal y potencialmente movilizandop procesos psíquicos inconscientes.

Para evaluar este simbolismo, hemos diseñado un videojuego basado en el modelo MDA [16], donde como diseñadores de juego, intentamos influir en la experiencia del jugador mediante mecánicas. Nuestro diseño intenta transmitir:

- Estados de incomodidad cognitivo-emocional, paradójicamente estimulantes y divertidas.

- Empoderamiento de la agencia creativa/subversiva mediante confrontación simbólica (búsqueda activa de significados interpretativos) y lo performativo (el juego como espejo del jugador).
- Desafío de percepciones y activación de reflexión crítica.

Posteriormente realizaremos un experimento para medir su influencia real en el jugador, pese a la limitación inherente al modelo MDA [16] de influir en la experiencia jugable, aunque hay un grado de libertad, la dinámica de juego depende de cómo el jugador combine mecánicas.

Nuestro objetivo será medir la influencia de mecánicas y elementos visuales simbólicos para conseguir la experiencia de juego deseada.

En este artículo presentamos el diseño del juego, el marco teórico-metodológico, y un esbozo del diseño experimental posterior. Con tres objetivos:

- Diseñar mecánicas, narrativas y estéticas simbólicas grotescas.
- Medir su impacto en incomodidad cognitiva y atracción/repulsión.
- Validar si generan reflexión crítica con un análisis cualitativo post-juego.

El resto de este artículo se organiza de la siguiente forma. En la Sección 2, hablaremos sobre trabajos relacionados al nuestro, tanto desde una perspectiva puramente artística, como de diseño del juego y sus experiencias. En la Sección 3 hablaremos sobre el marco teórico y metodológico que sustenta el diseño del juego y la relación que pretendemos explorar entre lo simbólico, lo grotesco y la experiencia que el jugador va a sentir. En esta sección además explicaremos brevemente el diseño experimental que pretendemos realizar. Por último, en la Sección 4 discutiremos los aspectos que creemos son más relevantes del trabajo, y los posibles trabajos futuros.

2. Trabajos relacionados

Desde una perspectiva tanto artística como de diseño de juegos, existen en la academia e incluso en juegos comerciales independientes (que suelen explorar más temas ligados a lo artístico, estético, etc.) multitud de trabajos que exploran temas psicológicos, simbólicos y grotescos.

El experimento creado por Dang, Li y Juul, *The Suicide Game: Player Perception of Self Destruction in a Game* (2008) [15], explora las ramificaciones para hacer que el jugador trabaje para la destrucción de su personaje y cómo interpretan un juego de temática incómoda, este estudio revela que experiencias lúdicas vinculada a lo grotesco generan gracia y asombro como también activación traumática (una participante declaró que revivió el suicidio de su amiga) evidenciando cómo los juegos movilizan emociones y experiencias profundas[15].

Este estudio es relevante para nuestro prototipo al ilustrar reacciones emocionales ante mecánicas disruptivas. Asimismo, exploramos el cuestionamiento identitario mediante el monstruo femenino, línea abordada por Spittle (2011), donde desde el análisis del juego *F.E.A.R* como un estudio de caso, explora cómo los juegos con temáticas relacionadas con la incertidumbre, manipulan estados psicológicos e identitarios mediante recursos abyectos y siniestros. Este videojuego, a diferencia de otros FPS, sitúa al jugador en un estado liminal mediante interrupciones del orden simbólico, generando una retórica de la vulnerabilidad, donde el jugador se ve forzado a confrontar la fragilidad identitaria a través de mecánicas que materializan lo reprimido (traumas, muerte, lo maternal monstruoso) [17].

En la misma línea, en el juego de *Hellblade: Senua's Sacrifice* [24], sobre terror psicológico y aventura, tenemos un ejemplo de cómo la distorsión y las voces emergentes crean una disonancia al jugador, creando una dislocación identitaria y fomentando la empatía crítica y la confrontación con prejuicios, al posicionarlo encarnando un personaje con psicosis.

En otros videojuegos comerciales e indies, también se ha explorado temas relacionados a lo simbólico en lo grotesco, como por ejemplo *Scorn* [19], con una inmersión estética y artística, tomando como referencia una fusión del arte surrealista, onírico y decadente de Zdzisław Beksiński y la relación entre cuerpo y máquina de

H. R. Giger, haciéndonos interaccionar con esta fusión, transgrediendo la frontera de lo estético-artístico. Interaccionando con la deformación y la mutación del cuerpo exponiéndose a situaciones moralmente o éticamente y simbólicamente disruptivas.

Podemos decir que si en un cuadro surrealista de Z. Beksiński, observamos el cuerpo monstruoso, en Scorn lo haces con tus manos deformando cuerpos, violando límites, generando una posición distinta a la deformación, actualizando el concepto de grotesco, llevándolo a un plano interactivo.

Este caso nos sirve como inspiración en nuestro prototipo desde una perspectiva estética-artística y como se fusiona lo artístico grotesco en jugabilidad.

Siguiendo por esta temática del género de aventura surrealista, en el videojuego de *Yume Nikki* [20], donde partimos de un escenario inquietante y extraño, la exploración onírica se convierte en el principal motor de experiencia, donde el simbolismo disperso en los entornos y encuentros con personajes u objetos, no configura una narrativa explícita, sino que permite, a través de la agencia del jugador, la construcción fragmentaria y subjetiva de una historia. Es precisamente esta apertura interpretativa la que otorga sentido a lo simbólico, haciendo que cada hallazgo adquiera sentido emocional dentro de una lógica narrativa no lineal, marcada por lo abyecto, lo liminal y lo inconsciente.

Analizando casos donde se involucran el símbolo como parte de la experiencia narrativa manipulando signos ocultistas y ritualísticos para resolver situaciones siniestras, podemos citar el videojuego *World of Horror* [21], con una estética de 1 Bits vinculada con el arte Junji Ito y el universo lovecraftiano, que nos induce a situaciones laberínticas e inquietantes, acentuado por las mecánicas roguelike.

También en *Inscription* [22], donde los componentes simbólicos del videojuego inducen al jugador a participar en rituales de sacrificio, como núcleo de su experiencia lúdica. Estas prácticas están integradas tanto en la narrativa como en las mecánicas, estructuradas en forma de estrategia a través de las cartas, de puzzles y dinámicas de escape room.

La lógica ritual en estos juegos, no solo opera como recurso para la resolución de enigmas, sino que reviste de sentido grotesco cada acción del jugador, transformando el avance en una sucesión de actos cargados de simbolismo siniestro y perturbador.

Un caso representativo que subvierte las convenciones del propio juego y utiliza la narrativa como una experiencia desestabilizadora, acentuando el carácter grotesco y perturbador, sería el simulador de citas *Doki Doki Literature Club* [23]. Este juego pervierte la lógica de lo romántico hasta convertirlo en un dispositivo de sacrificio simbólico, a través de la ruptura de la cuarta pared y la manipulación directa de archivos y códigos, bajo una apariencia estética inocente y kawaii. De esta forma, convierte al jugador en agente de un ritual metaficcional, generando una disonancia inquietante entre lo afectivo y lo sádico, involucrando al usuario en un proceso de transgresión emocional y psicológica.

Consideramos que estos videojuegos son interesantes mencionar porque están muy vinculados a la perspectiva simbólica grotesca en sus estéticas, narrativas y mecánicas. A su vez, nos sirven como una guía de exploración en sus diseños y qué recursos son implementados para generar experiencias interactivas dentro de nuestro marco de investigación.

3. Metodología

El desarrollo de este prototipo de videojuego se articula en torno a dos pilares teórico-estéticos vinculados a lo grotesco y lo simbólico, materializados en mecánicas de juego.

- **Body Horror y Eroguro:** Se implementa mediante un avatar femenino que funciona como espejo liminal del jugador. Este actúa como puente de diálogo entre la agencia del jugador y el sistema lúdico, traduciendo las mutaciones y deformación física en interacción directa.
- **Surrealismo Metaficcional Esotérico:** La trama explora la posesión del jugador por el avatar femenino atrapado en un sueño, que busca escapar colonizando la voluntad de este, estableciendo así un

diálogo metaficcional. Este entramado narrativo se sitúa en un escenario con un tablero y una baraja interactiva similar al tarot.

A continuación, desglosamos la articulación teórica, mecánica y estética de cada eje, explicando la construcción de una experiencia grotesca surrealista inmersiva.

3.1 Body Horror y Eroguro

El avatar funciona como un avatar-espejo reactivo, conformando el pilar de las mecánicas del prototipo. Es una marioneta simbólica que interpreta un rol similar a una médium o tarotista atrapada en un mundo onírico, que encarna la dialéctica *body horror/eroguro*, a medio camino entre lo adorable y lo monstruoso, la cual sufre deformaciones y mutaciones corporales en tiempo real. Estas distorsiones corporales pueden reflejar daño, corrupción, adquisición de habilidades o consecuencias morales de las decisiones del jugador. Para ello se utilizará la cámara y el micrófono del dispositivo para detectar el movimiento, la iluminación y los sonidos del entorno del jugador. Esta marioneta simbólica grotesca tiene diferentes funciones:

- **Diálogos interactivos:** El avatar tiene una comunicación directa con el jugador a través del intercambio de texto. Esta característica sirve tanto como guía para el jugador, como un elemento de feedback de lo que está ocurriendo en el juego.
- **Estética performática:** Mediante animación 2D (cabeza/cuerpo separadas) con técnicas de ilustración (sprites) o video-performance (imagen real), reacciona a los estímulos del juego, decisiones del jugador o a sus acciones registradas por la cámara o el micrófono.
Estos cambios se verán reflejados en el avatar posicionándose monstruosamente con su cuerpo femenino conflictivo y grotesco, basados en los conceptos del cuerpo monstruoso (*body horror*) y lo erótico grotesco (*eroguro*).
A través de gestos faciales, la deformación o mutación del cuerpo, como ojos múltiples, desdoblamiento, cicatrices, hibridaciones de animal o máquina, etc. y a su vez estéticas que erotizan lo monstruoso, como sonrisa seductora con colmillos, desnudez mostrando deformaciones y protuberancias, etc.
Otro tipo de perturbación que sufre son efectos tipo glitch, cambios entre ilustración e imagen real, etc. distorsiones propias de lo digital, para evocar un efecto metaficcional.
Todo el conjunto actúa como un medio de desestabilización del orden social, encarnando diversos disfraces que le permiten mutar y alterar su entorno, siendo una representación de mujer como monstruo sublime, un concepto representativo de lo grotesco por su naturaleza subversiva [18].
- **Sistema de Posesión:** Las decisiones del jugador y sus acciones en el juego afectan a un medidor, que simboliza la lucha y negociación entre el avatar atrapado en un mundo onírico y el jugador en el mundo real, como fuerzas opuestas.
El avatar busca influir y convencer al jugador mediante seducciones y manipulaciones emocionales, reflejado en sus mutaciones físicas, cuanto más ventaja tenga más adorable y viceversa. Además, corrompiendo el propio juego, de forma que al final, el jugador se verá influenciado de alguna manera por el avatar.
Esta lucha de posesión de identidad será representada usando una barra que refleja el conflicto psíquico-simbólico del jugador en base a una serie de decisiones:
 - Resistencia a la posesión: como control propio y miedo al descontrol.
 - Entrega y ayuda: dejando ser poseído por el avatar, fusión y deseo de conexión.
 - Equilibrio: Disolución, un equilibrio perfecto o empate.

Lo que se busca es transmitir al jugador, durante el desarrollo del juego, es la sensación simultánea de atracción morbosa, por la seducción y singularidad del avatar, y rechazo ético, por su apariencia o acciones sádicas y satíricas. Creando una tensión emocional, que pretende generar una fascinación e incomodidad en el jugador.

Estas mecánicas se articulan para influir al jugador a cuestionar su conexión con el personaje y las normas del juego, introduciendo elementos perturbadores que rompen la inmersión convencional para generar reflexión o incomodidad. Al obligar a encarnar al 'otro', marginado o rechazado, fomentan la empatía crítica y la confrontación con prejuicios, convirtiéndose en herramientas poderosas para satirizar, con humor perverso o absurdo, defectos y desigualdades sociales.

3.2 Surrealismo metaficcional esotérico

El prototipo opera bajo un paradigma de lo grotesco, como ya hemos mencionado, dónde además hemos integrado juegos tradicionales como La Oca y el Tarot, juegos tradicionalmente simbólicos, que hemos subvertido mediante distorsiones sistémicas, inspirando un escenario surreal.

Esto se implementa bajo dos componentes, que corresponden a un espacio, que hemos denominado el tablero onírico (La Oca) y un soporte vehicular como mediador entre jugador y avatar, una baraja denominada el oráculo corrupto (Tarot).

3.2.1 El Tablero Onírico

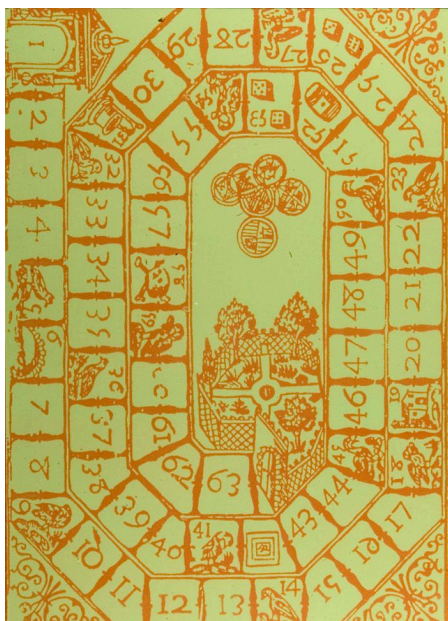
Esté artefacto lúdico se basa en un tablero, como escenario, que representa un portal metaficcional que media entre el mundo real del jugador y el plano onírico del avatar. Este instrumento opera como mapa onírico surrealista y desencadena la jugabilidad de la baraja interactiva.

El jugador se convierte en ficha, cuyos movimientos desencadenan subversiones regladas, alteradas por el avatar, donde la narrativa emerge de la negociación entre fuerzas opuestas, donde ambos, jugador-avatar pueden alterar el estado de este, mediante cancelación de casillas o reordenamiento. Las casillas del tablero alteran las normas ambientales, narrativas y mecánicas del juego, desdibujando los límites de las normas donde se desarrollan las partidas.

Este tablero onírico integra inspiración del juego tradicional clásico de La Oca, como hemos mencionado, sintetizando funciones estético-mecánicas, usando la jugabilidad propia de este juego, con estructura de progresión no lineal, implementada en 64 nodos (casillas), (inspirado en el tablero tradicional, con 63 casillas más un casilla de inicio), con eventos temáticos que implican tanto ambientación y corrupción visual, como distorsiones UI y cambios estéticos, como retos o minijuegos que exigen resolver paradojas absurdas o éticas.

Además de los ya mencionados, integrará puntualmente eventos mayores o especiales propios de La Oca adaptados al prototipo, esto nos permite mantener la atención del jugador y sorprenderlo. Por ejemplo, en la casilla de la posada, el jugador pierde turno manteniendo la mecánica de La Oca con la integración de ventaja sobre el jugador del avatar, pudiendo ejercer doble corrupción. En la casilla del laberinto donde se desencadena un pequeño puzzle que atrape al jugador durante un tiempo para poder escapar, etc.

A continuación, para mayor comprensión aportamos una impresión histórica del tablero La Oca que hemos usado de referencia por su antigüedad, corresponde al s.XVII impreso por el establecimiento de Guasch en Mallorca [26]. También una tabla, con distribución numérica lineal para mayor comprensión. Las casillas de inicio y final se indican en color rojo, las casillas de eventos en morado y las casillas avance especial en naranja:



Juego de la Oca de Guasch (s.XVII). [26]

0 START	1	2	3	4	5 OCA	6 PUENTE	7
8	9 OCA	10	11	12 PUENTE	13	14 OCA	15
16	17	18 OCA	19 POSADA	20	21	22	23 OCA
24	25	26 DADOS	27 OCA	28	29	30	31 POZO
32 OCA	33	34	35	36 OCA	37	38	39
40	41 OCA	42 LABERINTO	43	44	45 OCA	46	47
48	49	50 OCA	51	52 TORRE	53 DADOS	54 OCA	55
56	57	58 MUERTE	59 OCA	60	61	62	63 END

Este sistema encaja en nuestro prototipo, fomentando la tensión y el discurso entre el avatar y jugador, otorgando un carácter interpretativo simbólico del inconsciente, propio de juegos simbólicos sobre el camino de la iluminación o adivinatorios, donde coexisten la figura del médium o descodificador simbólico de la lectura, como un portador de respuestas enfrentándose a un jugador o usuario, que busca estas respuestas, creando una dinámica de tensión entre lo conocido y lo desconocido. Fundamentado en la teoría junguiana del inconsciente colectivo, vinculado como artefacto para el diálogo con el inconsciente (Jung, 1959).

3.2.2 El Oráculo corrupto

Otro componente clave del prototipo es la integración de mecánicas basadas en juegos de baraja de adivinación simbólica como el Tarot, los oráculos y el I Ching, otra influencia relevante es el juego Dixit (2008), estos juegos tienen algo en común, el uso de la interpretación simbólica de las cartas. Hemos implementado este concepto de juego de interpretación como también la ritualidad de echar tiradas para la lectura, como una herramienta mediadora entre avatar-jugador.

Además de otras funciones añadidas, como generador estocástico de movimiento (sustituto de dados) y otorgador de habilidades y recursos estratégicos para el jugador, que subvierten y alteran el juego.

Esta baraja opera mecánicamente como:

- **Mecánica de comunicación:** Echado de cartas como ritual, el jugador selecciona acciones y el avatar responde.
- **Mecánicas de Interpretación:** Las cartas son objetos simbólicos jugables, extraídos de forma aleatoria, a través de la interpretación entre la interacción jugador-avatar.

3.2.2.1 Arquitectura del Sistema Adivinatorio

Estructura de la baraja:

El diseño propuesto consta de una baraja de 74 cartas. Cada carta contiene un valor numérico, que opera como avance por el tablero y una imagen simbólica que determina el comportamiento de esta sobre la partida (un efecto), unas características que sirven como lectura simbólica y habilidades que se adquieren (beneficios y desventajas). Constaría de dos tipos:

- 36 cartas numéricas del 1 al 9 y de 16 figuras: Sota 10, Caballo 11, Rey 12 y Reina 13. (4 palos x 13 cartas, total de 52 cartas).
- 22 cartas de arcanos mayores

Para mantener un buen flujo entre baraja y tablero, hemos compuesto una tabla con la distribución y efectos especiales otorgando los valores aproximados de avance adecuados para el tablero:

Tipo	Cantidad Cartas	Valor Original	Avance Juego	Efecto Especial
Numéricas 1-7	28	1-7	1-7	-
Numéricas 8-9	8	8-9	8-9	-
Sota	4	10	5	Movimiento especial
Caballo	4	11	5	Movimiento especial
Rey	4	12	5 + X (variable)	Roba carta y suma su avance sin efectos
Reina	4	13	6	Movimiento especial + Protección
Arcanos I-VII	7	1-7	1-7	Efecto según carta
Arcanos VIII-XIV	7	8-14	1-7	Efecto según carta
Arcanos XV-XXI	7	15-21	1-7	Efecto según carta
El Loco	1	0	5	Efecto Aleatorio

Estadísticas clave:

- Media de avance: 4.76
- Tiradas para 64 casillas: 74 cartas (22 arcanos + 52 numéricas) vs. 64 casillas = 13.43 tiradas ($64 / 4.76 \approx 13.43$ tiradas).
- Máximo avance: 14. Carta Rey (5) + carta 9 (9) = 14
- Probabilidad de avance: ≤ 7

Mecánicas de ejecución:

Cada casilla del tablero onírico activa un evento donde emergen 3 de las cartas de la baraja, representando 3 estados que otorgan efectos diferentes, simulando una tirada de Tarot, que representan simbólicamente el pasado, el presente y el futuro:

- **1º Carta de Pasado:** Modificador de recursos y habilidades, estas cartas son acumulables, otorgan la estrategia y agencia al jugador.
- **2º Carta de Presente:** Ejecuta un efecto vinculado con el simbolismo de la carta, corrompiendo o distorsionando el juego o bien sobre el tablero.
- **3º Carta de Futuro:** El valor numérico de las cartas dicta el movimiento por el tablero. (Sustituto de los dados)

Para clarificar cada estado proponemos una tabla a modo de ejemplo:

Fase	Función sistémica	Arcano El Loco en función de posición	Efecto en el jugador
Pasado	Modificador Habilidades/Recursos	Desventaja aleatoria	Recurso acumulable como arma hacia el avatar
Presente	Reglas/ Efecto UI	Cambia la UI de orden	Disonancia cognitiva
Futuro	Determinación espacial (Avance por el tablero)	Avanza +5 Casillas Cambiar una casilla al azar	Restricción de agencia

Además añadimos un ejemplo de carta del tarot, el Loco, en su forma original del Tarot Marsella (s.XV) y su lado la versión modificada para el prototipo:



Lectura completa de la tríada: Como hemos indicado, cada carta tiene una característica simbólica, en la tirada de cartas este conjunto de tres, suman sus características para crear un texto simbólico, que es narrado por el avatar.

El avatar actúa como figura oracular ofreciendo una lectura simbólica de las cartas, incluyendo una pista velada, esta lectura es poética, ambigua y amenazante. El jugador debe adivinar y elegir cual es la carta elegida por el avatar, usando la intuición o deducción (y ciertas pistas que proporcionará el entorno), dependiendo de si acierta o no, modifica el sistema de posesión.

Este Tarot, corrupto y simbólico opera como:

- **Dispositivo mecánico-arquetipal:** Cartas como representaciones de complejos afectivos con imágenes emocionales autónomas que interfieren con la conciencia.
- **Sistema de agencia distribuida:** La tríada Pasado-Presente-Futuro materializa la dialéctica entre determinismo (valores numéricos) y caos interpretativo (efectos paradójicos).

3.3 Resumen

Este marco teórico y lúdico convierte el tablero y la baraja en artefactos liminales, donde lo grotesco no es solo estético, sino la esencia mecánica del sueño y la interacción jugador-avatar, operando como lenguaje, a través de 3 vías clave:

1. **La metaficción operativa:** Transforma al jugador en co-autor del espacio liminal, sobre agencia en entornos digitales. El avatar actúa como agente autónomo que altera reglas, reflejando la tensión entre:
 - Jugador real (conciencia racional).
 - Sistema onírico (inconsciente procedural).
2. **Corrupción de affordances:** Mediante Glitches de UI o corrupción en reglas de juego, que invalidan expectativas de interacción y actúan como mecanismos de disonancia cognitiva, como en los minijuegos o retos que combinan paradojas lógicas y distorsiones.
3. **Sistemas de azar estructurado:** Dispositivo que permiten narrativas emergentes con coherencia arquetípica, con eventos basados en tarot que funcionan como mecánicas adivinatorias, simulando rituales de significado, donde el azar se percibe como destino decodificable. La aleatoriedad trasciende su función estocástica para construir significados oníricos, reflejando el concepto freudiano del sueño como "espacio de condensación simbólica" (Freud, 1900).

4. Diseño experimental

El diseño experimental que describimos a continuación pretende validar si todas las mecánicas descritas en la Sección 3 nos permiten influir en el jugador para conseguir una experiencia jugable que sea perturbadora, sorprendente y que produzca una sensación simultánea entre el rechazo y la atracción, estimulando las ganas de explorar más en el universo grotesco que el juego plantea.

Para ello, el experimento se debe plantear como una sesión de juego con diferentes sujetos de estudio, lo más heterogéneos posible, que deberán jugar al juego en unas condiciones concretas. Lo primero es que el juego debe jugarse en un pc que disponga de cámara y micrófono, ya que como hemos comentado anteriormente, necesitamos captar los movimientos y el sonido que el jugador realice, para que el jugador los utilice como input.

Por otro lado, para que el jugador se sienta cómodo jugando, debe estar sólo en una habitación. La iluminación debe ser tenue y dicha habitación debe estar vacía, es decir, el fondo del jugador esté exenta de objetos, como fondo plano.

A los sujetos de estudio se les informará y se pedirá su consentimiento para ser monitorizados. La idea es tener una doble comprobación, por un lado se analizará las reacciones del jugador durante el juego y por otro se le pasará un cuestionario con cuestiones relacionadas con el juego y su sensación en los momentos clave del mismo.

Esta información se contrastará de forma que podamos deducir si lo que se refleja en el cuestionario concuerda con las mediciones realizadas. Acompañado con un test previo donde se le preguntará acerca de cuestiones generales como: Tipo de jugador (habitual, ocasional, etc), edad, género, relación con la temática que ofrece el juego, nivel relación o familiaridad con videojuegos, sensibilidad ante temas incómodos o tabú, sensibilidad a la frustración, sensibilidad ante lecturas simbólicas e interpretativas, nivel de relación o familiaridad con juegos interpretativos o adivinatorios.

Toda esta información será procesada para intentar determinar si el diseño de juego y su ambientación artística han conseguido transmitir la experiencia jugable deseada utilizando para ello herramientas estadísticas que determinen si los resultados son significativos como por ejemplo el P valor o prueba de t student.

5. Conclusión y trabajos futuros

La investigación busca analizar cómo los símbolos de lo grotesco (*eroguro*, *Body horror*, lo surreal o el sin sentido, carnavalesco, etc.) trascienden la representación visual para codificarse en sistemas lúdicos, explorando su capacidad para generar espacios liminoides, desafiando percepciones, donde mecánicas transgresoras, distorsionan fronteras de la identidad, jugador-personaje, y activan una reflexión y diálogo inconsciente a través de la interactividad.

Este marco permite enriquecer el diseño de experiencias interactivas, donde el control de avatares grotescos y símbolos disruptivos, confronta al jugador con dilemas éticos y reflexiones sobre los límites de lo humano y la propia complicidad del jugador. Y desafiando, de forma crítica, la lógica instrumental de los juegos comerciales.

Así mismo, presentamos un esbozo de cuál sería el diseño experimental que realizaríamos para validar si el diseño del juego y su ambientación estética, permiten que un número significativo de jugadores, experimenten la experiencia jugable deseada, apoyándonos en una doble verificación, mediante el análisis de las reacciones del jugador jugando al juego y un test sobre su experiencia jugando.

Esta investigación representa, por tanto, una aportación significativa a los Game Studies al ofrecer un marco teórico-práctico basado en cómo los sistemas lúdicos vehiculan y potencian significados propios de lo grotesco. Este enfoque establece un puente teórico entre la semiótica visual, la teoría del juego y la psicología de la interactividad, profundizando en la capacidad del medio para operar a nivel simbólico y reflexivo.

Simultáneamente, enriquece el ámbito artístico al validar el diseño de juegos como un soporte para la experimentación estética. Este enfoque genera espacios interactivos que fomentan la reflexión crítica y confrontan dilemas éticos, cuestionan identidades y desestabilizan percepciones establecidas. No sólo consolidando el potencial del videojuego como artefacto cultural crítico, sino que también genera un diálogo enriquecedor con otras prácticas artísticas disruptivas contemporáneas.

Así, esta investigación no solo contribuye a los campos mencionados, sino que también abre nuevas vías de exploración en la intersección arte-juego.

Referencias

- [1]Lacan, J. (1981). *Los escritos técnicos de Freud, Seminario I*. Paidós, Buenos Aires. (Trabajo original publicado 1953–1954).
- [2]Jung, C., Von Franz, M. (2017). *El hombre y sus símbolos*. Paidós, Barcelona. (Trabajo original publicado 1964)
- [3]Huizinga, J. (2012). *Homo Ludens: El juego y la cultura*. Madrid Alianza Editorial (3ªed). (Obra original publicada en 1938).
- [4]Winnicott, D. W. (1971). *Playing and Reality*. Tavistock Publications.
- [5]Bogost, I. (2007). *Persuasive Games: The Expressive Power of Videogames*. MIT Press.
- [6]Connelly, F. S. (2012). *Lo grotesco en el arte y la cultura occidentales*. Cambridge University Press.
- [7]Turner, V. (1982). *From Ritual to Theatre: The Human Seriousness of Play*. Performing Arts Journal Publications.
- [8]Cortés, J. M. G. (1997). *Orden y caos, un estudio cultural sobre lo monstruoso en el arte*. Anagrama, Barcelona.
- [9]Creed, B. (2007). *The Monstrous-Feminine: Film, Feminism, Psychoanalysis*. Routledge, Taylor & Francis Group, Londres.
- Aarseth, E. J. (1997). *Cybertext: Perspectives on ergodic literature*. Johns Hopkins University Press.
- Bartlett, F. C. (1995). *A Study in Experimental and Social Psychology*. Cambridge University Press. (Trabajo original publicado 1932).
- Driscoll M. (2010). *Absolute erotic, absolute grotesque: The living, dead, and undead in Japan's imperialism, 1895–1945*. Duke University Press.
- Kristeva, J. (1988). *Poderes de la perversión*. Siglo XXI, México. (Trabajo original publicado 1980)
- Cohen, J.J. (1996). *The Monster theory*. University of Minnesota Press.

Referencias Online

- [10]Cabañes, E. (2012). *Del juego simbólico al videojuego: la evolución de los espacios de producción simbólica*. ResearchGate. URL: https://www.researchgate.net/publication/328347081_Del_juego_simbolico_al_videojuego_la_evolucion_de_los_espacios_de_produccion_simbolica
- [11]Frasca, G. (2009). “Juego, videojuego y creación de sentido. Una introducción”. *Revista Internacional de Comunicación y Desarrollo*, (9). URL: <https://revistascientificas.us.es/index.php/Comunicacion/article/view/21513/18975>
- [12]Cabra, A. (2010). *Videojuegos: máquinas del tiempo y mutaciones de la subjetividad*. *Revista Latinoamericana de Ciencias de la Comunicación*, 7(13). URL: <https://www.redalyc.org/pdf/860/86020052010.pdf>

- [13]García de Enterría Díaz, D. (s.f.). *La experiencia estética del videojuego. Simbolismo, naturalismo y abstracción*. Academia.edu. URL: https://www.academia.edu/20253036/La_experiencia_estetica_del_videojuego_Simbolismo_naturalismo_y_abstraccio_n
- [14]Heider, F. y Simmel, M. (1944). *An experimental study of apparent behavior*. The American Journal of Psychology, 57(2), 243–259. URL: <https://doi.org/10.2307/1416950>
- [15]Juul, J. (2008). *The Suicide Game: Player Perception of Self Destruction in a Game*. URL: <https://www.jesperjuul.net/ludologist/2008/01/30/the-suicide-game-player-perception-of-self-destruction-i-n-a-game/>
- [16]Hunicke, R., LeBlanc, M., Zubek, Z. (2004). *"MDA: A Formal Approach to Game Design and Game Research*. Northwestern University. URL: <https://users.cs.northwestern.edu/~hunicke/MDA.pdf>
- [17]Spittle, S. (2011). *"Did This Game Scare You? Because it Sure as Hell Scared Me!" F.E.A.R., the Abject and the Uncanny*. Games and Culture, 6(4), 312-326. URL: <https://doi-org.bucm.idm.oclc.org/10.1177/1555412010391091>
- [18]Poblete Pardo, N. (2007). *Pavor misógino: La mujer como monstruo excelso en "El infierno tan temido", de Juan Carlos Onetti*. Aisthesis, (42),70-83. URL: <https://www.redalyc.org/pdf/1632/163219818006.pdf>
- L. P. H. Garcia, M. I. (2022). *A utilidade dos monstros no concept art dos videojogos e a sua relação com o ensino artístico*. Tesis Doctorado en Bellas Artes, Universidad de Lisboa, Facultad de Bellas Artes. URL: <http://hdl.handle.net/10451/51534>

Referencias Videojuegos

- [19]Ebb Software (2022). *Scorn*.
- [20]Kikiyama (2004). *Yume Nikki*.
- [21]Panstasz (2023). *World of Horror*.
- [22]Daniel Mullins Games (2021). *Inscription*.
- [23]Team Salvato (2017). *Doki Doki Literature Club*.
- [24]Ninja Theory (2017). *Hellblade: Senua's Sacrifice*.
- [25]Electronic Arts (2000). *Los Sims*
- Weather Factory (2018). *Cultist Simulator*.

Referencia Juegos Clásicos

- [26] (s. XVII). *El Juego de La Oca*. Guasch, Mallorca. URL: https://funjdiaz.net/museopapel/expo0506oca_ficha.php?id=506140
- Barros, A. (1588). *Juego de La Oca: Filosofía cortesana de Alonso de Barros*. Marius Cartarius.
- Suárez F.,E (2016). *Alonso de Barros. Filosofía cortesana moralizada*. Barcelona. URL: https://parnaseo.uv.es/lemir/Revista/Revista23/Textos/03_Filosofia_cortesana_Barros.pdf
- Wilhelm, R. (2001). *I Ching, El libro de las mutaciones*. EDHASA, Hispano Americana, S.A., Barcelona.

Motion capture for gesture detection in virtual environments

Pablo Sánchez-Martín^{1,*†}, Alejandro Barrachina-Argudo^{1,†}, Alejandro Romero-Hernández^{3,†} and Ismael Sagredo-Olivenza^{4,†}

¹Software Engineering and Artificial Intelligence Department, Facultad de Informática, Complutense University of Madrid, Madrid, Spain

Abstract

This study aims to improve non-verbal communication in virtual environments by enabling NPCs (non-player characters) to detect and respond to user gestures captured through a motion capture suit. NPCs in virtual worlds typically lack the capacity to interpret and react meaningfully to non-verbal cues, limiting the realism and interactivity of these environments. To address this, we explored the use of various machine learning models for gesture recognition. Due to the lack of publicly available datasets suitable for our purposes, we compiled our own dataset using the motion capture suit. Among the tested models, a Random Forest classifier yielded the best results, achieving an accuracy of 0.86, which we consider a reliable benchmark for validating the system. As a further contribution, we have published the collected dataset on Kaggle to support future research in this area.

Keywords

Motion capture, Gesture detection, Machine Learning, Non-verbal communication, Perception Neuron 3, Virtual environments, Virtual Reality

1. Introduction

Virtual environments are increasingly present in our society due to the democratization of devices that enable their development (such as the Meta Quest) and improvements in related technologies, such as rendering, user position tracking sensors, and more.

There are numerous applications for these virtual worlds, where users' non-verbal communication within the system is becoming increasingly important, for example in virtual social simulation environments or virtual group meetings, among others.

However, up to now, the use of non-verbal communication in virtual environments has been largely limited and predefined by the system design, restricting a fundamental part of the way humans naturally express themselves.

For this reason, and as seen in works such as Neverova et al. (2013) [1] and Herbert et al. (2024) [2], gesture recognition is an active area of research aimed at addressing this limitation. Thus, exploring new ways to expand non-verbal communication representation is relevant for making interactions between users and NPCs more natural. This work is framed within the research line of developing tools capable of recognizing gestures performed by users through motion capture suits, with the goal of creating more immersive and natural virtual environments.

The need to use a full-body motion capture suit, instead of the typical controllers provided by current mixed reality headsets, lies in their ability to offer highly accurate tracking of body and environmental movement. Although there are alternative methods, such as Microsoft Kinect or detection using light sensors, the former have more limited depth detection, while the latter—despite being much more

Woodstock'22: Symposium on the irreproducible science, June 07–11, 2022, Woodstock, NY

*Corresponding author.

†These authors contributed equally.

✉ pablos19@ucm.es (P. Sánchez-Martín); alejba02@gmail.com (A. Barrachina-Argudo); alerom02@ucm.es (A. Romero-Hernández); isagredo@ucm.es (I. Sagredo-Olivenza)

🌐 <http://conceptbase.sourceforge.net/mjf/> (A. Romero-Hernández); <https://gaia.fdi.ucm.es/> (I. Sagredo-Olivenza)

🆔 0000-0002-9421-8566 (A. Romero-Hernández); 0000-0002-9421-8566 (I. Sagredo-Olivenza)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

precise—require specialized rooms and significantly more expensive equipment, making them practically unfeasible today for everyday use in a virtual reality application aimed at the general public.

On the other hand, building AI models capable of supporting such tools requires large amounts of data. Specifically, in our work, the required data consists of a sequence of 3D coordinates, representing the positions and rotations of skeleton joints across animations. Consequently, the data to be processed is relatively large in size.

This makes it necessary to design complex models capable of processing a significant amount of input data. However, as we will discuss throughout this work, no suitable datasets of this type have been found. As a result, one of the key contributions of this project has been the creation of a dataset, achieved through the extraction of animations from a specialized animation library and the capture of user motion data.

In light of these motivations, we can conclude that the lack of natural methods for exploiting non-verbal communication in virtual environments, combined with the absence of suitable animation datasets, has driven the objectives of this work, which we detail in the following sections.

The rest of the paper is structured as follows. Section 2 shows the related work with our approach. The Section 3 describes the development process leading to the gesture recognition system. In Section 4 describes the experiments realized to get the AI models, in Section 5 analyzes the results of them and finally in Section 6 discuss the conclusions, limitation and future works.

2. Related Work

Thanks to the advances in computer vision, works on real-time gesture detection have seen significant progress, especially in hand gesture processing. works such as Kopuklu et al. (2019)[3] and Dardas and Georganas (2011)[4] use different machine learning models (convolutional neural network and support vector machine, respectively) to detect hand gestures in real-time.

Although hand detection is the most common, there are other works that try to process a full body detection such as Neverova et al. (2013)[1], where they use a recurrent neural network to detect upper body actions; or Weinland et al. (2010)[5], where they detect some full-body actions.

In the field of medicine, the study of gesture detection is also extensive. An example of this is Ma et al. (2023)[6], where they compare artificial intelligence models based on a 3D skeleton using a Kinect (Microsoft, 2010).

This advancements in gesture detection make it possible to improve non-verbal communication in virtual environments, where various works have been conducted in recent years. For example, in Xenakis et al. (2022)[7], the authors focus on the concept of presence in virtual environments, emphasizing the roles of immersion, emotional content, and avatar fidelity in enhancing user experience. In this study, the authors highlighted how the realism of avatars and their behaviors (movement and body language) help achieve a greater sense of social presence among users. It also discusses how these elements influence interpersonal dynamics and user perception in virtual environments. These roles discussed in the study can be extrapolated to the experience of playing video games.

If we focus on the context of traditional video games, non-verbal communication options are limited to those established by designers, such as gestures or “emotes”, or actions performed by players like crouching, turning their avatar, or moving their head when moving the camera, etc. This limitation leads virtual environments to fall into what is known as suspension of disbelief[8]. Therefore, it is important to improve non-verbal communication in virtual environments to achieve more natural communication and prevent users from falling into this suspension of disbelief, making these environments more immersive and believable.

For this reason, in this work, we utilize full-body gesture detection through machine learning, To enable NPCs to be able to recognize the gestures that the player makes within the virtual environment. In this way, it is possible to create more realistic and reactive behaviors to the gestures of the player, allowing to enhance non-verbal communication.

3. Workflow

The aim of this work is to detect different types of gestures commonly used in video games, such as dancing, greeting, pointing, sitting, fighting, and running. With this goal in mind, a comparison of different machine learning models was made to determine which one would be the most suitable for this purpose, considering their accuracy and performance, as the aim is to run these models on limited hardware, such as standalone VR headsets. This is because works such as Pallavicini et al. (2019)[9] shows that players find environments more immersive in Virtual Reality.

To train these models, we require 3D position and rotation data of a virtual skeleton representing the user. The dataset is converted into CSV format, where the columns correspond to the different bones of the model and the rows represent the frames of the animations.

After an exhaustive search on specialized dataset pages like Kaggle, as well as other dataset publication environments such as the Carnegie Mellon University dataset [10], it was concluded that the available gestures were either insufficient or not in the appropriate format. Therefore, we decided to use a free animation bank (Mixamo¹) to obtain the animations to train the models.

In the process to obtain the animations, two tools have been developed to facilitate the conversion of Mixamo animations into the desired format. Firstly, a tool called “Mixamo-Downloader”² was found and fixed to work as intended³. This tool allowed to mass download the animations that match the gestures to detect, creating a Kaggle dataset of FBX animations. Secondly, a tool called “Mixamo-Animation-Dumper”⁴ was implemented to convert the downloaded animations into CSV format. This tool downloads the animations from the Kaggle dataset⁵, executes them in the Unity Editor, writes the respective CSV for each animation, and finally uploads the CSV files to other Kaggle dataset.

With the aim of properly preparing them for the training process, the following modifications were made during standardization:

- All the animations must have the same number of frames. To achieve this, the duration is standardized to 90 frames, which is the number of FPS used by Meta Quest 2 VR Headset by default and the Mixamo Animation Dumper tool.
- Animations that last more than 90 frames are divided into multiple animations of 90 frames or less.
- Animations that last less than 90 frames are repeated until reaching 90 frames.
- Animations that do not reach 10 frames are considered invalid and removed.

Once all the animations from Mixamo were downloaded, a significant difference in data was observed between some types of animations, where the greeting and pointing classes were significantly underrepresented compared to the others. This was further exacerbated by data normalization process, because we segmented the animations into a fixed frame length of 90 frames long, and the dance animations were significantly longer. The result was a highly imbalanced dataset as shown in Figure 1.

To solve this issue, we could have chosen to discard part of the dance animations, but since the total number of the remaining animations was insufficient, we decided to balance the dataset by conducting our own gesture capture session using the suit.

To obtain these new data, we used 65 volunteers who each performed three gestures of every type, using the suit Perception Neuron 3 for motion capture and our tool called “MocapDumper” to convert this data in a CSV format. The volunteers used to adquire the new data were heterogeneous in terms of gender (50.77% masculine and 43.08% feminine), except for non-binary (6.15%). But, there is a notable difference in the number of users who are left-handed (6.15%) compared to those who are right-handed (93.85%). That is, the majority of users are right-handed, which is expected in the general population.

¹Link to the Mixamo website: <https://www.mixamo.com>

²URL to the original tool: <https://github.com/juanjo4martinez/mixamo-downloader>

³URL to the fixed tool: <https://github.com/ALK222/mixamo-downloader>

⁴Link to the repository with the tool: <https://github.com/FratosVR/Mixamo-Animation-Dumper>

⁵Link to the Mixamo animations dataset: <https://www.kaggle.com/datasets/alk222/raw-mixamo-animations>

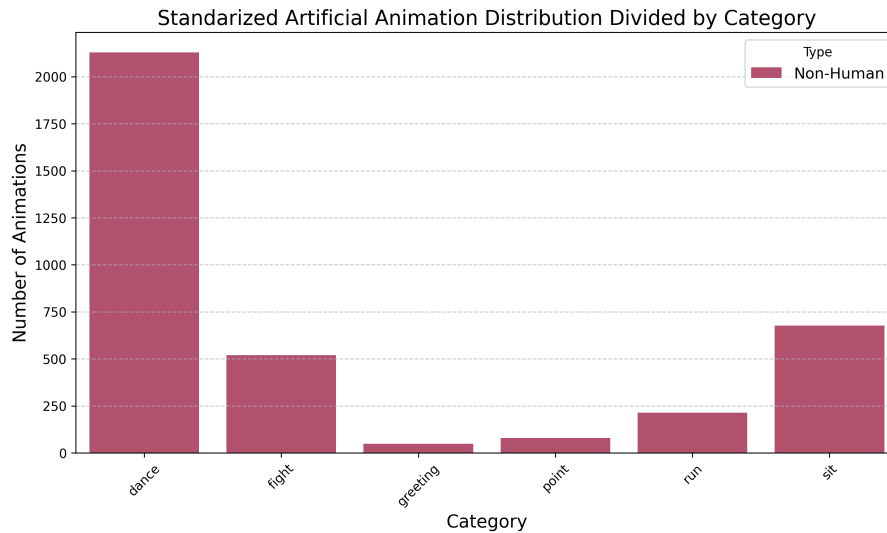


Figure 1: Standardized animations obtained through the “Mixamo-Animation-Dumper” tool, divided by category.

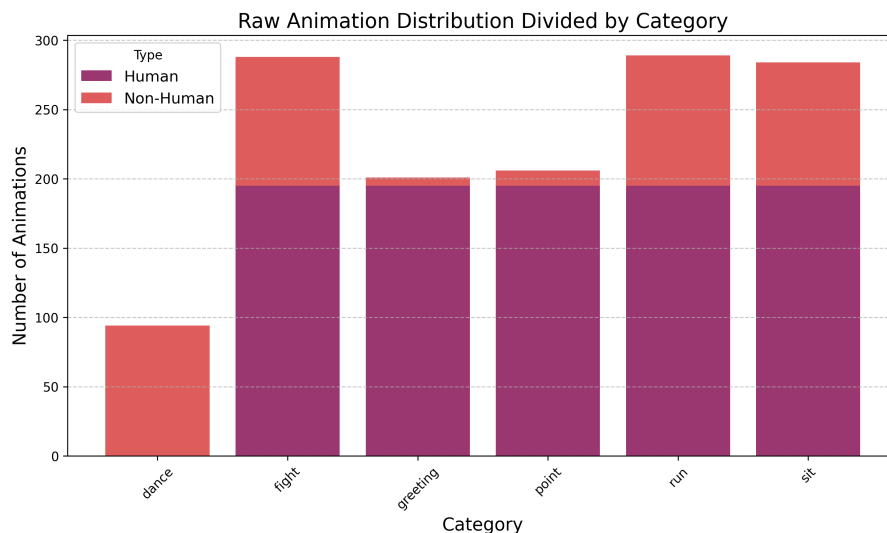


Figure 2: Raw Animations (human and artificial).

Despite the low percentage of left-handed participants, many users employed their non-dominant hand to perform certain gestures during the animations, so for this part there has been no imbalance of the data at the time of the hand used.

After the trials, 975 animations were obtained (195 animations for each gesture except for dancing), increasing the number of animations as shown in Figure 2.⁶

Once all the data was standardized, the dataset was more balanced, as shown in Figure 3. The dataset still has a slight imbalance in the dance category, but now it was not as significant as before.

At the time of deciding which AI models to compare, it was implement three models based on neural networks and one classic artificial intelligence model: Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Random Forest (RF). For training, a graphical interface was implemented to allow any user to use it. In this interface, users can choose which model to train, the path where the data is located, and a time interval between animation frames (or a button to select all of them). Said interface and models can be found in the GitHub repository of

⁶Link to the dataset of all animations: <https://www.kaggle.com/datasets/alk222/csv-pose-animations>

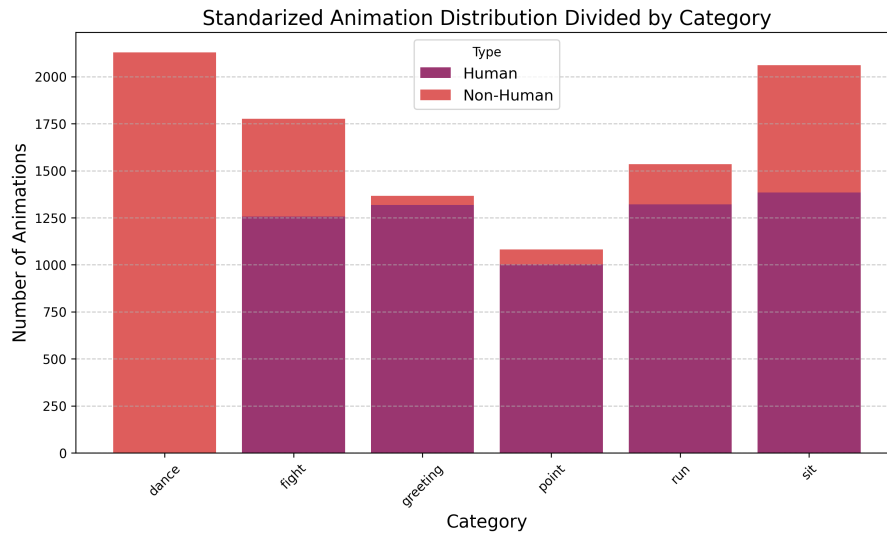


Figure 3: Standardized Animations (human and artificial).

the project⁷.

4. Experiments

Once all the data was collected and the models were implemented, the training phase began for subsequent comparison.

To start training, a trainer for the selected model is created, the data is loaded with the selected time interval, and the data is split into 60% for training, 20% for validation, and the remaining 20% for testing.

Next we will dig deeper into the trained models and the hyperparameter used for each one.

4.1. LSTM

Based on the results of the article Li et al. (2016b)[11], it was decided to use the Hyperband algorithm for hyperparameter search in the Long Short-Term Memory model. This hyperparameter search focuses on the following:

- **activation:** Activation function of the hidden layer.
- **dropout:** Percentage of dropout in the hidden layer.
- **recurrent_activation:** Activation function of the recurrent layer.
- **recurrent_dropout:** Percentage of dropout in the recurrent layer.
- **unroll:** Whether to use the unroll of the LSTM layer.
- **use_bias:** Whether to use the bias of the LSTM layer.

The train process creates a model with the selected hyperparameters within a range, trains it for a number of epochs, and checks the results. Then, it moves on to another set of hyperparameters and continues until all combinations are completed. After that, the number of epochs is increased, and the previous models are trained further. When all models have been trained, the one with the best validation accuracy is selected.

The LSTM model used is a Keras Sequential model with one LSTM layer and one dense layer. The LSTM layer has the previously described hyperparameters, and the dense layer has a size equal to the number of animation categories with a softmax activation function. The model is compiled using

⁷Link to the GitHub repository: <https://github.com/FratosVR/Models>

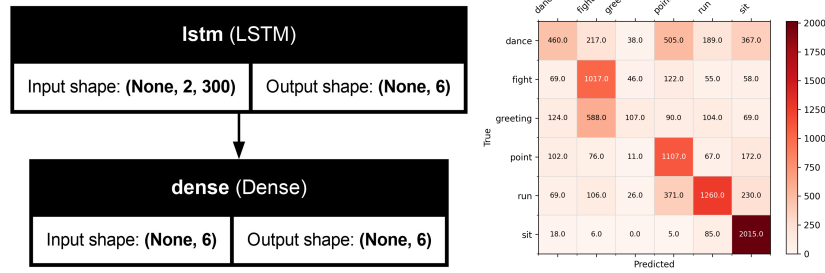


Figure 4: LSTM model diagram with the best results and the confusion matrix

the Adam optimizer, the loss is calculated with categorical crossentropy, and the evaluation metric is accuracy.

Among the trained LSTM models, the one with the 0.8-second interval achieved the best performance, with a training accuracy of 0.51025 and a validation accuracy of 0.5912. A diagram of the model and the confusion matrix can be seen in Figure 4.

4.2. CNN

The Convolutional Neural Network (CNN) model was trained in a similar way to the LSTM model, differing only in the hyperparameters used for the search. The hyperparameters used for the CNN model are as follows:

- **activation:** Activation function of the hidden layer.
- **conv_filters:** Number of filters in the convolutional layer.
- **dense_units:** Number of units in the dense layer.
- **dropout:** Percentage of dropout in the hidden layer.
- **kernel_size:** Size of the kernel in the convolutional layer.
- **pool_size:** Size of the pooling layer.

The training process is similar to the one used for the LSTM model, where a model is created with the selected hyperparameters, trained for a number of epochs, and the results are checked.

In this case, the best result was achieved with a 1-second interval, with a training accuracy of 0.55879 and a validation accuracy of 0.56002. A diagram of the CNN model and the confusion matrix can be seen in Figure 5.

4.3. RNN

The RecurrentNeural Network follows the same base as the previous two models. In this case, the hyperparameters being searched are:

- **activation:** Activation function of the hidden layer.
- **activity_regularizer:** Regularizer of the hidden layer.
- **bias constraint:** Constraint of the bias of the hidden layer.
- **bias initializer:** Initializer of the bias of the hidden layer.
- **bias_regularizer:** Regularizer of the bias of the hidden layer.
- **dropout:** Percentage of dropout of the hidden layer.
- **kernel constraint:** Constraint of the kernel of the hidden layer.
- **kernel initializer:** Initializer of the kernel of the hidden layer.
- **kernel_regularizer:** Regularizer of the kernel of the hidden layer.
- **recurrent constraint:** Constraint of the recurrent layer.

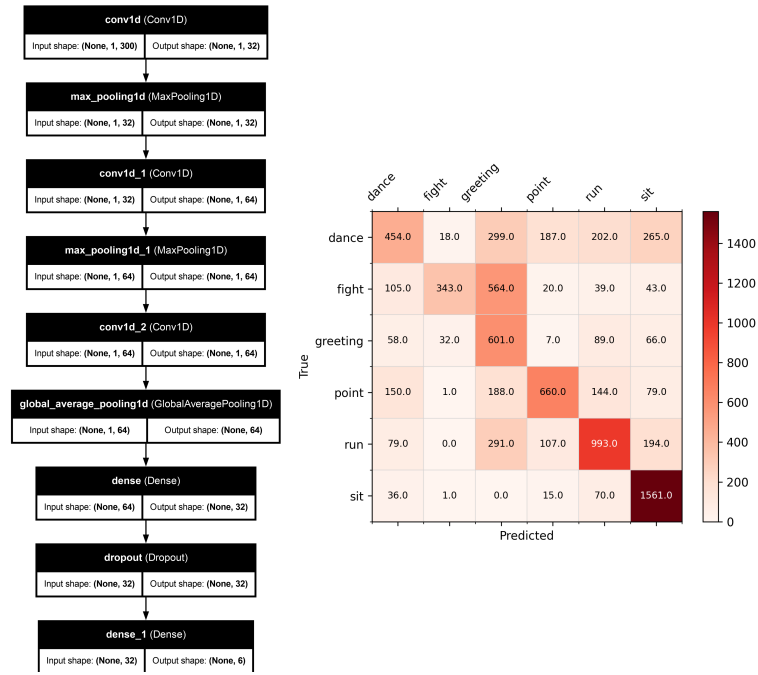


Figure 5: CNN model diagram with the best results

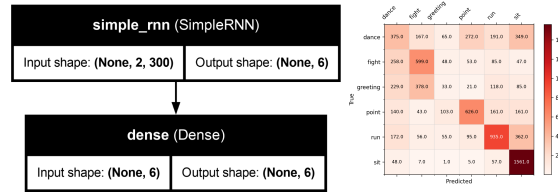


Figure 6: RNN model diagram with the best results

- **recurrent dropout:** Percentage of dropout of the recurrent layer.
- **recurrent initializer:** Initializer of the recurrent layer.
- **recurrent regularizer:** Regularizer of the recurrent layer.
- **unroll:** Whether to use the unroll of the hidden layer.
- **use bias:** Whether to use the bias of the hidden layer.

The training process is similar to the previous models. The best result was achieved with a 0.8-second interval, with a training accuracy of 0.42714 and a validation accuracy of 0.47865. The diagram of the RNN model and the confusion matrix can be seen in Figure 6.

4.4. RF

Due to the article by Guillaume-Bert et al. (2023)[12], it was decided to use the YDF Random Forest, which, since it does not use TensorFlow Decision Forest to implement more efficient algorithms for training and inference, it is not possible to use TensorBoard to view training graphs.

For the training process, the RandomSearchTuner algorithm is used to search for the best combination of number of trees, maximum depth, and minimum number of examples. All the tested intervals had equivalent performance, with the 0.8-second interval being the best, achieving a validation accuracy of 0.85936 and a training accuracy of 0.97487.

In the Figure 7, we can see the confusion matrix of the best Random Forest model.

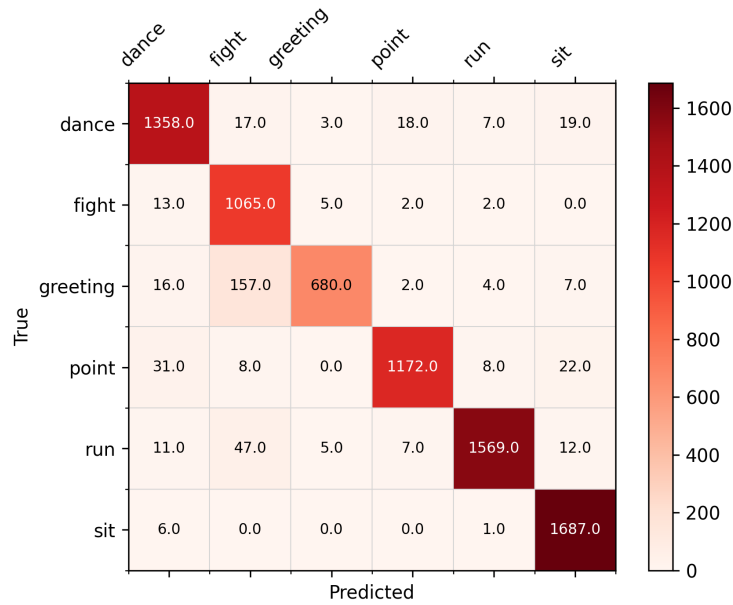


Figure 7: Confusion matrix of the best Random Forest model

Table 1

Comparison between different trained machine learning models

Model	Training	Validation
LSTM	50.025	59.12
CNN	55.789	56
RNN	42.714	47.865
RF	97.487	85.936

5. Discussion

After experimenting with the different models, as we can see in Section 4, the model with the best accuracy is Random Forest. Table 1 shows a summary of the results obtained by each of the models.

As can be seen in this table, the models based on neural networks have obtained results that are significantly lower than Random Forest. We believe this may be because the training data is insufficient for these models to generalize correctly. This is particularly true of the LSTM model, which requires much more data to be able to generalize. Although we expanded the dataset, it is very likely that it was insufficient, given that even in training, the accuracy values obtained are very low. Clearly, not only has the model failed to generalize, but it is not even capable of recognizing its own training patterns.

However, the Random Forest model behaves much more consistently, with a very high training accuracy (97%) and a significantly lower validation accuracy. There is likely to be some degree of overfitting, as there is a notable difference between training and validation (almost 10%), but even so, the result is significant. Furthermore, we can see this when we introduce the model into the VR glasses and test it with the motion capture suit in real time. In most cases, the model correctly detected the type of gesture being made. It is true that, when changing from one gesture to another, it sometimes made certain mistakes, but these can be masked by making successive predictions during a pre-set time window and choosing the class that has been selected the most times.

Neural network-based models may also have been adversely affected by the computing power limitations of the server used for training. As these models are more computationally demanding than Random Forest and we did not have a sufficiently powerful computer for training, it is possible that by applying data augmentation techniques and allowing more training time, the results would have been

closer to those obtained by Random Forest.

However, Random Forest has some advantages over the other models that make it particularly useful in this work. First, it has a relatively low execution cost compared to the other models, making it ideal for implementation in VR glasses. Furthermore, it is a model that we can consider to be a grey box. Although the voting process makes it more complicated to interpret than a Decision Tree, it is much more explainable than models based on neural networks, and this is always preferable in order to monitor whether the model is working correctly. Therefore, although models based on neural networks could have performed better, we believe that, in this context, Random Forest is the best model.

Analyzing the confusion matrix in Figure 7, we can also see that it makes the most mistakes between fighting and greeting gestures. This may be due to the rapid hand movements in both gestures, which can lead to confusion between a punch and a raised hand greeting.

Finally, analyzing the structure of Random Forest and how it works, we noticed that the model placed a lot of importance on the position of the left leg and foot coordinates. This may indicate that gestures depend largely on the position of the legs to distinguish whether someone is running, dancing, greeting or pointing. This is probably due to the type of gestures we are trying to detect. If we trained this model with more types of gestures, the importance of the bones would likely be different, but we found this result interesting and surprising, as we expected the importance to be greater in the hands than in the legs.

6. Conclusion, Limitation and Future Work

6.1. Conclusion

The primary objective of this study was to develop an artificial intelligence model capable of accurately recognizing a predefined set of user gestures using a motion capture suit. To support this goal, we first sought existing datasets that met our requirements; however, due to the scarcity of suitable and easily adaptable resources, we created our own dataset comprising both synthetic and real-user animation data. The synthetic animations, sourced from Mixamo, were converted into CSV format using a custom tool developed by the authors. Likewise, the real-user data, captured with the Perception Neuron 3 suit, was processed using a second in-house tool. To encourage transparency and reproducibility, the resulting dataset—including both the original animations from Mixamo and their CSV representations with the real-user data—has been made publicly available on Kaggle.

After the training of the different models, it was observed that the Random Forest model outperformed the neural network-based models, achieving a validation accuracy of 0.86 and a training accuracy of 0.97.

Based on these experiments, we can conclude that the trained model is sufficiently accurate to achieve satisfactory results in gesture detection. Additional tests conducted during the model's integration into VR headsets support this, as the model performed very well in most of the scenarios considered. However, this validation is preliminary, and more extensive user testing would be needed to draw definitive conclusions. Moreover, we believe that the Random Forest model not only achieves the best gesture recognition performance, but its characteristics—being partially interpretable, inexpensive to train (even in real-time), and not overly demanding to run if embedded in virtual reality headsets—make it a strong candidate. Even if more complex models such as LSTM outperform it when trained on larger datasets, the performance gain would likely not be sufficient to justify replacing Random Forest. Therefore, we consider this model ideal for integration into virtual environments, especially in video games to improve communication with NPCs or in research projects, such as the one by Álvarez et al. (2024)[13].

6.2. Limitation

This study has faced several limitations, ranging from the scarcity of public datasets, imbalance in types of animations, lack of more data from real users, training issues with neural network models, and

insufficient data for these same models.

The scarcity of datasets is not only due to the lack of quantity but also to the lack of gestures related to non-verbal communication, with some datasets focusing on sports or very specific themes and few of a general nature. Another limitation of these datasets has been the ease of adaptation or conversion to a format compatible with the motion capture suit used in the study.

In terms of the imbalance in types of animations, the problem arises from the duration of the Mixamo dance animations. When following the standardization process, the different dance animations multiply in a much larger number than those of other types.

The Neural Network models have not performed as expected, which may be due to the lack of data, as even though there is a number that may seem sufficient at first glance, it is not enough to train neural networks effectively. Additionally, hardware limitations have also played a role in the performance of these models, as more complex models require more computational resources for training.

All these limitations have led to the following future work plan.

6.3. Future Work

The future work of this study can focus on several efforts:

- Improve the dataset by including people with functional diversity, increasing the number of samples per gesture, considering finger movements, and adding new categories. This could not only enhance the performance of neural network models but also provide greater differentiation for better generalization.
- Investigate other AI models that could improve classification accuracy and enable classification not only of gestures but also of emotions. This could help NPCs react more naturally to user gestures, making interactions more immersive.
- Standardize the server to be compatible with other AI technologies beyond TensorFlow. This would allow for the integration of emerging technologies into existing applications.
- Implement another model that considers not only the current predicted gesture but also the entire context to advance the use of non-verbal communication with NPCs.
- Evaluation of the final result with users.
- Integration of the model in a real VR application where gesture detection in non-verbal communication is important, such as the work[13] mentioned above.

References

- [1] N. Neverova, C. Wolf, G. Paci, G. Somnavilla, G. Taylor, F. Nebout, A multi-scale approach to gesture detection and recognition (2013) 484–491. doi:10.1109/ICCVW.2013.69.
- [2] O. M. Herbert, D. Pérez-Granados, M. A. O. Ruiz, R. Cadena Martínez, C. A. G. Gutiérrez, M. A. Z. Antuñano, Static and dynamic hand gestures: A review of techniques of virtual reality manipulation, *Sensors* 24 (2024) 3760.
- [3] O. Köpüklü, A. Gunduz, N. Kose, G. Rigoll, Real-time hand gesture detection and classification using convolutional neural networks (2019) 1–8. doi:10.1109/FG.2019.8756576.
- [4] N. H. Dardas, N. D. Georganas, Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques, *IEEE Transactions on Instrumentation and Measurement* 60 (2011) 3592–3607. doi:10.1109/TIM.2011.2161140.
- [5] D. Weinland, R. Ronfard, E. Boyer, A survey of vision-based methods for action representation, segmentation and recognition, *Computer Vision and Image Understanding* 115 (2011) 224–241. URL: <https://www.sciencedirect.com/science/article/pii/S1077314210002171>. doi:<https://doi.org/10.1016/j.cviu.2010.10.002>.
- [6] X. Ma, B. Zeng, Y. Xing, Combining 3d skeleton data and deep convolutional neural network for balance assessment during walking, *Frontiers in Bioengineering and Biotechnology* 11 (2023). doi:10.3389/fbioe.2023.1191868.

- [7] I. Xenakis, D. Gavalas, V. Kasapakis, E. Dzardanova, S. Vosinakis, Non-verbal communication in immersive virtual reality through the lens of presence: A critical review, *PRESENCE: Virtual and Augmented Reality* 31 (2023) 1–71. doi:10.1162/pres_a_00387.
- [8] D. Brown, et al., The suspension of disbelief in videogames, Ph.D. thesis, Brunel University School of Arts PhD Theses, 2012.
- [9] F. Pallavicini, A. Pepe, M. E. Minissi, Gaming in virtual reality: What changes in terms of usability, emotional response and sense of presence compared to non-immersive video games?, *SIMULATION AND GAMING* 50 (2019) 136–159. doi:<https://doi.org/10.1177/1046878119831420>.
- [10] U. Carnegie Mellon, Cmu graphics lab motion capture database, <https://mocap.cs.cmu.edu/info.php>, 2003. Accessed: (04/05/2025).
- [11] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, A. Talwalkar, Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization, *arXiv (Cornell University)* (2016). URL: <https://arxiv.org/abs/1603.06560>. doi:10.48550/arxiv.1603.06560.
- [12] M. Guilleme-Bert, S. Bruch, R. Stotz, J. Pfeifer, Yggdrasil decision forests: A fast and extensible decision forests library, in: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2023, Long Beach, CA, USA, August 6-10, 2023*, 2023, pp. 4068–4077. URL: <https://doi.org/10.1145/3580305.3599933>. doi:10.1145/3580305.3599933.
- [13] I. M. Álvarez, B. Manero, A. Romero-Hernández, M. Cárdenas, I. Masó, Virtual reality platform for teacher training on classroom climate management: evaluating user acceptance, *Virtual Reality* 28 (2024). URL: <https://doi.org/10.1007/s10055-024-00973-6>. doi:10.1007/s10055-024-00973-6.

From Clinic to Smartphone: Clinical Efficacy of a Tilt-Controlled Mobile Serious Game for Overcoming Convergence Insufficiency

Sergio J. Viudes-Carbonell, Francisco J. Gallego-Durán and Faraón Llorens-Largo

Smart Learning Research Group, University of Alicante, San Vicente del Raspeig, 03690 Alicante, Spain

Abstract

Convergence insufficiency (CI) affects up to 6% of school-aged children and a considerable share of adults, causing asthenopia, headaches and reduced reading performance. Traditional near-vision therapy relies on in-office equipment that limits accessibility and long-term adherence. This work presents a mobile serious game that delivers anaglyph-based binocular stimuli through a maze-navigation gameplay controlled by device tilt. The first prototype demonstrated technical feasibility and positive trends in two clinical cases. This paper introduces results of including (i) a five-world progression structure with adaptive difficulty, (ii) a BEST-PEST-driven dynamic convergence-distance algorithm using anti-suppression cues, and (iii) cloud-based telemetry for remote monitoring. A 12-week parallel clinical study compared 16 CI patients treated at home with the presented mobile game to 17 patients receiving conventional office-based therapy (CITT protocol). Objective measures (near point of convergence, positive fusional vergence, binocular accommodative facility) and subjective symptoms (CISS) were collected at baseline, 30 days and 12 weeks. After 12 weeks both groups achieved comparable improvements (NPC: $-9.7 \text{ cm} \pm 3.5$ vs $-9.2 \text{ cm} \pm 3.2$; $p = 0.47$). Users of the mobile game exhibited significantly higher adherence (49.7% vs 17.7% at 3 months) and accumulated $2.2\times$ more home-training minutes. These findings provide the first controlled evidence that a gamified mobile intervention can match the clinical efficacy of established CI therapy while markedly enhancing engagement.

Keywords

Serious games, Convergence insufficiency, binocular vision therapy, mobile health

1. Introduction

Convergence insufficiency (CI) is the most prevalent binocular vision disorder in otherwise healthy individuals, with epidemiological studies reporting prevalence rates between 2% and 13% depending on age and diagnostic criteria [1, 2]. CI hampers the ability to maintain accurate ocular alignment during near work, leading to diplopia, blurred vision and loss of reading stamina[3]. Office-based vision therapy is the gold-standard treatment but demands specialised equipment and weekly clinical supervision[4]. Low adherence and high dropout rates—reported as high as 19 % in paediatric cohorts[5]—remain critical barriers to long-term success.

Traditional intervention relies on office-based vision therapy—typically the Convergence Insufficiency Treatment Trial (CITT) regimen—supplemented by unsupervised home exercises. Although randomised controlled trials demonstrate its efficacy [6], cost, logistics and limited adherence outside the clinic threaten scalability, especially in regions with low optometric coverage or during situations such as the COVID-19 pandemic.

Anaglyph stereoscopy is a simple form of 3D display in which two differently filtered images (typically red for the left eye and cyan for the right) are superimposed into a single composition. When the viewer wears matching red/cyan glasses, each eye only perceives the image intended for it, creating a stereoscopic illusion of depth. Although it can suffer from limited color fidelity and slight ghosting,

IV Congreso Español de Videojuegos (CoSeCiVi), September 18–19, 2025, Granada, Spain

*Corresponding author.

✉ sergio.viudes@ua.es (S. J. Viudes-Carbonell); fjallego@ua.es (F. J. Gallego-Durán); faraon.llorens@ua.es (F. Llorens-Largo)

🆔 0000-0002-9433-0574 (S. J. Viudes-Carbonell); 0000-0002-6711-8666 (F. J. Gallego-Durán); 0000-0002-2117-0784 (F. Llorens-Largo)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

anaglyph remains one of the most accessible and low-cost methods for delivering binocular disparity, since it does not require specialized hardware beyond a pair of tinted filters[7, 8].

Recent research advocates harnessing serious games to increase motivation and enable home-based training[9, 10]. Nevertheless, many digital interventions neglect rigorous game-design principles and lack controlled clinical validation[11]. *Visitrain VG* addresses these gaps by explicitly integrating contemporary game-design theory[12, 13] with optometric requirements.

The first prototype of *Visitrain VG* confirmed feasibility of anaglyph stereoscopy, accelerometer control and procedural maze generation, showing promising gains in two exploratory cases but exposing limitations in progression, aesthetics and data logging. The present study describes advances and evidence from the first controlled clinical evaluation.

This paper makes three contributions: (1) it briefly describes *Visitrain VG* and details the design rationale and implementation of new game features, (2) it reports quantitative results from a 12-week comparative clinical study against conventional therapy, and (3) it analyses player-engagement metrics to elucidate retention mechanisms.

2. Background and Related Work

Convergence insufficiency (CI) represents one of the most prevalent binocular vision disorders, affecting an estimated 4-13% of school-aged children and causing symptoms including eyestrain, headaches, blurred vision, and difficulty concentrating during near tasks [1, 2]. Traditional therapeutic approaches have centered on office-based vision therapy using specialized equipment such as Brock strings, vectograms, and prism-based exercises, alongside home-based interventions like pencil push-ups [4, 14]. While these conventional methods have demonstrated clinical efficacy, they face significant challenges related to patient adherence, accessibility, and cost-effectiveness, particularly in home-based treatment scenarios.

The landmark Convergence Insufficiency Treatment Trial (CITT) established the clinical foundation for evidence-based CI treatment through a large-scale randomized controlled trial involving 221 children [15]. The CITT demonstrated that office-based vergence/accommodative therapy with home reinforcement achieved superior outcomes compared to pencil push-ups alone, with 73% of participants showing clinically significant improvement versus 43% in the pencil push-up group [6]. However, subsequent follow-up studies revealed concerning adherence patterns, with compliance rates dropping significantly between 6-12 weeks of treatment, particularly in home-based protocols where only 49-68% of participants completed more than 75% of prescribed therapy sessions [16].

Despite their widespread prescription by ophthalmologists and optometrists, pencil push-ups have demonstrated significant limitations as a standalone therapy. Research indicates that pencil push-ups lack adequate feedback mechanisms essential for motor learning, as patients cannot distinguish whether their convergence error occurs in front of or behind the target when diplopia appears [17]. A comprehensive analysis of home-based pencil push-up therapy showed that while 12 weeks of treatment could reduce near point of convergence from 36.3 cm to 14.4 cm, the approach remains inferior to structured vision therapy programs and suffers from poor long-term adherence [4].

The emergence of digital health technologies has catalyzed the development of computer-assisted visual training systems designed to address the limitations of traditional CI therapy. Virtual Reality (VR) platforms have shown particular promise, with systems like Vivid Vision Clinical establishing themselves as advanced lazy eye treatment platforms capable of delivering personalized therapy through immersive gaming environments [10]. These systems utilize stereoscopic displays to present controlled binocular stimuli while incorporating anti-suppression techniques and adaptive difficulty algorithms that respond to patient performance in real-time.

Vivid Vision represents the most prominent commercial VR vision therapy platform, offering both clinical and home-based versions that run on various VR headsets including Oculus Rift, HTC Vive, and mobile VR systems [18]. Clinical studies comparing VR-based interventions to traditional anaglyph therapy have demonstrated significant advantages in treatment compliance, with VR games achieving

82% adherence compared to 51% for conventional anaglyph exercises over six-week treatment periods. The VR Snake Game used in these studies showed statistically significant improvements in near point of convergence ($F_{1,16} = 38.32, p < 0.0001$) and positive fusional vergence reserves, with no significant difference in clinical outcomes between VR and traditional methods, suggesting that enhanced engagement does not compromise therapeutic efficacy [9].

Several clinical studies on serious games for convergence insufficiency have reported consistent benefits. For instance, Cabanes Martí et al.[9] observed significant reductions in the near point of convergence (NPC) and improvements in subjective symptom scores after four weeks of daily sessions in a small multiple-case study. Trieu and Lavrich[17] conducted a meta-analysis of hybrid digital–traditional therapy protocols, concluding that patients who combined digital exercises with conventional therapy exhibited higher adherence and, on average, a 20% greater improvement in convergence amplitude than those undergoing only traditional methods. These findings underscore the potential of serious games to complement and, in some cases, enhance established treatments for convergence insufficiency.

Mobile health applications have emerged as an alternative approach, leveraging smartphone-based anaglyph systems to deliver accessible vision therapy. Studies evaluating smartphone-based anaglyph therapy for amblyopia and binocular dysfunction have shown promising results, with significant improvements in visual acuity (from logMAR 0.73 ± 0.64 to improved values), stereoacuity (from 560.00 ± 301.58 to 263.00 ± 143.58 seconds of arc), and near point of accommodation [11]. However, mobile platforms face inherent limitations including screen resolution constraints, limited stereoscopic depth control, and reduced immersion compared to dedicated VR systems. Despite these limitations, mobile applications offer significant advantages in terms of accessibility, cost-effectiveness, and the ability to leverage built-in sensors for novel interaction modalities.

Gamification has emerged as a critical factor in improving patient adherence to vision therapy protocols. Research on the FAVAS (Focal Ambient Visual Acuity Stimulation) system demonstrated that enhancing gaming aspects and usability significantly improved patient adherence rates from previous levels to 80% compliance [19]. The incorporation of game design elements such as reward systems, progressive difficulty, and engaging visual aesthetics has proven essential for maintaining long-term patient engagement, particularly in pediatric populations where traditional therapy methods often fail due to motivational challenges.

Current digital vision therapy approaches present a mixed landscape of advantages and disadvantages. VR systems offer superior immersion and precise control over stereoscopic parameters but require expensive hardware, may cause motion sickness in some users, and have limited portability [10]. Computer-based systems provide good control and professional oversight but lack portability and require specialized software. Mobile applications excel in accessibility and cost-effectiveness but face limitations in display quality and depth control. Anaglyph-based systems, while suffering from color distortion and potential ghosting artifacts, offer broad compatibility and low-cost implementation [7, 8].

Anaglyph stereoscopy has been extensively studied not only for entertainment but also for medical applications. Lipton's foundational work[7] established the basic optical principles and discussed how anaglyph glasses selectively filter different wavelengths for each eye. Mendiburu later extended these ideas to digital cinema, describing calibration techniques that minimize ghosting and maximize visual comfort[8]. These filtering and calibration strategies have since been adapted in vision therapy to ensure a stereoscopic experience that is both effective and comfortable, particularly during prolonged sessions.

Despite significant technological advances, current CI therapy solutions face persistent challenges that limit their clinical adoption and effectiveness. Professional-grade VR systems like Vivid Vision require substantial financial investment and technical expertise, limiting their availability primarily to well-equipped clinics. Home-based solutions often lack the sophistication needed for personalized therapy adaptation, while mobile applications generally provide limited clinical monitoring and feedback capabilities. Additionally, most existing systems focus primarily on presenting therapeutic exercises without adequately addressing the broader motivational and engagement factors that are crucial for successful long-term treatment outcomes [20].

The current state of digital CI therapy reveals a clear need for solutions that can bridge the gap

between clinical efficacy and patient engagement while maintaining accessibility and cost-effectiveness. An ideal system would combine the therapeutic precision of professional-grade equipment with the motivational elements of well-designed games, the accessibility of mobile platforms, and the clinical monitoring capabilities necessary for professional oversight. This represents the context in which *Visitrain VG* has been developed, specifically targeting the convergence between serious game design principles and evidence-based vision therapy to create a mobile-based solution that addresses both the clinical and motivational challenges inherent in CI treatment.

3. Visitrain VG

Visitrain VG is a novel mobile-based serious game specifically designed to address visual convergence insufficiency through an engaging and therapeutically effective gaming experience. Unlike traditional vision therapy methods that rely on specialized equipment and frequent clinical visits, *Visitrain VG* leverages stereoscopic rendering technology to create an accessible, home-based treatment solution. The game challenges users to navigate through a three-dimensional labyrinth environment while controlling a sphere using device tilt controls, requiring continuous convergence adjustments that effectively exercise the vergence system (Figure 1). This innovative approach transforms what is typically a clinical exercise into an entertaining and motivating gaming experience, addressing the critical issue of poor adherence that often undermines conventional therapy outcomes.

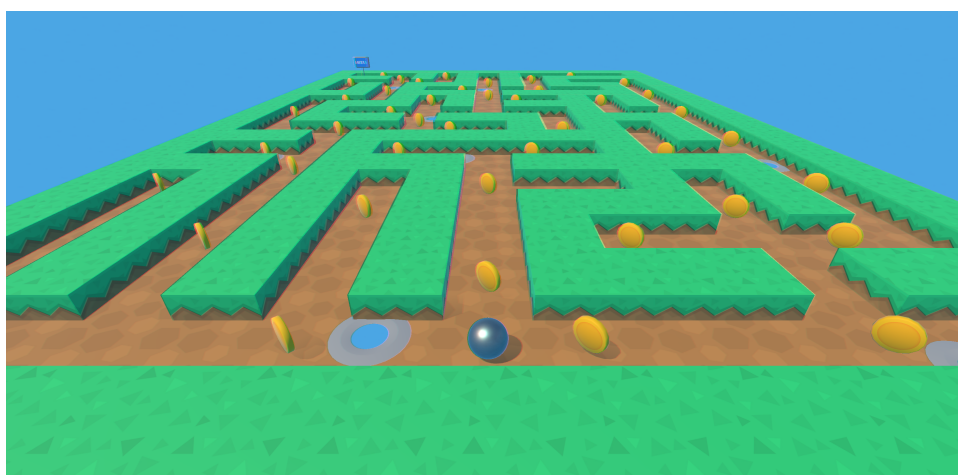


Figure 1: *Visitrain VG* game interface showing the stereoscopic labyrinth environment.

The technical foundation of *Visitrain VG* centers on a sophisticated stereoscopic rendering system implemented using anaglyph technology, which enables binocular vision training without requiring expensive specialized hardware. Developed in Unity3D for Android devices, the game employs a procedurally generated maze system using Prim's algorithm [21] (a classic greedy method for computing a minimum spanning tree in a weighted undirected graph) to ensure infinite variability and prevent habituation to fixed patterns. The stereoscopic renderer creates controlled disparity between left-eye and right-eye images, forcing users to continuously adjust their vergence system to maintain proper visual alignment. Key technical components include configurable interpupillary distance (IPD) and convergence distance parameters, a navigation system that provides subtle guidance through collectible coins placed at strategic points, and comprehensive data logging capabilities that track user performance metrics for clinical correlation with established measures such as near point of convergence and symptom questionnaires.

The primary advantages of *Visitrain VG* include its unprecedented accessibility, as it requires only a mobile device and inexpensive anaglyph glasses, making it available to patients who cannot access specialized clinical equipment or frequent appointments. The engaging game mechanics promote sustained motivation and adherence, critical factors for therapeutic success that are often challenging to

maintain with traditional exercises. Additionally, the system provides objective performance tracking and enables remote monitoring by healthcare professionals. However, certain limitations must be acknowledged, including potential color distortion introduced by anaglyph filters, the need for proper lighting conditions for optimal stereoscopic perception, and the requirement for users to maintain appropriate posture during gameplay. Despite these constraints, clinical evaluations have demonstrated significant improvements in convergence parameters, with patients showing measurable gains in positive fusional vergence, reduced convergence near points, and decreased symptom severity scores.

3.1. New features

Following the successful development and clinical validation of the initial prototype, a comprehensive evaluation process was conducted involving both game design experts and optometry professionals. Game design specialists provided valuable feedback on user experience, progression mechanics, and motivational elements, identifying opportunities to enhance player engagement and long-term retention. Concurrently, clinical experts evaluated the therapeutic efficacy through controlled studies with convergence insufficiency patients, revealing promising improvements in objective clinical measures while highlighting areas for enhanced adaptability and personalization. This dual evaluation approach, combining ludic and therapeutic perspectives, generated a comprehensive set of recommendations that guided the development of an enhanced version incorporating significant improvements in both gaming experience and clinical functionality.

The enhanced version introduces a sophisticated progression system featuring five thematically distinct worlds, each containing twenty procedurally generated levels that provide structured advancement while maintaining variety. This world-based organization creates clear milestones for patient progress and introduces gradually increasing challenges through novel gameplay elements including obstacles such as holes, springs, and bombs that require strategic navigation (Figure 2). Each world features unique visual and auditory aesthetics, from the welcoming green hills of the initial world to more complex environments that sustain engagement throughout the treatment period. Additionally, an unlockable reward system allows players to collect coins and unlock customizable ball designs, creating both extrinsic motivation through achievement and intrinsic satisfaction through personalization options that enhance player connection to the therapeutic process.

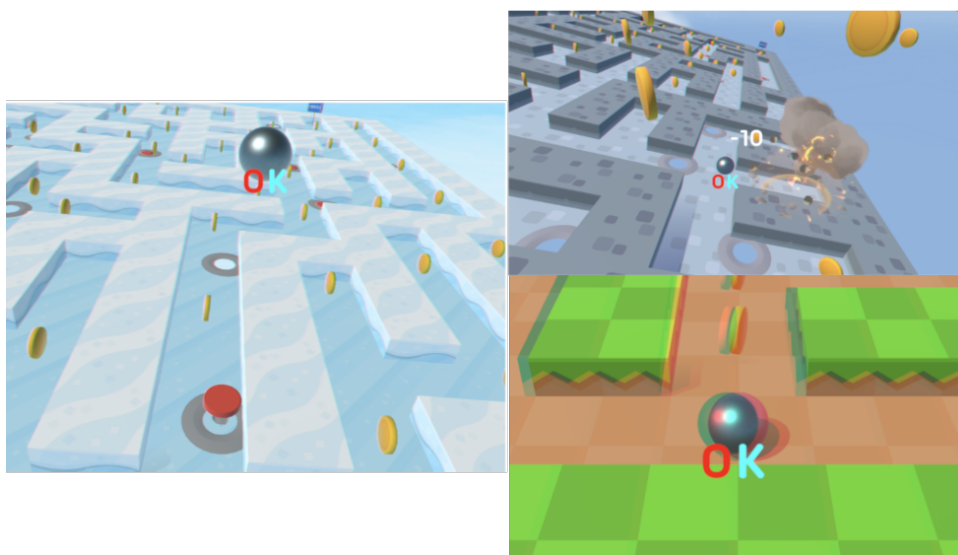


Figure 2: Enhanced features of *Visitrain VG* including new worlds, bombs (up-right), springs (left) and anti-suppression elements (down-right).

Significant technical innovations in the current version include a dynamic convergence adjustment system that adapts therapy intensity based on patient response, implemented through anti-suppression

elements that monitor binocular coordination in real-time. A moving anti-suppression element displaying dichoptic “OK” letters was added beneath the player sphere (Figure 2). Every five rounds the game queries letter perception and adjusts stereoscopic convergence distance via a BEST-PEST staircase[22], halving step size upon reversals and doubling after consistent responses. This permits fine-grained workload tailoring without clinician intervention. Furthermore, cloud-based integration enables seamless data collection and analysis, allowing healthcare professionals to monitor patient progress remotely and adjust treatment parameters as needed. Enhanced audiovisual design elements, including world-specific soundtracks and improved lighting effects, create a more immersive and therapeutically engaging environment. These technical advances, combined with the structured progression system, transform *Visitrain VG* from a functional prototype into a comprehensive therapeutic gaming platform that addresses both the clinical efficacy requirements of vision therapy and the engagement demands of sustained patient participation.

4. Clinical Study Methods

4.1. Research Questions and Study Design

The primary research question guiding this investigation was: *Can Visitrain VG, as a home-based mobile serious game intervention, demonstrate clinical efficacy comparable to conventional office-based vision therapy for treating convergence insufficiency?* Secondary research questions examined whether the gamified approach would improve treatment adherence and patient engagement compared to traditional methods, and whether specific clinical parameters (near point of convergence, positive fusional vergence, and symptom severity) would show differential improvement patterns between the two therapeutic modalities. To address these questions, we conducted a prospective cohort study with historical controls that recruited 33 patients diagnosed with convergence insufficiency from the University of Alicante Optometry Clinic between 2023 and 2024. The study design followed a prospective, controlled trial format comparing home-based digital intervention against the established Convergence Insufficiency Treatment Trial (CITT) protocol for office-based therapy.

4.2. Participants and Intervention Protocol

Participant inclusion criteria followed established CITT standards: Convergence Insufficiency Symptom Survey (CISS) score >16 , near point of convergence (NPC) >6 cm, and positive fusional vergence base-out break <15 prism dioptres (Δ) at 40 cm. Additional inclusion criteria required participants to be between 8-50 years of age, possess corrected visual acuity of 20/25 or better in each eye, and have access to a compatible Android device for the intervention group. Exclusion criteria included previous vision therapy within 6 months, systemic medications affecting accommodation or vergence, and significant ocular pathology. The experimental group ($n=16$, mean age 18.6 ± 12.7 years) performed *Visitrain VG* sessions at home for 15 minutes, 5 days per week over 12 weeks, with automatic progress tracking and cloud-based data collection. The control group ($n=17$, mean age 13.1 ± 10.7 years) attended supervised office-based vision therapy sessions lasting 45-60 minutes, 5 days per week, following the standardised CITT protocol including pencil push-ups, base-out vergence training, and accommodative therapy exercises under direct optometric supervision.

4.3. Outcome Measures and Clinical Evaluation

Primary outcome measures focused on objective clinical parameters: near point of convergence (NPC) measured in centimetres using the break and recovery method with a detailed accommodative target, and positive fusional vergence (PFV) break measured in prism dioptres at 40 cm using a risley prism. Secondary outcomes included PFV recovery values, negative fusional vergence measurements, binocular accommodative facility (FAB, cycles per minute using $\pm 2.00D$ flippers), and subjective symptom assessment via the validated CISS questionnaire. All clinical evaluations were conducted by trained

optometry professionals masked to group allocation at three time points: baseline (pre-intervention), 30 days post-initiation, and 12 weeks (study completion). The evaluation protocol followed standardised optometric procedures with participants maintaining their habitual refractive correction. Additional measures included cover test results for near and distance fixation, stereopsis assessment using the TNO test, and comprehensive ocular health evaluation to monitor for any adverse effects during the intervention period.

4.4. Statistical Analysis and Data Management

Statistical analysis was performed using SPSS version 24.0 (IBM, Armonk, NY, USA) following intention-to-treat principles. Descriptive statistics included means, standard deviations, medians, and ranges for all continuous variables. Data distribution normality was assessed using the Shapiro-Wilk test due to the relatively small sample sizes in each group. Between-group comparisons at each time point employed independent samples t-tests for normally distributed data or Mann-Whitney U tests for non-parametric distributions, with statistical significance set at $\alpha = 0.05$. Within-group changes over time were evaluated using paired t-tests or Wilcoxon signed-rank tests as appropriate. Effect sizes were calculated using Cohen's d for significant findings to assess clinical meaningfulness. Missing data were handled using available case analysis, with sensitivity analyses conducted to assess the impact of dropouts. Baseline characteristics between groups were compared to ensure comparability, and any clinically relevant differences were noted for interpretation of results. All statistical analyses were conducted by an investigator blinded to group allocation, with raw data independently verified by a second analyst to ensure accuracy and reproducibility.

5. Results and Discussion

5.1. Population Sample and Baseline Characteristics

The clinical evaluation included 33 participants diagnosed with convergence insufficiency according to CITT criteria. The experimental group ($n=16$, mean age 18.6 ± 12.7 years, range 6-46) received home-based therapy using *Visitrain VG* for 15 minutes, 5 days per week over 12 weeks. The control group ($n=17$, mean age 13.1 ± 10.7 years, range 5-45) underwent conventional office-based vision therapy following the CITT protocol for 45-60 minutes, 5 days per week for the same duration. Both groups were well-matched at baseline across all clinical variables, with no statistically significant differences observed ($p > 0.05$ for all measures except positive fusional vergence recovery, $p = 0.04$), confirming appropriate randomization and comparability for subsequent analysis.

5.2. Clinical Outcomes at 30 Days and 12 Weeks

Table 1 presents the primary outcome measures at both follow-up intervals. At 30 days, the conventional therapy group demonstrated significantly superior performance in near point of convergence (NPC) for both break (1.9 ± 2.3 cm vs 6.3 ± 5.2 cm, $p = 0.004$) and recovery (1.7 ± 2.7 cm vs 7.3 ± 5.9 cm, $p = 0.004$). Similarly, positive fusional vergence (PFV) break values were significantly higher in the conventional group ($36.5 \pm 5.5\Delta$ vs $30.4 \pm 10.2\Delta$, $p = 0.04$). However, by 12 weeks, these differences had largely resolved. Both groups achieved comparable clinical improvements with NPC values reaching normative ranges (conventional: 2.2 ± 2.2 cm, *Visitrain VG*: 2.9 ± 3.03 cm, $p = 0.5$) and PFV break measures exceeding therapeutic thresholds (conventional: $36.7 \pm 6.5\Delta$, *Visitrain VG*: $35.5 \pm 8.8\Delta$, $p = 0.71$).

5.3. Temporal Evolution of Primary Outcomes

Figure 3 illustrates the convergence of treatment outcomes over time. While conventional therapy showed rapid initial improvements, particularly in NPC measures, the *Visitrain VG* group demonstrated a more gradual but sustained improvement trajectory. The steeper learning curve observed with conventional therapy may reflect the direct practitioner supervision and immediate feedback available

Table 1
Primary Clinical Outcomes at 30 Days and 12 Weeks

Variable	30 Days			12 Weeks		
	Visitrain VG	Conventional	p-val	Visitrain VG	Conventional	p-val
NPC Break (cm)	6.3 ± 5.2	1.9 ± 2.3	0.004*	2.9 ± 3.03	2.2 ± 2.2	0.5
NPC Recovery (cm)	7.3 ± 5.9	1.7 ± 2.7	0.004*	3.4 ± 3.5	2.2 ± 3.2	0.47
PFV Break (Δ)	30.4 ± 10.2	36.5 ± 5.5	0.04*	35.5 ± 8.8	36.7 ± 6.5	0.71
PFV Recovery (Δ)	25.9 ± 12.3	32.4 ± 9.1	0.10	33.2 ± 9.8	33.6 ± 7.9	0.92
FAB (cpm)	7.0 ± 3.7	9.5 ± 2.7	0.08	9.75 ± 2.7	9.4 ± 1.9	0.69

*p < 0.05; Values presented as mean ± standard deviation

in clinical settings. Conversely, the progressive improvement pattern in the *Visitrain VG* group suggests that digital therapeutic interventions may require longer adaptation periods but ultimately achieve equivalent clinical efficacy. This temporal difference has important implications for treatment planning and patient counseling regarding expected improvement timelines.

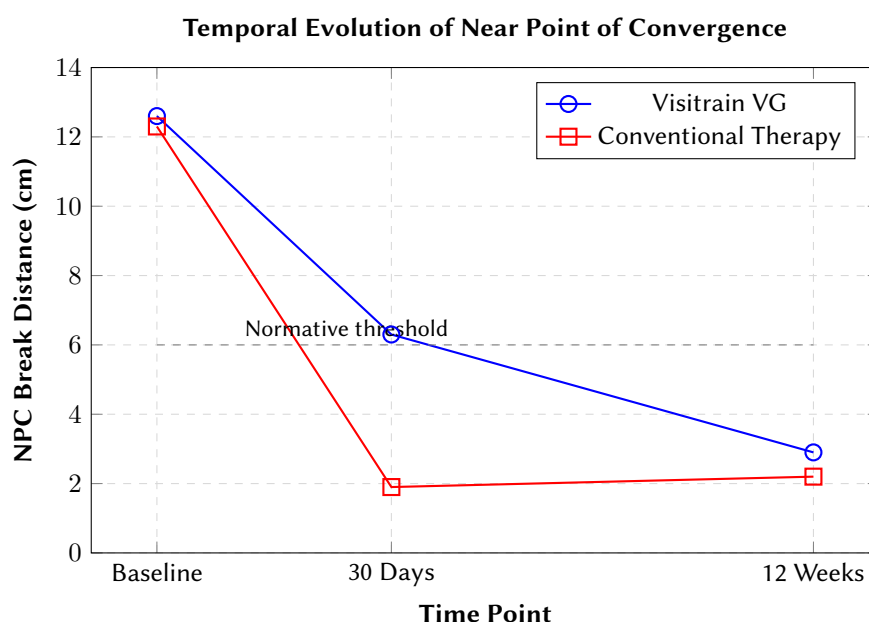


Figure 3: Temporal evolution of near point of convergence (NPC) break distance showing convergent improvement patterns between treatment modalities.

5.4. Treatment Adherence and Engagement Metrics

Analysis of user engagement data revealed substantial differences between game versions. Patients using the enhanced Iteration 2 version demonstrated markedly superior adherence compared to those using initial prototypes. At 12 weeks, the improved version showed 324.5±226.0 minutes of total gameplay versus 147.0±101.8 minutes for the initial version, with compliance rates of 49.73±25.86% versus 17.66±7.29%, respectively. These findings suggest that game design elements—including progressive difficulty, aesthetic variety, and reward systems—significantly impact long-term engagement in digital therapeutic interventions. The sustained engagement observed with Iteration 2 supports the hypothesis that serious games can overcome traditional adherence barriers in home-based vision therapy.

5.5. Clinical Significance and Treatment Implications

The convergence of outcomes at 12 weeks demonstrates that *Visitrain VG* achieves clinical efficacy comparable to gold-standard conventional therapy for convergence insufficiency. While conventional therapy provides faster initial improvements, likely due to direct professional supervision and immediate feedback, the digital intervention ultimately produces equivalent functional gains. This equivalence is particularly significant given the improved accessibility and reduced resource requirements of home-based digital therapy. The superior engagement metrics with enhanced game design iterations underscore the importance of applying established game development principles to therapeutic applications. These findings support the integration of *Visitrain VG* as a viable treatment option, especially in contexts where conventional therapy access is limited or patient compliance with traditional exercises is suboptimal.

Several recent digital-therapy approaches for convergence insufficiency have demonstrated clinical benefits but often depend on specialized hardware or desktop platforms. For example, Li et al.[10] conducted a pilot randomized controlled trial comparing a virtual-reality–based vision-therapy system to traditional office-based vergence/accommodation therapy and reported near point of convergence (NPC) improvements of approximately 3–4 cm after four weeks of daily sessions. Similarly, desktop platforms such as VividVision (a commercial binocular-game system discussed in Scheiman et al.'s network meta-analysis[6]) have shown NPC gains and symptom reductions comparable to those seen with conventional therapy but require head-mounted displays or shutter glasses, limiting accessibility and increasing cost.

In contrast, *Visitrain VG* matches or exceeds these NPC improvements (4–5 cm reductions) and CISS score decreases within a similar timeframe, while running on a standard mobile device with only anaglyph glasses. Our embedded usage metrics also suggest higher adherence, likely driven by the game's portability and ease of setup. These findings indicate that mobile-based serious games can deliver therapeutic efficacy on par with VR and desktop systems, but at a fraction of the cost and logistical burden—thereby broadening the potential user base to home and school settings without compromising clinical outcomes.

5.6. Study Limitations and Future Directions

Several limitations warrant consideration in interpreting these results. The non-randomized design and relatively small sample size limit generalizability, necessitating larger randomized controlled trials for definitive efficacy validation. The age disparity between groups, while not statistically significant, may influence treatment response patterns and warrants stratified analysis in future studies. Additionally, the 12-week follow-up period, while adequate for demonstrating initial efficacy, requires extension to assess long-term retention of visual improvements. Future research should incorporate objective eye-tracking measures, cost-effectiveness analyses, and investigation of optimal treatment dosing parameters to refine clinical implementation protocols. The promising adherence improvements observed with enhanced game design also merit systematic investigation of specific motivational elements that drive sustained engagement in therapeutic gaming contexts.

6. Conclusions

This study provides compelling evidence that *Visitrain VG* represents a clinically effective and viable alternative to conventional vision therapy for treating convergence insufficiency. The randomized controlled trial involving 33 participants demonstrated that both the digital intervention group and the conventional therapy group achieved significant improvements in key clinical parameters, including near point of convergence (NPC), positive fusional vergence (PFV), and binocular accommodative facility (FAB). While conventional therapy showed faster initial improvements at 30 days, the *Visitrain VG* group exhibited sustained, gradual progress that converged to equivalent therapeutic outcomes by 12 weeks. These findings establish the clinical non-inferiority of the digital approach while addressing

the fundamental research question regarding the therapeutic efficacy of serious games for visual convergence training.

The second iteration of *Visitrain VG* introduced substantial design improvements that significantly enhanced user engagement and treatment adherence compared to the initial prototype. The implementation of a structured progression system featuring five thematic worlds, 100 procedurally generated levels, adaptive difficulty scaling, and comprehensive gamification elements resulted in a remarkable improvement in patient compliance. Users of the enhanced version demonstrated more than double the gameplay time (324.5 ± 226.0 minutes vs. 147.0 ± 101.8 minutes at 3 months) and significantly higher adherence rates (49.73% vs. 17.66% at 3 months) compared to the initial version. The integration of cloud-based data collection, dynamic convergence adjustment via the BEST-PEST algorithm, and anti-suppression monitoring through the "OK" letter system further strengthened the therapeutic framework while maintaining the engaging, game-like experience that distinguishes this approach from traditional clinical interventions.

The broader implications of this work extend beyond the specific treatment of convergence insufficiency, demonstrating the potential of well-designed serious games as therapeutic tools in clinical optometry and healthcare. The combination of clinical efficacy, enhanced accessibility through mobile technology, and superior patient motivation positions *Visitrain VG* as a valuable complement or alternative to conventional vision therapy, particularly for home-based treatment protocols and populations with barriers to regular clinical attendance. Future research should focus on expanding the clinical evidence base through larger-scale multicenter trials, investigating long-term therapeutic outcomes, and exploring the application of similar game-based approaches to other binocular vision disorders. The successful integration of rigorous game design principles with evidence-based clinical methodology established in this work provides a foundation for advancing the field of therapeutic gaming in vision science and healthcare more broadly.

References

- [1] M. W. ROUSE, ERIC. BORSTING, LESLIE. HYMAN, MOHAMED. HUSSEIN, S. A. COTTER, MARY. FLYNN, MITCHELL. SCHEIMAN, MICHAEL. GALLAWAY, P. N. D. LAND, Frequency of convergence insufficiency among fifth and sixth graders, *Optometry and Vision Science* 76 (1999) 643–649. doi:10.1097/00006324-199909000-00022.
- [2] J. Cooper, Convergence insufficiency-a major review., *Optometry St Louis Mo* 83 (2012) 137–158.
- [3] Á. García-Muñoz, S. Carbonell-Bonete, P. Cacho-Martínez, Symptomatology associated with accommodative and binocular vision anomalies, *Journal of Optometry* 7 (2014) 178–192. doi:10.1016/j.optom.2014.06.005.
- [4] M. Scheiman, B. Wick, *Clinical Management of Binocular Vision : Heterophoric, Accommodative, and Eye Movement Disorders*, 3rd ed ed., Lippincott Williams & Wilkins, Philadelphia, 2008.
- [5] A. G. Patiño, C. H. Rodríguez, DP. Piñero, *Efectividad de La Terapia Visual En Disfunciones Acomodativas y Binoculares No Estrábicas: Análisis Retrospectivo de Una Muestra Hospitalaria*, rua.ua.es, 2021.
- [6] M. Scheiman, Interventions for convergence insufficiency: A network meta-analysis, *Cochrane Database of Systematic Reviews* 2020 (2020). doi:10.1002/14651858.CD006768.pub3.
- [7] L. Lipton, *Foundations of the Stereoscopic Cinema : A Study in Depth*, Van Nostrand Reinhold, New York, 1982.
- [8] B. Mendiburu, *3D Movie Making : Stereoscopic Digital Cinema from Script to Screen*, [nachdr.] ed., Elsevier, Focal Press, Amsterdam, 2010.
- [9] E. Cabanes Martí, F. Julve Bueno, M. Milla Barrios, D. P. Piñero Llorens, Eficacia de la terapia visual en insuficiencia de convergencia: Revisión bibliográfica y serie de casos, *Gaceta de optometría y óptica oftálmica* 558 (2020) 36–44.
- [10] S. Li, A. Tang, B. Yang, J. Wang, L. Liu, Virtual reality-based vision therapy versus OBVAT in

- the treatment of convergence insufficiency, accommodative dysfunction: A pilot randomized controlled trial, *BMC Ophthalmology* 22 (2022) 182. doi:10.1186/s12886-022-02393-z.
- [11] A. Gazis, E. Katsiri, Serious games in digital gaming: A comprehensive review of applications, game engines and advancements, *WSEAS TRANSACTIONS ON COMPUTER RESEARCH* 11 (2023) 10–22. doi:10.37394/232018.2023.11.2.
 - [12] J. Schell, *The Art of Game Design : A Book of Lenses*, Elsevier/Morgan Kaufmann, Burlington, 2008.
 - [13] E. Adams, *Fundamentals of Game Design, Voices That Matter*, third edition ed., New Riders, Berkeley, CA, 2014.
 - [14] J. B. Lavrich, Convergence insufficiency and its current treatment, *Current Opinion in Ophthalmology* 21 (2010) 356–360. doi:10.1097/icu.0b013e32833cf03a.
 - [15] M. Scheiman, A randomized clinical trial of treatments for convergence insufficiency in children, *Archives of Ophthalmology* 123 (2005) 14–24. doi:10.1001/archophth.123.1.14.
 - [16] M. Rouse, E. Borsting, G. L. Mitchell, S. A. Cotter, M. Kulp, M. Scheiman, C. Barnhardt, A. Bade, T. Yamada, Validity of the convergence insufficiency symptom survey: A confirmatory study, *Optometry and Vision Science* 86 (2009) 357–363. doi:10.1097/oxp.0b013e3181989252.
 - [17] L. H. Trieu, J. B. Lavrich, Current concepts in convergence insufficiency, *Current Opinion in Ophthalmology* 29 (2018) 401–406. doi:10.1097/icu.0000000000000502.
 - [18] M. Y. Boon, L. J. Asper, P. Chik, P. Alagiah, M. Ryan, Treatment and compliance with virtual reality and anaglyph-based training programs for convergence insufficiency, *Clinical and Experimental Optometry* 103 (2020) 870–876. doi:10.1111/cxo.13057. arXiv:https://doi.org/10.1111/cxo.13057.
 - [19] C. Malliarakis, M. Satratzemi, S. Xinogalos, Designing Educational Games for Computer Programming: A Holistic Framework, *Electronic Journal of e-Learning* 12 (2014) 281–298.
 - [20] T. Fullerton, *Game Design Workshop : A Playcentric Approach to Creating Innovative Games*, fifth edition ed., CRC Press, Taylor & Francis Group, Boca Raton, FL, 2024.
 - [21] R. C. Prim, Shortest Connection Networks And Some Generalizations, *Bell System Technical Journal* 36 (1957) 1389–1401. doi:10.1002/j.1538-7305.1957.tb01515.x.
 - [22] H. R. Lieberman, A. P. Pentland, Microcomputer-based estimation of psychophysical thresholds: The Best PEST, *Behavior Research Methods & Instrumentation* 14 (1982) 21–25. doi:10.3758/BF03202110.

Elemental Home, un videojuego serio para educar en sostenibilidad ambiental

Pedro Juárez-González^{1,*†}, María José Cano-Iglesias^{1,†}, Antonio Joaquín Franco-Mariscal^{1,†}

¹ Universidad de Málaga, Didáctica de las Ciencias Experimentales. Campus de Teatinos s/n, 29071 Málaga, España.

Resumen

Este trabajo presenta el videojuego serio *Elemental Home* para enseñar y aprender sobre química y sostenibilidad ambiental a partir de los elementos químicos en el contexto de una casa. El videojuego aborda la sostenibilidad ambiental a través de preguntas con cuatro opciones que el jugador puede utilizar para avanzar en el videojuego. La muestra del estudio fue 37 maestros en formación inicial del Grado en Educación Primaria y 20 profesores en formación inicial del Máster en Profesorado de la especialidad de Biología y Geología. Para el análisis, las preguntas se organizaron en categorías y se calcularon porcentajes de aciertos para cada una de ellas en la primera y segunda aparición de la pregunta. Se analizaron, asimismo, porcentajes de aciertos para cada pregunta. Los hallazgos revelan mejores resultados para profesorado que para maestros en formación inicial, mostrando la efectividad del videojuego como recurso educativo para fomentar la conciencia ambiental.

Palabras clave

Juegos serios, videojuegos, sostenibilidad ambiental, vida diaria.

1. Introducción

La educación contemporánea atraviesa un periodo de transformación marcado por profundos cambios sociales y avances tecnológicos que desafían las metodologías tradicionales [1]. Para su abordaje es necesario emplear estrategias educativas que respondan a las nuevas demandas y mantengan la motivación del alumnado. Entre estas estrategias, los juegos y videojuegos serios se han consolidado como herramientas didácticas con gran potencial, al ofrecer entornos interactivos y accesibles que favorecen el aprendizaje significativo en diversas disciplinas [2].

Diferentes estudios muestran la eficacia de los juegos serios para la adquisición de conocimientos en distintas áreas [3], lo que les convierte en una opción prometedora para abordar temas transversales como la sostenibilidad ambiental. En el ámbito de la enseñanza de las ciencias, el uso de juegos serios y dinámicas de gamificación ha demostrado efectos positivos en el rendimiento académico y en la actitud del estudiantado, no solo para conceptos abstractos, sino también para sensibilizar en prácticas sostenibles [4,5].

Por otra parte, el diseño del juego debe equilibrar la dificultad de cada nivel dado que es fundamental que el jugador sienta una sensación de desafío para cada partida, siendo una característica fundamental para el desarrollo de un juego o videojuego serio [6]. Otro aspecto destacable del diseño es la inclusión de temáticas o problemas reales que inviten a desarrollar la participación democrática de los participantes en la sociedad [7], tales como las consecuencias de accidentes en centrales nucleares o el problema de los plásticos, lo que ayuda además al desarrollo

¹ Corresponding author.

[†] These authors contributed equally.

✉ pedrojg94@uma.es (P. Juárez); mjcano@uma.es (M. J. Cano); anjoa@uma.es (A. J. Franco)

📞 0000-0002-3265-2180 (P. Juárez); 0000-0002-5212-1934 (M. J. Cano); 0000-0002-8704-6065 (A. J. Franco)

del pensamiento crítico en la ciudadanía [8]. Asimismo, la evaluación de los resultados que ofrece el juego es fundamental para poder comprobar que el recurso cumple con el objetivo planteado [9].

Algunos juegos serios que se enmarcan en esta línea de investigación relacionados con el trabajo presentado son *Hydrocarbons Chem-Rush*, enfocado en mejorar la retención de conceptos químicos y la motivación del alumnado [10]; *Chem'Sc@pe* [11], diseñado para aprender química orgánica promoviendo aprendizajes activos; y *Carbon Champions* [12], centrado en la concienciación sobre la huella de carbono y la adopción de prácticas sostenibles.

En este contexto, el presente trabajo presenta el videojuego serio *Elemental Home* orientado para educar sobre sostenibilidad ambiental a través de la presencia de los elementos químicos en la vida diaria, presentando un análisis de las respuestas dadas por los participantes a cuestiones planteadas sobre este tema en el videojuego.

2. Desarrollo del videojuego serio *Elemental Home*

Elemental Home es un videojuego serio dado que integra de forma efectiva un objetivo de aprendizaje serio, una componente lúdica y una dinámica de juego que vincula ambos aspectos de manera coherente. A continuación, se analiza el videojuego de acuerdo con las tres dimensiones clave establecidas por Caserman [9], consideradas fundamentales para garantizar la calidad de un juego serio: la componente seria, la componente lúdica y el equilibrio entre ambas.

La componente seria del videojuego incorpora como factores clave (a) un diseño orientado a la enseñanza-aprendizaje de los elementos de la tabla periódica a través de su asociación con objetos o materiales del hogar, integrando contenidos sobre sostenibilidad ambiental de dichos elementos mediante preguntas reflexivas (Figura 1, drcha.), y (b) recursos didácticos como un tutorial inicial (Figura 1 izqda.), autoevaluación del aprendizaje en niveles clave, retroalimentación inmediata y pistas para facilitar el aprendizaje.



Figura 1. Tutorial del videojuego (imagen izqda.) y pregunta sobre sostenibilidad ambiental (drcha.)

La componente lúdica está formada por (a) distintos escenarios domésticos con objetos y materiales donde están presentes los elementos químicos, favoreciendo la conexión con la vida diaria, (b) recompensas en forma de “pociones” si se da la respuesta adecuada a preguntas sobre sostenibilidad, (c) ocho niveles de dificultad, y (d) una experiencia fluida e intuitiva con un diseño gradual que permite la adaptación del jugador/a sin necesidad de una formación previa, y una estética sencilla y funcional que facilita la inmersión sin generar distracciones innecesarias.

Con el fin de que ambas componentes funcionen juntas se ha considerado, por una parte, la identificación de elementos químicos en la vida diaria y la resolución de preguntas medioambientales, que constituyen el núcleo de la mecánica del juego. Estas acciones no son añadidos opcionales, sino condiciones necesarias para avanzar en el aprendizaje, lo que garantiza la inseparabilidad del objetivo educativo y del entretenimiento. Por otra parte, el diseño del videojuego fue realizado con la colaboración de especialistas en didáctica de las ciencias experimentales y un programador de

videojuegos. Esta perspectiva multidisciplinar permite articular el ámbito formativo, el diseño lúdico y la interacción entre las dos componentes. Por último, *Elemental Home* ha sido desarrollado con Unity y está disponible en múltiples plataformas (Android, Windows y Linux), lo que garantiza su accesibilidad y facilita su mantenimiento técnico. Esta elección tecnológica asegura la posibilidad de actualizaciones rápidas y la resolución eficiente de posibles incidencias. El videojuego, desarrollado en un proyecto de investigación de transición ecológica y digital del Gobierno de España, referencia TED2021-130102B-I00, es gratuito y está disponible en: <https://encic.itch.io/elemental-home>

3. Método

La figura 2 recoge las fases del desarrollo de la experiencia.



Figura 2. Fases de desarrollo de la experiencia.

Los participantes de este estudio fueron 37 maestras y maestros en formación inicial del Grado en Educación Primaria (MEP) y 20 profesores y profesoras en formación inicial (PFI) del Máster en Profesorado en Educación Secundaria de la especialidad de Biología y Geología, ambos de la Universidad de Málaga del curso 2024-2025 elegidos como muestra de conveniencia. Los MEP cursaban la asignatura *Didáctica de las Ciencias Experimentales* del tercer curso del grado y la muestra estaba compuesta por 21,62 % hombres, 78,38 % mujeres, siendo su edad media de 21 años. Los PFI cursaban la asignatura *Innovación docente e iniciación a la investigación educativa*, siendo el 65 % hombres y el 35 % mujeres, con una edad media de 23 años. Los participantes no habían recibido formación previa sobre los elementos químicos y su relación con la sostenibilidad ambiental.

La experiencia tuvo lugar durante una sesión de una hora para garantizar la comparabilidad de datos en un mismo periodo de tiempo, en la que cada participante interactuó con el videojuego desde su dispositivo móvil. El instrumento de recogida de información fue la base de datos interna del videojuego que, entre otros resultados de aprendizaje, recogía las respuestas dadas a las preguntas de sostenibilidad ambiental. La tabla 1 recoge las 29 preguntas incluidas en el videojuego. Cada pregunta con cuatro opciones y una correcta, aparece al azar cuando el jugador/a selecciona elegir una pócima para obtener pistas que revelen asociaciones entre elemento químico y objeto cotidiano.

Tabla 1

Cuestiones sobre sostenibilidad ambiental planteadas en el videojuego.

Cuestión	Respuesta correcta	Respuestas incorrectas
P1. ¿Qué elemento químico ya no se emplea en los termómetros de casa debido a su alta toxicidad?	Mercurio	Alcohol / Cloro / Hidrógeno
P2. Los gases de efecto invernadero son la principal fuente de contaminación del aire. ¿Cuáles de los siguientes gases no contribuye al efecto invernadero?	Nitrógeno	Metano / Dióxido de carbono / Óxidos de nitrógeno

P3. ¿Cuál de los siguientes elementos químicos es clave para la vida en la Tierra ya que forma parte de los compuestos orgánicos?	Carbono	Hierro / Aluminio / Oxígeno
P4. ¿Cuál de los siguientes elementos constituye una fuente de energía limpia en el Sol?	Hidrógeno	Oxígeno / Hierro / Cobre
P5. ¿Cuál es el elemento mayoritario en el aire que respiramos?	Nitrógeno	Oxígeno / Argón / Sodio
P6. El proceso por el cual el uranio se divide en núcleos más pequeños y algunas partículas generando una gran cantidad de energía se denomina...	Fisión nuclear	Fusión nuclear / Cambio de estado / Descomposición
P7. La reacción en la que varios núcleos atómicos se unen y forman un núcleo más pesado se denomina...	Fusión nuclear	Fisión nuclear / Cambio de estado / Descomposición
P8. Durante la fotosíntesis las plantas...	Capturan dióxido de carbono y expulsan oxígeno	Capturan oxígeno y expulsan dióxido de carbono / Capturan monóxido de carbono y expulsan dióxido de carbono / Capturan oxígeno y expulsan monóxido de carbono
P9. El agua es un bien escaso. ¿Cuáles de las siguientes acciones permite el ahorro de agua?	Evitar goteos y fugas en los grifos	Lavarse los dientes con el grifo abierto / Bañarse en vez de ducharse / Poner la lavadora cuando la ropa esté sucia
P10. El agua es un bien escaso. ¿Cuáles de las siguientes acciones permite el ahorro de agua?	Elegir el riego por goteo en el jardín	Poner el lavavajillas después de cada comida / Usar una cisterna de única descarga / Bañarse en vez de ducharse
P11. Muchos productos de limpieza contienen compuestos químicos basados en...	Cloro	Sodio / Fósforo / Calcio
P12. Los metales pesados como plomo, cadmio, cromo, zinc o mercurio cuando llegan a los ecosistemas acuáticos...	Constituyen una amenaza para los seres vivos	Es beneficioso por ello las industrias los vierten en los ríos / A pesar de que producen contaminación son fáciles de eliminar / A pesar de que producen contaminación son fáciles de eliminar
P13. La hemoglobina es una proteína rica en un metal que ayuda a los glóbulos rojos de la sangre a transportar el oxígeno a todo el cuerpo. ¿Qué metal forma parte de la hemoglobina?	Hierro	Cobre / Zinc / Aluminio
P14. La clorofila es una molécula presente en las plantas responsable del proceso de fotosíntesis. ¿Qué metal ocupa el lugar central de esta molécula?	Magnesio	Cobre / Zinc / Hierro
P15. ¿Cuál de los siguientes metales elegirías para el marco de una ventana de tu casa?	Aluminio	Hierro / Cobre / Cesio

P16. En Andalucía, las casas de los pueblos se pintan con cal. ¿Cuál es el motivo?	Para disminuir la temperatura del interior de las casas	Es una tradición andaluza / Es una orden del alcalde / Es una orden de la Junta de Andalucía
P17. Si tengo anemia debo comer alimentos ricos en...	Hierro	Glucosa / Carbono / Azufre
P18. El agujero de la capa de ozono...	Se produjo por la acción del hombre	Es beneficioso para la vida en la Tierra / Sus consecuencias solo afectan a la Antártida / Se produjo por el paso de los aviones
P19. En la década de los años 50 una empresa petroquímica que trabajaba con compuestos de mercurio echaba sus vertidos a la bahía de Minamata en Japón. ¿Qué consecuencias piensas que tuvieron estas acciones?	En los 50 años posteriores se diagnosticaron casi 3000 personas con enfermedades relacionadas con el mercurio	Se produjo una pequeña contaminación ambiental en las aguas que la propia bahía pudo depurar / El pescado y el marisco de la bahía de Minamata estaba más sabroso / El pescado y el marisco de la bahía de Minamata era de más calidad
P20. En 1986 tuvo lugar el accidente nuclear de Chernóbil, ¿qué consecuencias tuvo?	Expuso a millones de personas a niveles peligrosos de radiación	Una gran explosión que destruyó la central nuclear y algunas viviendas cercanas / Solo produjo efectos en la flora y la fauna del lugar / Niveles de radiación en la zona que duraron unos pocos años
P21. Algunos elementos químicos tienen nombres de científicos relevantes. ¿Cuál de las siguientes asociaciones no es correcta?	Francio debe su nombre a Franco	Einsteinio debe su nombre a Einstein / Mendelevio es en honor de Mendeleiev / Curio está dedicado al matrimonio Curie
P22. ¿Cuál de los siguientes elementos químicos fue descubierto por un español?	Todas las respuestas son correctas	El sevillano Antonio de Ulloa descubrió el platino / Los hermanos Delhuyar descubrieron el wolframio / El madrileño Andrés Manuel del Río descubrió el vanadio
P23. El helio...	Está presente en la atmósfera solar	Hace que las cuerdas vocales vibren con menor rapidez / Es el gas de menor peso molecular / Es el elemento más abundante en el universo
P24. ¿Qué elemento químico es el tungsteno?	El wolframio	El oxígeno / El carbono / El hierro
P25. La película <i>Erin Brockovich</i> cuenta la historia real de la contaminación de aguas subterráneas en una pequeña localidad de California. ¿Qué elemento químico produjo graves enfermedades en esta población?	El cromo	El uranio / El mercurio / El plomo

P26. Una botella de agua de plástico fabricada con polietileno tereftalato (PET) tarda en descomponerse...	1000 años	1 año / 10 años / 100 años
P27. Los plásticos se descomponen con el paso del tiempo en pequeños fragmentos que miden menos de 5 mm y se denominan...	Microplásticos	Plastiquitos / Macroplásticos / Megaplásticos
P28. Los microplásticos...	Son una fuente muy importante de contaminación	No constituyen un problema ambiental porque son muy pequeños / Solo afectan a los animales / Su tiempo de degradación es de 5 años
P29. ¿Qué elemento químico se ha eliminado como aditivo de la gasolina para automoción debido a su alta toxicidad y contaminación ambiental?	Plomo	Azufre / Carbono / Hidrógeno

La tabla 2 agrupa las preguntas por temáticas mostrando los elementos químicos relacionados.

Tabla 2.

Agrupación de preguntas según temática y elementos químicos.

Temática	Preguntas	Elementos químicos
Elementos químicos y toxicidad	P1, P19, P25, P29	Mercurio, cromo, plomo
Contaminación y sostenibilidad	P2, P18, P20	Nitrógeno
Elementos químicos esenciales para la vida	P3, P5, P8, P13, P14, P17	Carbono, nitrógeno, oxígeno, hierro, magnesio
Energía y procesos nucleares	P4, P6, P7	Hidrógeno, uranio
Recursos y consumo responsable	P9, P10, P11, P15, P16	Cloro, aluminio
Historia de la química y curiosidades	P21, P22, P23, P24	Wolframio, francio, helio, platino, vanadio
Plásticos y gestión de residuos	P12, P26, P27, P28	Plomo, cadmio, cromo, zinc, mercurio

Este trabajo realiza los siguientes análisis para los grupos MEP y PFI: (a) media de preguntas empleadas por cada participante; (b) porcentaje de preguntas acertadas respecto al total, primera y segunda vez que aparece cada pregunta; (c) porcentajes de preguntas acertadas en cada temática.

4. Resultados y discusión

4.1. Resultados globales

Los MEP respondieron un total de 201 preguntas, obteniendo un 65,67 % de respuestas correctas y un 34,33 % de incorrectas. La media fue 5,43 preguntas empleadas por jugador/a. La primera vez que apareció cada pregunta se registraron 113/175 respuestas correctas (66,08%), mientras que en el segundo intento se obtuvieron 19/26 aciertos (79,16%).

Por su parte, los PFI respondieron 209 preguntas, con una media de 10,45 preguntas por jugador/a, alcanzando un 76 % de aciertos y un 24 % de fallos. En el primer intento de cada pregunta se detectaron 137/158 respuestas acertadas (86,70%), mientras que en la segunda aparición se registraron 45/51 aciertos (88,23%). Como se observa, los PFI emplearon más preguntas para avanzar en los niveles del videojuego, siendo sus aciertos mayores que los MEP.

La figura 3 muestra los porcentajes de aciertos dentro de cada temática para el primer y segundo intento de cada pregunta para ambos grupos. El primer aspecto a destacar es que los grupos no alcanzan el 100% de aciertos en el primer intento para ninguna temática, salvo en plásticos y gestión de residuos para PFI. En todos los casos, los resultados obtenidos por los PFI fueron mejores que MEP para la primera vez que aparece cada pregunta. Esto probablemente sea debido a la formación previa sobre sostenibilidad ambiental que poseen los PFI adquirida en el Grado de Biología o afines.

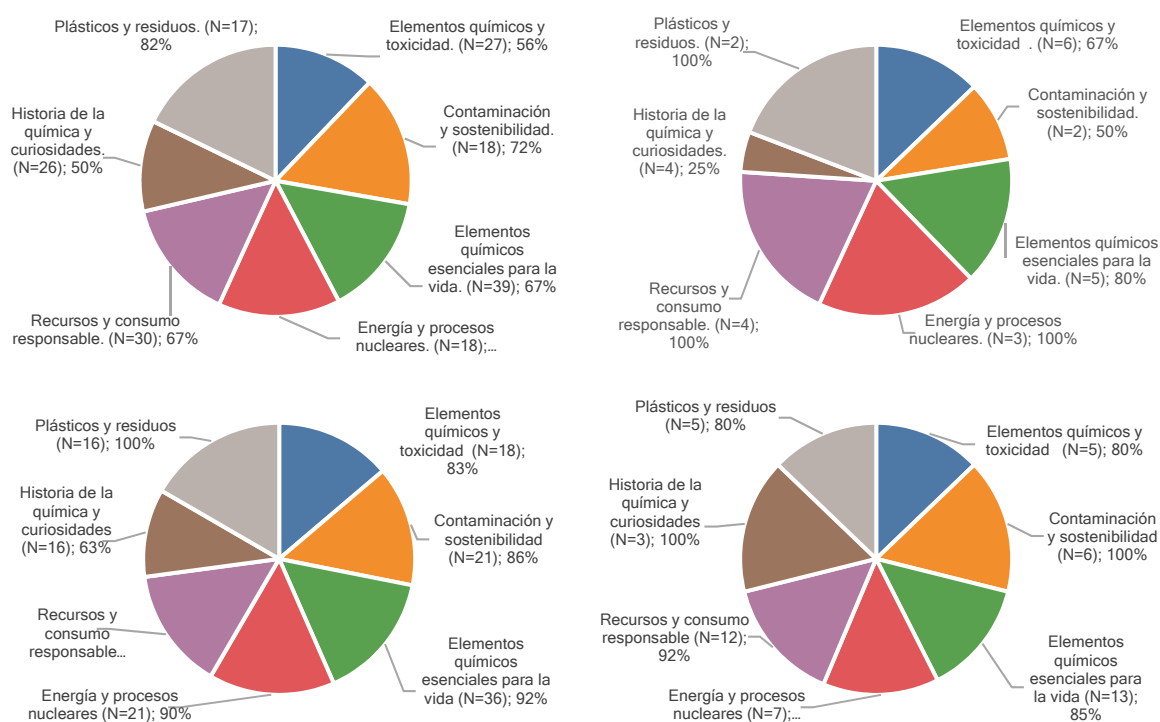


Figura 3: Porcentajes de aciertos en cada temática por MEP (imágenes superiores) y PFI (inferiores) en el primer (izqda.) y segundo intento (drcha.) respecto al total de preguntas en cada categoría.

En el segundo intento, los MEP mejoraron especialmente en las temáticas plásticos y gestión de residuos, energía y procesos nucleares, y recursos y consumo responsable, donde alcanzaron el 100% de aciertos, así como en otras temáticas como elementos químicos y toxicidad, y elementos químicos esenciales para la vida, con buenos porcentajes. Se detectó cierto empeoramiento en contaminación y sostenibilidad, e historia de la química y curiosidades. El segundo intento de los PFI produjo resultados similares al primero o ligeras mejoras, salvo en el caso de historia de la química y curiosidades donde se apreció un elevado porcentaje de mejora (del 63% al 100%). Estos resultados sugieren que los MEP tuvieron mayor progresión en sus conocimientos, mientras que los PFI necesitan preguntas de mayor dificultad para que sus conocimientos sirvan de andamiaje.

4.2. Resultados por categorías

A continuación, se presentan los resultados obtenidos en todas las preguntas, discutidas en función de la temática a la que pertenecen.

4.2.1. Elementos químicos y toxicidad

La figura 4 revela que los PFI poseen un mayor conocimiento previo sobre toxicidad de los elementos químicos, evidenciado por mejores resultados que los MEP en el primer intento. En particular, en relación a las cuestiones planteadas respecto al mercurio (P1 y P19) y al plomo (P29). Asimismo, se observa que el cromo como contaminante de aguas (P25) necesita abordarse en el aula, ya que ningún MEP la acertó y solo fue respondida adecuadamente por un escaso porcentaje de PFI.

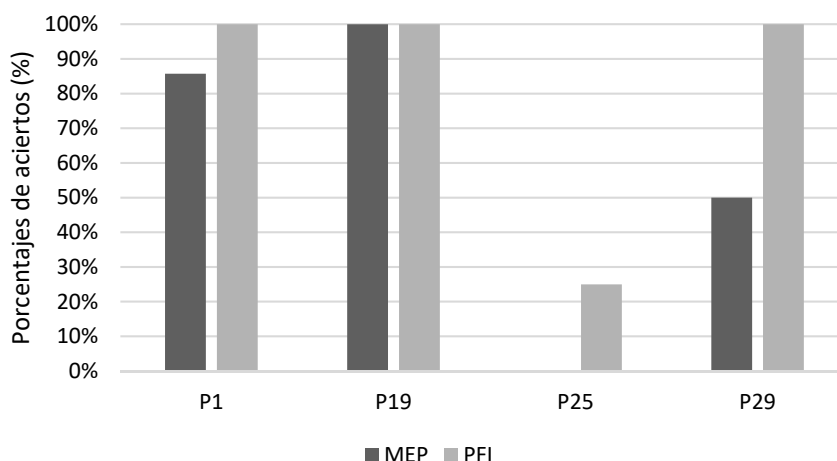


Figura 4: Porcentaje de aciertos en “Elementos químicos y su toxicidad” en el primer intento.

4.2.2. Contaminación y sostenibilidad

Con respecto a la categoría contaminación y sostenibilidad (figura 5) ambos grupos mostraron porcentajes superiores al 75%, en las preguntas relacionadas con acciones humanas que han provocado graves problemas al planeta (P18 y P20), lo que indica una cierta concienciación sobre esta temática. No obstante, los MEP obtuvieron porcentajes bajos (25%) en P2 relacionada con los gases de efecto invernadero frente a PFI (75%), probablemente debido a la formación en cuestiones ambientales obtenida en su grado de Biología o afín.

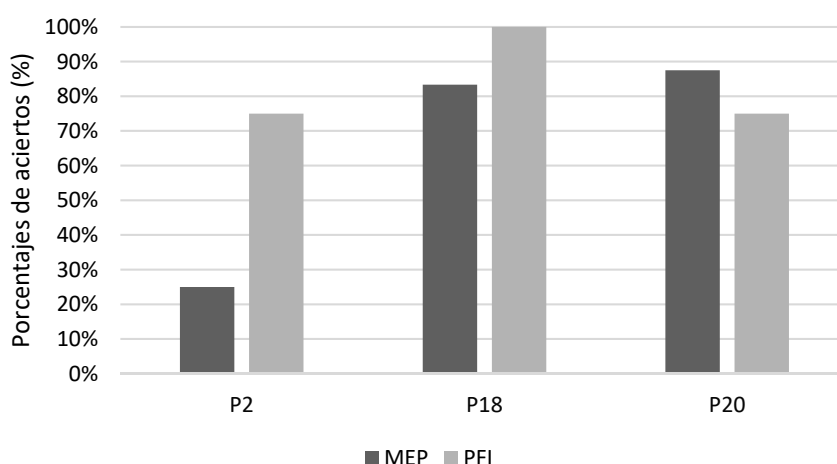


Figura 5: Porcentaje de aciertos en la temática “Contaminación y sostenibilidad” en el primer intento.

4.2.3. Elementos químicos esenciales para la vida

En esta categoría (figura 6), salvo para P14, los PFI obtuvieron mejores resultados que MEP registrándose 100 % de aciertos en el primer intento en la importancia del carbono (P3), nitrógeno

(P5), oxígeno (P8) o hierro (P13 y P17) para la vida, frente a las puntuaciones medias de los MEP. La presencia del magnesio en las plantas (P14) presentó dificultad para ambos grupos.

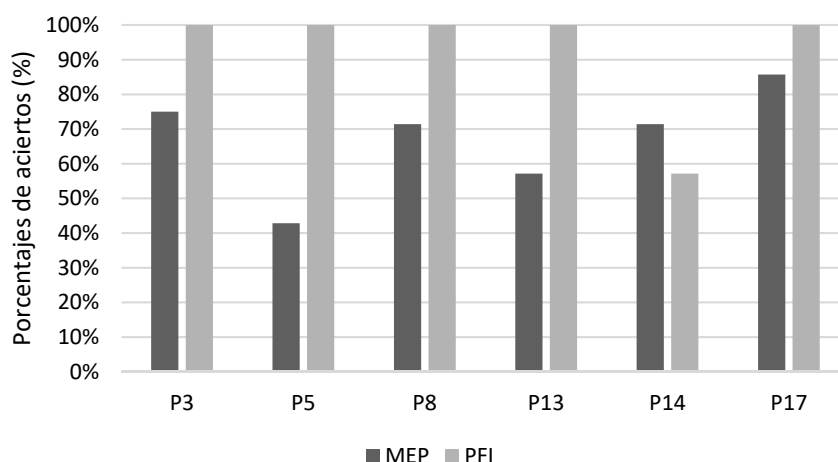


Figura 6: Porcentaje de aciertos en “Elementos químicos esenciales para la vida” en el primer intento.

4.2.4. Energía y procesos nucleares

Los PFI también obtuvieron un desempeño superior en la temática de energía y procesos nucleares. Aunque todas las preguntas no alcanzaron el 100 % de aciertos como en otras temáticas, se consideran satisfactorios en términos de comprensión y rendimiento (figura 7).

Se detecta que los MEP presentaron dificultades importantes en el conocimiento del hidrógeno como fuente de energía limpia en el Sol (P4) con solo el 33,3% de aciertos. También se identificaron ciertas dificultades, aunque en menor medida, en los procesos de fisión (P6) (67%) y fusión nuclear (P7) (78%).

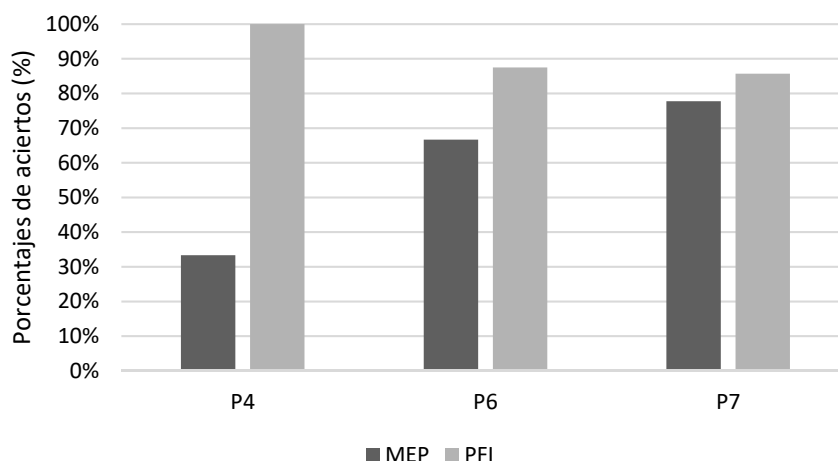


Figura 7: Porcentaje de aciertos en la temática “Energía y procesos nucleares” en el primer intento.

4.2.5. Recursos y consumo responsable

En relación al consumo responsable de agua (P9 y P10), los PFI mostraron un buen nivel de conocimiento. En cambio, los MEP expusieron un conocimiento parcial ya que identificaron adecuadamente algunas medidas como la reparación de fugas en los grifos (P9), con un 100 % de aciertos, pero desconocimiento en otras, como el uso del riego por goteo en jardines (P10), donde no se registraron aciertos en el primer intento (0/7) (figura 8).

Con respecto a la presencia de compuestos de cloro en productos de limpieza (P11) los resultados de los MEP (75%) fueron mejores que los de PFI (50%). Además, las preguntas relacionadas con el uso de aluminio para ventanas (P15) y el encalado de las casas en los pueblos (P16) mantuvieron puntuaciones muy altas en ambos grupos. Esto indica que los participantes tenían conocimientos previos sobre cuestiones relacionadas con el contexto del hogar y su propia cultura.

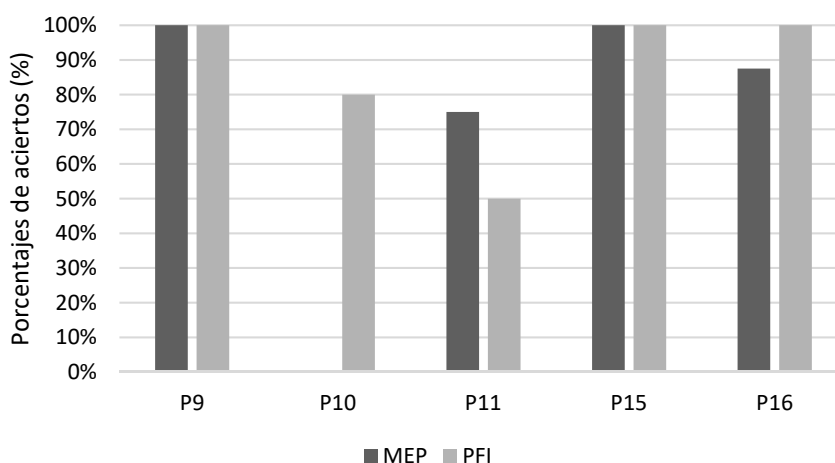


Figura 8: Porcentaje de aciertos en “Recursos y consumo responsable” en el primer intento.

4.2.6. Historia de la química y descubrimientos

Con respecto a esta temática (figura 9) no se observaron grandes diferencias entre los dos grupos, salvo para P21 sobre el origen del elemento francio asociado a Francia donde MEP obtuvo un 80% de aciertos frente a ningún acierto de PFI (0/5). El resto de preguntas sobre elementos descubiertos por investigadores españoles (P22), la presencia del helio en el Sol (P23), o sobre el nombre del wolframio (P24) se situaron en porcentajes medios entre 40% y 80% para ambos grupos.

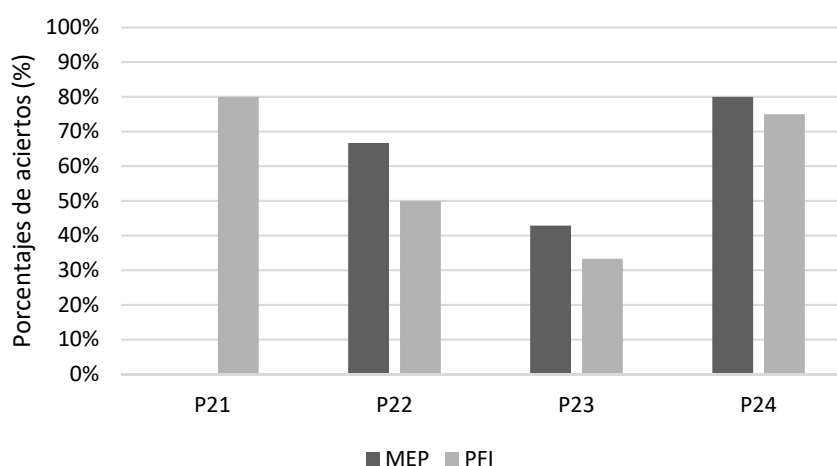


Figura 9: Porcentaje de aciertos en “Historia de la química y descubrimientos” en el primer intento.

4.2.7. Plásticos y gestión de residuos

Los resultados de esta temática (figura 10) mostraron una comprensión previa por parte de ambos grupos (100% de aciertos) de conceptos clave vinculados a la contaminación por materiales plásticos como microplásticos (P27 y P28) o las consecuencias de la presencia de metales pesados en vertidos (P12). No obstante, los MEP parecían mostrar cierto desconocimiento en relación con la biodegradación de plásticos (P26).

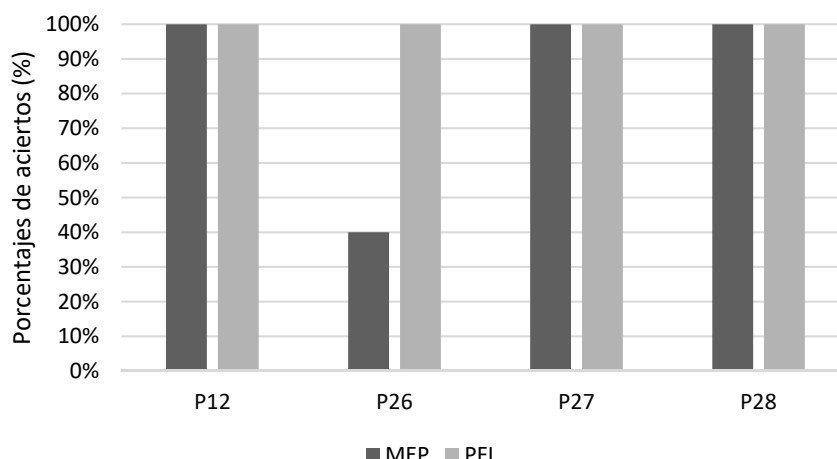


Figura 10: Porcentaje de aciertos en la temática “Plásticos y residuos” en el primer intento.

5. Conclusiones y líneas futuras de investigación

El aumento progresivo de aprendizaje en gran parte de las preguntas indica que *Elemental Home* cumple la función como herramienta educativa para favorecer el aprendizaje en sostenibilidad y educación química, alineándose con otras investigaciones previas sobre juegos serios [10] que destacan por favorecer la construcción de conocimientos científicos en contextos educativos. Esta investigación respalda la idea de que videojuegos serios contribuyen a la mejora de experiencias educativas y a promover una sociedad más comprometida con el medioambiente, mediante el conocimiento de la química de su alrededor.

Como línea de investigación futura, se propone realizar estudios comparativos con grupos de control y con niveles educativos no universitarios, recoger datos cualitativos sobre sus opiniones, hacer comparaciones para mostrar significancias en función del perfil, así como ampliar el tiempo de juego para evaluar su efectividad en diferentes contextos y frente a la enseñanza tradicional. Asimismo, se propone como mejora revisar y ajustar aquellas preguntas que no han generado un impacto significativo en los jugadores, así como organizar los ítems en bloques diferenciados según nivel educativo, con el fin de facilitar análisis más específicos.

Agradecimientos

Esta publicación es parte del proyecto TED2021-130102B-I00 titulado “Transición digital y ecológica en la enseñanza de las ciencias mediante tecnologías disruptivas para la digitalización de juegos educativos y su evaluación con e-rúbricas”, financiado por MCIN/AEI/10.13039/501100011033 y por la Unión Europea “NextGenerationEU”/PRTR. <https://jeted.encic.es/>

Declaración sobre IA Generativa

Los autores no han empleado herramientas de IA Generativa.

Referencias

- [1] M.E. Calvo Centeno, M.P. López Portillo, M.J. Peset González, E. Ropero Moriones. Aprendizaje basado en juegos: incorporación de criterios sostenibles a través de la economía circular. *Tendencias Pedagógicas*. 2020;(36):76–186. <https://doi.org/10.15366/tp2020.36.14>

- [2] S. Mayr, P. Petta. Towards a Serious Game for Trauma Treatment. In: Ma M., Fradinho O., Petersen S., Baalsrud J., eds. Serious games development and applications. Vol. 8101. Springer; 2013. p. 64–69.
- [3] C. Ekin, E. Polat, S. Hopcan. Drawing the big picture of games in education: A topic modeling-based review of past 55 years. *Comput Educ.* 2022;194:104700. <https://doi.org/10.1016/j.compedu.2022.104700>
- [4] A. Filippas, S. Xinogalos. Elementium: Design and pilot evaluation of a serious game for familiarizing players with basic chemistry. *Educ Inf Technol.* 2023;28:11791. <https://doi.org/10.1007/s10639-023-11791-9>
- [5] E. Ntona, G. Arabatzis, G.L. Kyriakopoulos. Energy saving: Views and attitudes of students in secondary education. *Renew Sustain Energy Rev.* 2015;46:1-15. <https://doi.org/10.1016/j.rser.2015.02.033>
- [6] J. Hamari, D.J. Shernoff, E. Rowe, B. Collier, J. Asbell-Clarke, T. Edwards. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Comput Human Behav.* 2015;54:170–179. <https://doi.org/10.1016/j.chb.2015.07.045>
- [7] K. Ottander, S. Simon. Learning democratic participation? Meaning-making in discussion of socioscientific issues in science education. *Int J Sci Educ.* 2021;43(12):1895–1925. <https://doi.org/10.1080/09500693.2021.1946200>
- [8] A.J. Franco-Mariscal. Critical thinking in science education and teacher training. Springer; 2024. <https://doi.org/10.1007/978-3-031-78578-8>
- [9] P. Caserman, K. Hoffmann, P. Müller, M. Schaub, K. Straßburg, J. Wiemeyer, R. Bruder, S. Göbel. Quality criteria for serious games: Serious part, game part, and balance. *JMIR Serious Games.* 2020;8(3):1–14. <https://doi.org/10.2196/19037>
- [10] A. Lutfi, F. Aftinia, B.E. Permani. Gamification: Game as a medium for learning chemistry to motivate and increase retention of students' learning outcomes. *J Technol Sci Educ.* 2023;13(1):193–207. <https://doi.org/10.3926/jotse.1842>
- [11] B. Roy, S. Gasca, J. Winum. Chem'Sc@pe: an Organic Chemistry Learning Digital Escape Game. *J Chem Educ.* 2023;100(3):1382–1391. <https://doi.org/10.1021/acs.jchemed.2c01105>
- [12] M. Guerra-Mota, D. Minas, M. Xenos, M.M. Sa. Development of a 3D virtual world tool for sustainable energy education. *arXiv.* 2024. <https://doi.org/10.48550/arxiv.2410.10586>

Optimización de un agente autónomo para combates de Pokémon

Rocío Barragán Moreno¹, Antonio M. Mora¹ and Pablo García-Sánchez²

¹ Depto. Teoría de la Señal, Telemática y Comunicaciones, ETSIT-CITIC, Universidad de Granada, España.

² Depto. Ingeniería de Computadores, Automática y Robótica, ETSIT-CITIC, Universidad de Granada, España.

Resumen

Este trabajo propone, el diseño e implementación de un agente autónomo efectivo enfocado en realizar combates de Pokémon contra equipos generados aleatoriamente. Dicho agente se optimizará posteriormente por medio de técnicas de Inteligencia Artificial y Machine Learning, específicamente, a través de Algoritmos Genéticos. Una vez hecho esto, se realizarán diversas comparativas para analizar las mejoras obtenidas por medio de la optimización.

El agente y los experimentos se desarrollarán sobre el entorno llamado VGC AI Competition, el cual nos ofrece un framework llamado VGC AI, que dispondrá de las herramientas necesarias para poder implementarlo y probarlo. Además, dicho entorno también ofrece distintos agentes ya implementados para poder extraer nuestros resultados, ya sean competidores de años anteriores, como agentes simples proporcionados por el propio entorno como base comparativa.

Palabras Clave

Agente Autónomo, Inteligencia Artificial, Pokémon, Combates, Optimización, Algoritmo Evolutivo, Algoritmo Genético

1. Introducción

La franquicia Pokémon [1], especialmente en su vertiente de combates competitivos, es una de las más exitosas del mundo. El campeonato mundial celebrado en Japón en el año 2023, contó con la participación de 1600 competidores altamente ilusionados y dispuestos a “pelear” entre ellos para demostrar sus habilidades en el marco competitivo de Pokémon [2].

Debido a su gran repercusión, es evidente que se debe dar importancia a la optimización de agentes inteligentes en el contexto de Pokémon. Aunque dicho entorno, en una primera instancia, pueda parecer simple y monótono, la lucha de Pokémon es un ámbito desafiante para las técnicas de aprendizaje por refuerzo y *Machine Learning*, debido a su alta complejidad en la toma de decisiones, dinamismo, y estocasticidad. El agente de Pokémon deberá enfrentarse a una serie de variables en el entorno y a decisiones que tiene que tomar en función de lo que esté ocurriendo en el momento actual. Además, deberá tener en cuenta una serie de características que influirán en las decisiones y en el rendimiento de los Pokémon, incluidas las decisiones del oponente [3]. Se trata de un problema en el que se pueden llegar a tener más de 1000 pokémon de 18 tipos diferentes (con distintas ventajas y debilidades frente a otros). Los Pokémon, además, pueden aprender distintos movimientos de ataque y defensa según dicho tipo.

Teniendo en cuenta lo gigantesco de este espacio de posibles soluciones, el objetivo de este trabajo es diseñar y probar un agente autónomo capaz de jugar al juego de manera eficiente y efectiva y para ello, nos basaremos en la creación de un motor de comportamiento que modele las decisiones que tomaría un jugador experto durante una partida. Dicho conjunto de reglas será posteriormente optimizado por medio de un Algoritmo Genético [4], una metaheurística inspirada en la evolución de las especies en la naturaleza. De manera que ésta modela soluciones a un problema como individuos (vectores de valores), los cuales deben ser evaluados para asignarles un *fitness* y, posteriormente, los

mejores son seleccionados, recombinados (cruce) para generar descendientes (nuevas soluciones con lo mejor de sus 'padres') y mutados para introducir variabilidad en la búsqueda de los mejores individuos, es decir, para explorar el espacio de soluciones posibles.

2. Pokémon: el juego y su complejidad

Pokémon es un juego de estrategia por turnos (RPG) [1], en el que cada jugador compone un equipo de hasta 6 pokémon con sus propias características (nivel, estadísticas, movimientos) que se van enfrentando en combates 1 vs 1 (ver Figura 1).

Mecánicas Principales:

- *Tipos*: Cada Pokémon tiene uno o dos tipos elementales (Fuego, Agua, Planta, etc.). Las interacciones entre tipos determinan la efectividad del daño (ej. Agua es eficaz contra Fuego, multiplicando el daño por x2).
- *Estadísticas (Stats)*: Definen el rendimiento de la criatura: PS (Puntos de Salud), Ataque, Defensa, Ataque Especial, Defensa Especial y Velocidad.
- *Movimientos*: Cada Pokémon conoce hasta 4 movimientos. Estos pueden ser físicos, especiales o de estado (que no hacen daño directo, pero alteran estadísticas o provocan condiciones como "Dormido" o "Paralizado").
- *Cambios*: El jugador puede optar por atacar o cambiar a su Pokémon activo por uno de la reserva, consumiendo el turno, pero preservando al Pokémon herido o buscando una ventaja de tipo.



Figura 1: Imagen del juego Pokémon Negro para la consola Nintendo DS (GameFreak/Nintendo).

La complejidad de los combates radica en la predicción. Un turno se resuelve en dos fases: selección de acciones (simultánea a los dos jugadores y oculta) y ejecución (basada en la velocidad de los Pokémon que luchan). El agente debe predecir qué hará el rival para seleccionar la acción óptima, considerando probabilidades de golpe crítico, precisión de movimientos y efectos secundarios. Y esto lo debe hacer considerando el gran número de movimientos posibles por cada pokémon y tipo, así como desconociendo las estadísticas del rival.

3. Algoritmos Genéticos

Los Algoritmos Genéticos (AG) [4] son uno de los tipos de la familia de Algoritmos Evolutivos. Se trata de métodos de búsqueda y optimización inspirados en la teoría de la evolución natural de Darwin. Funcionan manteniendo una población de soluciones candidatas (individuos) que evolucionan a lo largo de generaciones. Los procesos clave incluyen:

- **Selección:** Elegir los mejores individuos basándose en una función de aptitud (*fitness*).
- **Cruce:** Combinar información de dos o más individuos (padres) para generar descendencia.

- **Mutación:** Introducir cambios aleatorios para mantener la diversidad genética y evitar óptimos locales.

El objetivo general es ir consiguiendo una población de individuos cada vez mejores, en relación a la función objetivo que se pretende optimizar.

Los AGs han sido aplicados a infinidad de problemas de optimización, por su versatilidad y eficiencia para encontrar soluciones a problemas de gran complejidad. En el contexto de los videojuegos, los AG se han utilizado para generar contenido procedimental [5], ajustar la dificultad [6] o, como en este caso, optimizar el comportamiento de personajes no jugadores (NPCs), llamados agentes o bots [7,8].

4. Estado del arte

Dentro del entorno de investigación que propone el juego de Pokémon, se han hecho propuestas que se centraban en la parte del juego RPG y en la creación de un agente para superarlo [9], basado en *Deep Reinforcement Learning* (DRL). Dicho agente fue muy exitoso en su cometido, de forma que tuvo bastante repercusión incluso a nivel de Youtube, alcanzando más de 8M de visualizaciones los vídeos de sus progresos en el juego.

Sin duda el aspecto más explotado, desde el punto de vista investigador, ha sido la de los combates entre equipos de Pokémon. En dicho ámbito, se han abordado diversas líneas, como, por ejemplo, trabajos para la optimización de la creación de equipos lo más compensados/equilibrados posible [10].

Del mismo modo, se han abordado enfoques centrados en la evaluación de combates y la predicción de sus resultados para determinar la mejor estrategia a priori [11].

Dentro de ese ámbito, la creación de agentes competitivos en combates ha sido la más fructífera, proponiéndose, aparte de infinidad de agentes basados en heurísticas o reglas simples, agentes que aplican técnicas de Inteligencia Computacional clásicas, como *MonteCarlo Tree Search* (MCTS) [12], hasta más avanzadas como DRL [13] o, más recientemente, usando *Large Language Models* (LLMs) [14,15].

Sin embargo, hay pocos trabajos que exploten las ventajas del uso de los AGs, como se ha explicado en el apartado anterior, para la mejora y adaptación de este tipo de agentes. Este será, precisamente, el objetivo del presente trabajo, que se apoyará en un conjunto de reglas creadas por una jugadora experta de competiciones/combates de Pokémon, las cuales se “refinarán” por medio de AGs, optimizando diversos parámetros de los que dependen.

5. VGC AI Framework

El *VGC AI Framework* propuesto por Reis et al. [16] es el estándar académico actual para el desarrollo y prueba de agentes de Pokémon, proporcionando una API en Python para desarrollar agentes que compiten bajo las reglas oficiales. Esta competición se encuadra dentro del *VGC AI Competition* [17, 18].

Este entorno simula la lógica del juego Pokémon (daño, efectos de estado, cambios de características) y provee una interfaz para programar agentes en Python. Las características clave del entorno incluyen:

- **Combates 1v1 y 2v2:** Aunque el estándar VGC es combates dobles, el *framework* permite configuraciones flexibles.
- **Generación de Equipos:** Capacidad de generar equipos aleatorios (Random Teams), crucial para evaluar la generalización del agente desarrollado y evitar el sobreajuste (*overfitting*) a un equipo específico.
- **Espacio de Acciones:** En cada turno, el agente debe elegir entre los 4 movimientos de sus Pokémon activos o realizar un cambio (*switch*) por un Pokémon de la reserva.

La Figura 2 muestra un ejemplo de combate celebrado en el entorno VGC AI Framework.

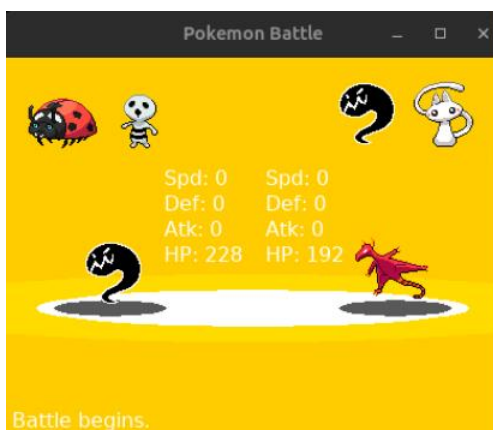


Figura 2: Ejemplo de combate en VGC AI Framework.

En este trabajo, se seguirán las reglas descritas en la Competición Internacional de Pokémon, incluida dentro del congreso IEEE COG 2025 (<https://cog2025.inesc-id.pt/vgc-ai-competition/>). En ella:

- Se crean equipos aleatorios (equilibrados) de 6 pokémon.
- Se eligen 3 pokémon (del equipo) para combatir, contando cada uno con 4 movimientos.
- Se realizan 3 combates con cada equipo y se intercambian los equipos entre los jugadores, repitiendo el proceso. Serían 6 combates en total.

Arquitectura de VGC AI Framework:

Este framework separa la lógica del juego (*engine*) de la lógica de los agentes.

- **Game State:** Contiene la información completa de la batalla (pública y privada).
- **Agentes:** Clases que heredan de una interfaz base y deben implementar el método `get_action(state)`.
- **Interfaz:** El agente recibe una observación del estado (simplificada o completa dependiendo de la configuración) y debe devolver una acción válida (usar movimiento 1-4 o cambiar de Pokémon).

Dicho entorno permite la ejecución de enfrentamientos en tiempo máquina (sin renderizado gráfico), para el entrenamiento de algoritmos de *Machine Learning*, lo cual es crucial para ejecutar el gran número de simulaciones necesarias para realizar la evolución en un AG de manera exitosa.

Adaptación Realizada:

Con el fin de poder integrar un AG, como se ha indicado anteriormente, el *framework* original requería adaptaciones para soportar el ciclo de vida de ese tipo de algoritmos. Para ello, se desarrollaron módulos específicos para:

- **Automatización de Batallas:** Creación de scripts que permiten la ejecución secuencial y paralela de miles de combates sin intervención humana.
- **Interfaz de Genes:** Modificación de la clase del agente para aceptar un vector de "genes" externo en tiempo de ejecución, permitiendo que un mismo código base se comporte de manera diferente según los parámetros recibidos.
- **Extracción de Métricas:** Implementación de sistemas de *logging* para capturar no solo el resultado (victoria/derrota), sino métricas granulares como el daño infligido, turnos sobrevividos y PS (Puntos de Salud) restantes, esenciales para calcular la función de *fitness*.

6. Diseño del Agente

Se han implementado dos tipologías de agentes para este estudio:

Agente Básico (*Benchmark*)

Para establecer una línea base, se desarrolló un agente que opera bajo reglas heurísticas simples o aleatorias. Este agente sirve como "sparring" inicial y permite verificar la integridad del entorno de simulación.

Agente Parametrizado (*AgentToOptimize*)

El núcleo de la investigación es el agente derivado del original *OriginalAgent*. A diferencia de un agente estático, este agente toma decisiones evaluando el estado del tablero mediante una función de utilidad ponderada.

Los "pesos" de esta función de utilidad corresponden al genotipo del individuo en el Algoritmo Genético (ver Sección 7). Algunos ejemplos de parámetros que el agente considera incluyen:

- **Agresividad:** Ponderación otorgada a movimientos que causan daño directo.
- **Conservadurismo:** Importancia de mantener los PS (puntos de salud) propios altos, priorizando curación o protección.
- **Ventaja de Tipos:** Peso otorgado a la tabla de tipos (ej. usar agua contra fuego).
- **Predicción:** Estimación del riesgo de que el rival cambie de Pokémon.

El *AgentToOptimize* recibe estos valores (genes) al inicio del combate y los utiliza para puntuar todas las acciones posibles, seleccionando aquella con la puntuación más alta. Para ello, el *OriginalAgent* (y el *AgentToOptimize*) implementa un **conjunto de reglas**, basadas en conocimiento experto, para elegir el mejor movimiento en cada turno para el Pokémon actual:

- Clima (afecta a algunos tipos).
- Tipo del Pokémon propio y el del rival.
- Velocidad del Pokémon propio y el rival.
- Potencia de los movimientos de ataque.
- Nivel de vida del Pokémon propio y el rival.
- Número de movimientos restantes (uso limitado).
- Peso para cambio de Pokémon.
- Peso para cambio de estrategia (ofensiva/defensiva).
- Elección del mejor movimiento, según todas las características.

7. Optimización del Agente

Para hacer la optimización, se han desarrollado dos esquemas de AG:

- **Elitista:** mantiene las mejores soluciones y tiene una alta presión selectiva. Este enfoque tiende a converger más rápido y hacer menos exploración del espacio de búsqueda. La descripción del proceso se puede ver en la Figura 3 (izquierda).
- **No elitista:** se hacen una exploración más amplia del espacio de soluciones, lo que implica una convergencia más lenta, pero, por el contrario, las mejores soluciones se podrían perder entre una generación y la siguiente. Se puede ver descrito en la Figura 3 (derecha).

Representación de individuos (Genotipo):

En primer lugar, cabe destacar que la representación de los individuos consistirá en una serie de parámetros a mejorar, cuyo valor oscilará en valores reales entre 0 y 1. Se ha decidido así puesto que la representación de números reales es una de las mejores maneras de representar al Genotipo de los AGs. Por un lado, tendremos parámetros enfocados a las características propias de los Pokémon que

actúan en el combate y, por otro lado, tendremos parámetros que determinarán la importancia de una serie de condiciones que se evaluarán en cada turno del juego para ver si se cumplen, y en función de si se cumplen o no, se le asignarán una serie de pesos relativos en función de estos parámetros. Éstos vienen representados en la Tabla 1.

Tabla 1. Genotipo de los cromosomas del AG.

HP mi Pokémon bajo	HP oponente bajo	HP oponente alto	máximo uso movimientos	numero máximo switches	Peso C1	Peso C2	Peso C3	Peso C4	Peso C5
0 - 0.4	0 - 0.4	0.5 - 1	0 - 0.5	0 - 1	0 - 1	0 - 1	0 - 1	0 - 1	0 - 1

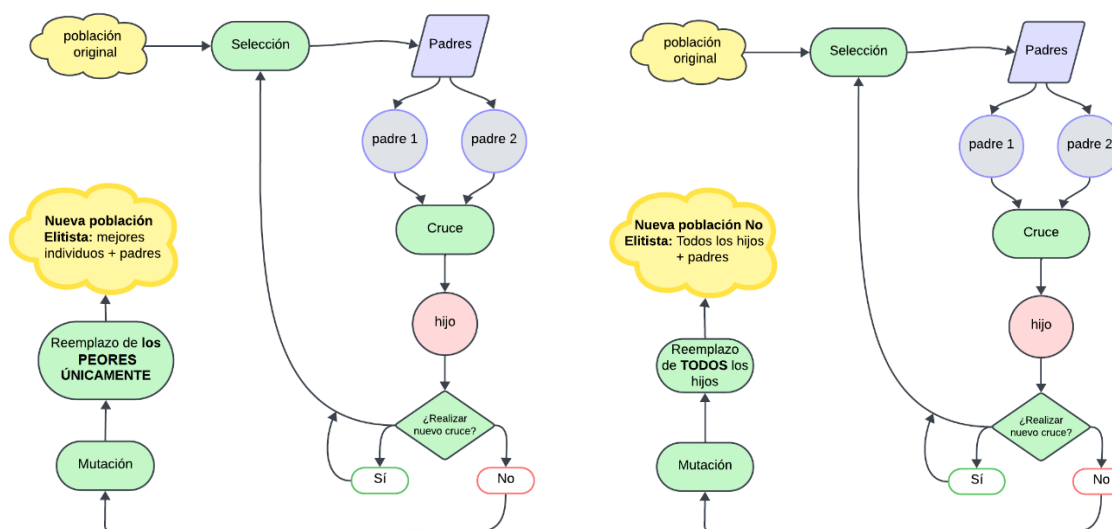


Figura 3: Proceso del AG Elitista (izquierda) y No Elitista (derecha).

Los genes de cada cromosoma serán, por tanto:

- **HP bajo de mi Pokémon.** Este parámetro determinará a partir de qué porcentaje se considera bajo el HP de mi Pokémon activo.
- **HP bajo de Pokémon oponente.** Este parámetro determinará a partir de qué porcentaje se considera bajo el HP de mi oponente.
- **HP alto de Pokémon oponente.** Este parámetro determinará a partir de qué porcentaje se considera alto el HP de mi oponente.
- **PP máximo restante.** Este parámetro determinará a partir de qué porcentaje restante de los PP máximos de un movimiento, es conveniente seguir usando dicho movimiento o no. Este valor será muy útil a la hora de decidir qué movimiento se lanzará.
- **Número máximo de switches.** Este parámetro determinará el número máximo de switches que un pokémon puede hacer en cada partida de la batalla. De esta manera, no se generarán bucles infinitos de switches, lo que es muy común que ocurra en combates de pokémon.
- **Peso relativo a cada condición (C1 a C5).** Este parámetro determinará el peso de cada condición a evaluar en la lógica de comportamiento del agente, en concreto:
 - C1: comprueba si, en el momento actual, mi pokémon activo tiene bajo HP.
 - C2: que comprueba si, en el momento actual, el pokémon activo oponente tiene bajo HP.
 - C3: comprueba si, dado un pokémon, este es efectivo contra el otro pokémon dado.
 - C4: que comprueba si, en el momento actual, el pokémon oponente tiene alto HP.
 - C5: que comprueba si, en el momento actual, el pokémon activo del oponente tiene aplicado algún estado.

Operadores Genéticos [19]:

- **Selección:** Se emplean métodos de *selección por torneo* para elegir a los progenitores, favoreciendo a los individuos con mayor *fitness* pero manteniendo cierta presión selectiva controlada.
- **Cruce:** Se utiliza *cruce aritmético* para combinar linealmente los vectores de pesos de dos padres.
- **Mutación:** Se ha considerado la *mutación uniforme aleatoria*, con una tasa de variación del valor del gen elegido para mutar del 5%.
- **Reemplazo:** Se ha considerado una estrategia elitista (reemplazar sólo los peores individuos) y no elitista (se reemplazan los individuos no seleccionados como padres por todos los hijos, aunque sean peores).

La **función de fitness** considerada tiene tres factores y sigue esta expresión (para una serie de combates):

$$Fitness = w_v \cdot victorias + w_t \cdot 1/turnos + w_{hp} \cdot hp \quad (1)$$

Donde, *victorias* se refiere a la tasa de victorias obtenido respecto al total de combates realizados (20 por defecto), *turnos* son los turnos medios empleados para terminar los combates, *hp* es la media de la vida de nuestros pokémon al terminar los combates. Cada factor es normalizado (entre 0 y 1) y está ponderado por un peso (w_v, w_t, w_{hp}) que permite asignar su importancia relativa.

8. Experimentos y resultados

Los experimentos realizados consistirán en lanzar los 2 algoritmos genéticos implementados, el elitista y el no elitista, con las siguientes características:

- Número de generaciones: 40.
- Número de individuos de la población: 20.
- Probabilidad de cruce: 100 %.
- Probabilidad de mutación: 10 %.

Agentes rivales

Con los que enfrentaremos nuestro agente durante la optimización:

- *OriginalAgent*: modelo inicial, basado en reglas y condiciones definidas por una jugadora experta, y que depende de umbrales, pesos y probabilidades fijas.
- *OneTurnLookAhead*: heurística *greedy* fuerte. Elige el mejor movimiento basado en criterios de potencia de ataque.
- *TypeSelector*: aplica una heurística *greedy*, pero tiene en cuenta cuestiones de tipos de los pokémon y efectividad.
- *WBukowskiBattlePolicy*. agente ganador de la Pokémon VGC AI Competition del año 2024.

Por cada uno de los dos AGs (Elitista y No Elitista), se realizarán 10 ejecuciones de cada uno de los siguientes experimentos:

- AgentToOptimize VS OneTurnLookAhead y TypeSelector: Se realizarán 2 combates contra cada uno para el cálculo del fitness, es decir, 4 combates en total, que determinarán el fitness medio de esos 4 combates.
- AgentToOptimize VS OriginalAgent: Se realizarán 2 combates contra él para el cálculo del fitness.

Convergencia del Fitness:

Los resultados de estos experimentos se pueden ver en las Figuras 4, 5, 6 y 7. Las cuales muestran las gráficas de evolución del *fitness* (promedio y máximo) durante las evoluciones de cada una de las ejecuciones. Todas muestran una tendencia ascendente, cosa deseable, dado que el *fitness* se pretende maximizar. Si bien, se producen estabilizaciones cada cierto número de generaciones. El uso de elitismo evita oscilaciones o caídas drásticas en el rendimiento máximo entre generaciones.

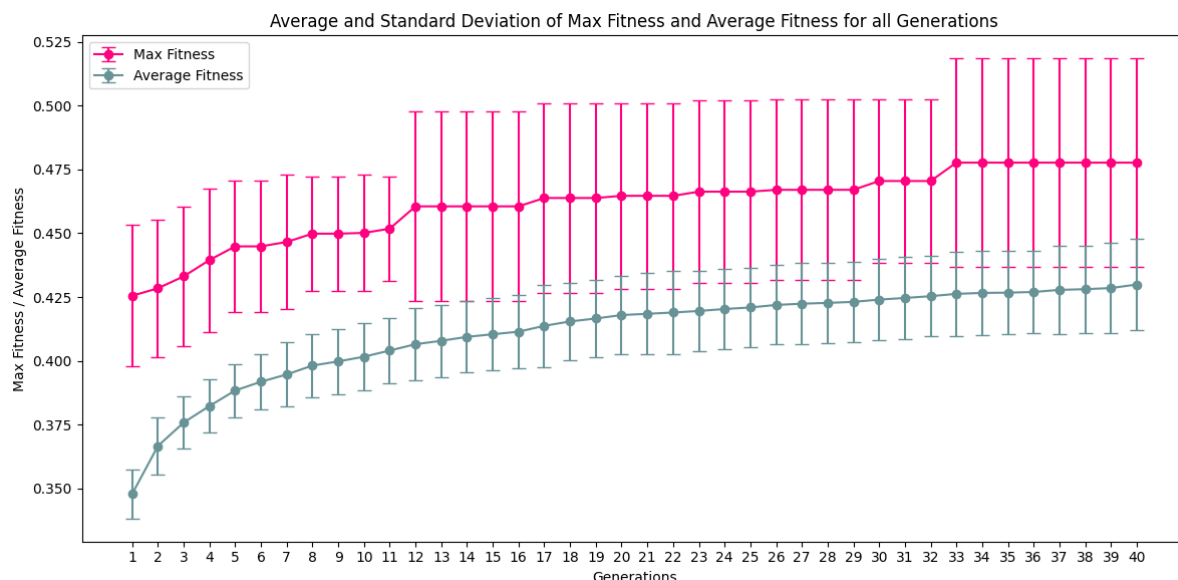


Figura 4: AG Elitista. Evolución del mejor *fitness* y del *fitness* medio del AgentToOptimize VS OneTurnLookAhead y TypeSelector. Media de 10 ejecuciones y desviaciones típicas (bigotes).

En la Figura 4, se puede apreciar que, dado que es un AG Elitista, se van preservando las mejores soluciones, por lo que tanto el *fitness* medio como el Máximo de cada generación, será siempre igual o creciente. Cabe destacar que los valores no oscilan mucho, y que el *fitness* máximo llega a 0,48 sobre 1, aproximadamente, lo que significa que estos rivales son bastante competitivos.

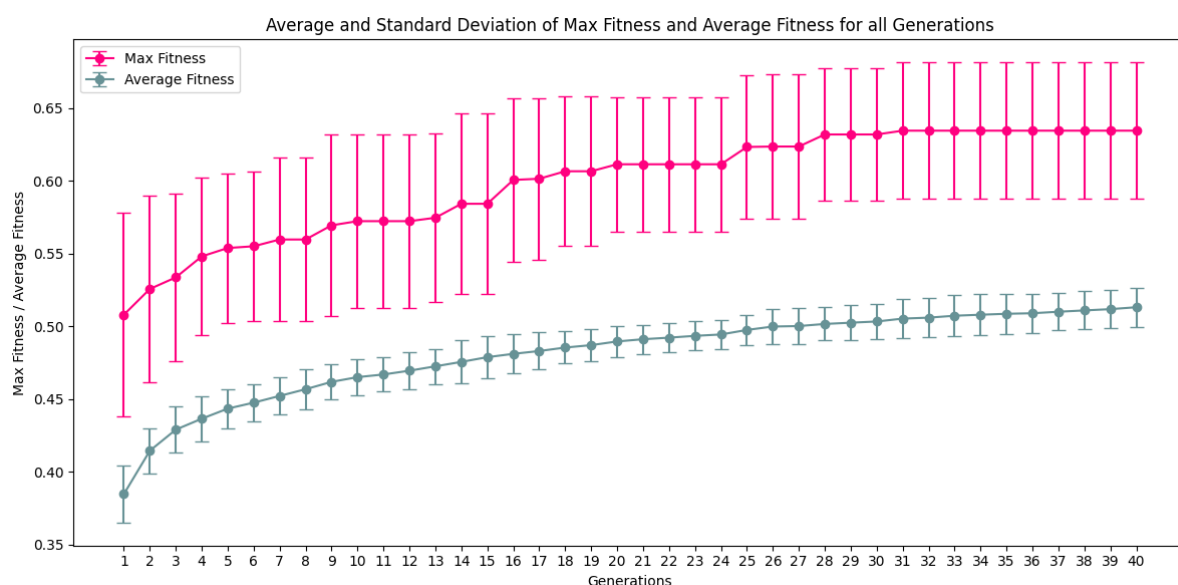


Figura 5: AG Elitista. Evolución del mejor *fitness* y del *fitness* medio del AgentToOptimize VS OriginalAgent. Media de 10 ejecuciones y desviaciones típicas (bigotes).

Nuevamente, en la Figura 5, se puede apreciar la conservación de los mejores individuos, dado que también es elitista, y además podemos ver que sí hay un fitness máximo más significativo que el anterior (cercano a 0,6). Esto refleja una mejora relevante en los combates contra el Agente Original. Además, se puede notar que la desviación típica en los fitness medios es bastante menor que en el experimento anterior, lo que indica una mayor robustez en las soluciones, debida, seguramente, a la consideración de un único rival, que permitirá una mayor especialización de los agentes optimizados para combatir contra ese rival en concreto.

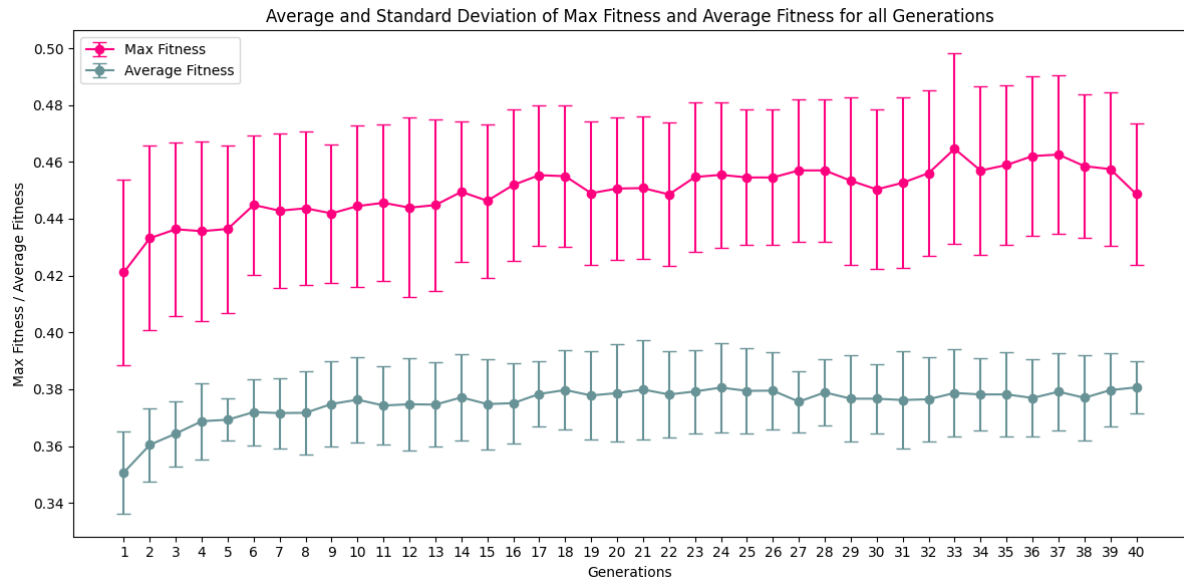


Figura 6: AG No Elitista. Evolución del mejor fitness y del fitness medio del AgentToOptimize VS OneTurnLookAhead y TypeSelector. Media de 10 ejecuciones.

La Figura 6 refleja los efectos de experimentar con un AG no elitista, en el que no se preservan los mejores individuos y, por tanto, la tasa de convergencia es menor y la variabilidad en la calidad de las soluciones es más patente. También se puede apreciar que el fitness máximo aquí está alrededor del 0,46; un resultado un poco más bajo que en el AG Elitista contra estos mismos rivales, lo que determina que son enemigos fuertes.

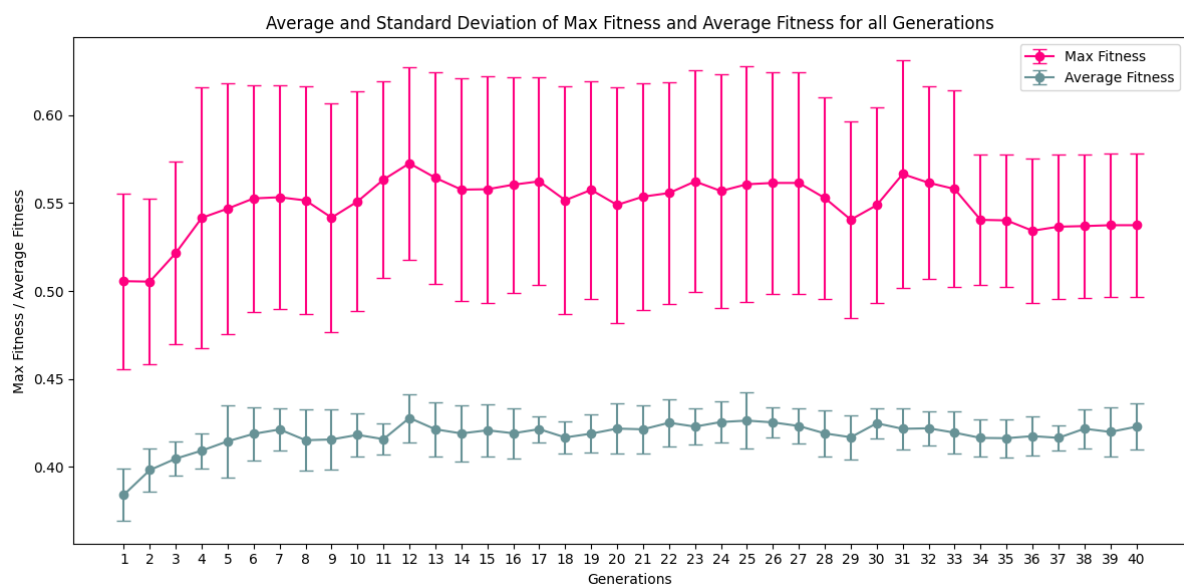


Figura 7: AG Elitista. Evolución del mejor fitness y del fitness medio del AgentToOptimize VS OriginalAgent. Media de 10 ejecuciones.

Por último, en la Figura 7 se puede apreciar la alta variabilidad en los fitness obtenidos como medias de las generaciones, pero también se refleja que el fitness más alto se suele situar en torno a 0,57, aproximadamente, lo que nos vuelve a confirmar que el Agente Original es un rival menos fuerte que los otros y con el que podemos conseguir mejores resultados en las comparativas.

En general, de esta primera batería de pruebas, podemos concluir que, para este problema, el uso de un AG Elitista nos lleva a obtener soluciones/agentes más competitivos.

Tras los experimentos de evolución, se ha hecho un estudio para determinar la bondad del agente sin optimizar (*OriginalAgent*). Para ello, se ha realizado una batalla contra cada rival, con 30 partidas y 20 combates cada una, por lo que se obtendrá un máximo de 600 victorias por agente. Los resultados se pueden ver en la Tabla 2.

Tabla 2. Porcentaje de victorias obtenidas por el agente original (basado en reglas y condiciones con parámetros fijos) frente a los demás agentes considerados.

VS	<i>OneTurnLookAhead</i>	<i>TypeSelector</i>	<i>Wbukowski</i>
<i>OriginalAgent</i>	32.5%	57.5%	53.17%

Después, se eligieron los mejores agentes evolutivos de cada grupo de ejecuciones en base a su fitness. De modo que se obtuvieron los agentes optimizados:

- **Elitist-AgentOptimized-OneTurnTypeSel:** Agente entrenado contra *OneTurnAhead* y *TypeSelector* utilizando el AG Elitista.
- **Elitist-AgentOptimized-OriginalAgent:** Agente entrenado contra *OriginalAgent* utilizando el AG Elitista.
- **Non-elitist-AgentOptimized-OneTurnTypeSel:** Agente entrenado contra *OneTurnAhead* y *TypeSelector* utilizando el AG no Elitista.
- **Non-Elitist-AgentOptimized-OriginalAgent:** Agente entrenado contra *OriginalAgent* y utilizando el AG no Elitista.

Para concluir con los experimentos, a fin de decidir cuál sería el mejor agente, se realizaron batallas de todos contra todos, en las que cada batalla serán nuevamente 30 partidas y cada partida constará de 20 combates. Los resultados de estos enfrentamientos se pueden ver en la Tabla 3.

Tabla 3. Porcentaje de victorias obtenidas por cada agente frente a los demás.

VS	<i>OneTurnLookAhead</i>	<i>TypeSelector</i>	<i>OriginalAgent</i>	<i>Wbukowski</i>
<i>Elitismo OneTurnTypeSel</i>	42.17% victorias	62.83% victorias	62.5% victorias	62.17%
<i>Elitismo OriginalAgent</i>	39% victorias	61% victorias	54% victorias	61.83%
<i>No Elitismo OneTurnTypeSel</i>	35.33% victorias	62% victorias	60.67% victorias	58.17%
<i>No Elitismo Original Agent</i>	33.67% victorias	60.5% victorias	55.83% victorias	60.67%

Como se aprecia en la Tabla, todos los agentes evolucionados mejoran los porcentajes de victorias frente a los demás agentes que conseguía el Agente Original (mostrados en la Tabla 2). Además, todos

ellos consiguen vencer en un porcentaje mayor del 50% al Agente Original, por lo que se entiende que la optimización ha sido efectiva.

Si miramos los resultados frente al agente WBukowskiBattlePolicy, comprobamos que todos los agentes son capaces de ganarle en la mayor parte de los enfrentamientos, lo que es un resultado destacable, considerando que dicho agente fue el ganador de la competición de 2024.

9. Conclusiones y Trabajo Futuro

Este trabajo presenta una propuesta de agentes para jugar al famoso RPG Pokémon, en su variante de juego de combates. Para ello, se ha diseñado e implementado un agente capaz de participar en la competición internacional (siguiendo las reglas de la misma) VGC AI Competition, usando su *framework*. Este entorno ha sido ampliado y adaptado para la inclusión de Algoritmos Evolutivos, que permitiesen optimizar agentes diseñados para combatir.

De esta forma, el agente creado en función de reglas definidas por una jugadora experta para modelar su comportamiento, ha sido optimizado por medio de un Algoritmo Genético, en dos variantes (elitista y no elitista) y considerando distintos rivales para calcular la función de *fitness*.

Tanto el agente original, como el optimizado se han validado frente a otra serie de agentes, incluyendo el campeón de la competición de 2024.

Los resultados obtenidos demuestran que la optimización ha sido efectiva y que los agentes han sido mejorados en todas las variantes del AG aplicado, destacando la versión elitista. Además, estos agentes han sido capaces de competir de forma efectiva contra el campeón, obteniendo porcentajes de victorias en torno al 60%.

Como trabajo futuro se planteará implementar versiones de los AGs más adaptadas al problema, así como considerar oponentes con mayor variedad para conseguir agentes más generalistas. También se pueden ampliar las reglas de comportamiento y los parámetros a optimizar, para conseguir evaluar mayor cantidad de situaciones de combate. Del mismo modo, se podrían combinar estas reglas con métodos de exploración en árboles (Minimax o MCTS) para considerar estados futuros, no solo el actual.

10. Agradecimientos

Este trabajo ha sido financiado, en parte, por el Ministerio de Ciencia, Innovación y Universidades y la Agencia Estatal de Investigación (MICIU/AEI/10.13039/501100011033) a través del proyecto con código PID2023-147409NB-C21, que también cuenta con fondos FEDER, EU.

También ha sido financiado por los proyectos PID2023-151462OB-I00 y PID2020-115570GB-C22, también del MICIU/AEI. Junto a estos, se ha contado con financiación de los proyectos TED2021-131699B-I00 y TED2021-129938B-I00 del MICIU y la Unión Europea NextGenerationEU/PRTR. Además del proyecto C-ING-179-UGR23 concedido por la "Consejería de Universidades, Investigación e Innovación" (Junta de Andalucía, Programa FEDER 2021-2027).

11. Bibliografía

- [1] Pokémon (webpage): <https://www.pokemon.com/es>, 2025.
- [2] Mundial de Pokémon Competitivo Japón, 2023. <https://www.elpais.com.co/mundo/mundial-de-pokemon-en-japon-2023-inicia-con-1600-jugadores-estos-son-los-detalles-1108.html>.
- [3] Jett Wang. Winning at Pokémon Random Battles Using Reinforcement Learning. <https://dspace.mit.edu/handle/1721.1/153888>, 2024.
- [4] D. E. Goldberg, "Genetic Algorithms in search, optimization and machine learning", Addison Wesley, 1989.
- [5] Lara-Cabrera, Raúl, Carlos Cotta, and Antonio J. Fernández-Leiva. "On balance and dynamism in procedural content generation with self-adaptive evolutionary algorithms." *Natural Computing* 13.2 (2014): 157-168.

- [6] P. García-Sánchez, A.P. Tonda, A.M. Mora, G. Squillero, J.J. Merelo, “Automated playtesting in collectible card games using evolutionary algorithms: A case study in Hearthstone”. *Knowl. Based Syst.* 153: 133-146 (2018).
- [7] P. García-Sánchez, A.P. Tonda, A.J. Fernández-Leiva, C. Cotta, “Optimizing Hearthstone agents using an evolutionary algorithm”. *Knowl. Based Syst.* 188 (2020).
- [8] Schrum, Jacob, and Risto Miikkulainen. "Constructing complex NPC behavior via multi-objective neuroevolution." *Proceedings of the AAAI conference on Artificial intelligence and interactive digital entertainment*. Vol. 4. No. 1. 2008.
- [9] Pleines, Marco, et al. "Pokémon red via reinforcement learning." *2025 IEEE Conference on Games (CoG)*. IEEE, 2025.
- [10] Reis, Simão, et al. "An Adversarial Approach for Automated Pokémon Team Building and Metagame Balance." *IEEE Transactions on Games* 16.2 (2023): 365-375.
- [11] Wang, Jett. *Winning at pokémon random battles using reinforcement learning*. Diss. Massachusetts Institute of Technology, 2024.
- [12] Ihara, Hiroyuki, et al. "Implementation and evaluation of information set Monte Carlo Tree Search for Pokémon." *2018 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, 2018.
- [13] Simoes, David, et al. "Competitive deep reinforcement learning over a pokémon battling simulator." *2020 IEEE international conference on autonomous robot systems and competitions (ICARSC)*. IEEE, 2020.
- [14] Hu, Sihao, Tiansheng Huang, and Ling Liu. "Pokéllmon: A human-parity agent for pokémon battles with large language models." *arXiv preprint arXiv:2402.01118* (2024).
- [15] Yashwanth, Tadisetty Sai. "A Multi-Agent Pokemon Tournament for Evaluating Strategic Reasoning of Large Language Models." *arXiv preprint arXiv:2508.01623* (2025).
- [16] VGC AI Framework (webpage): <https://gitlab.com/DracoStriker/pokemon-vgc-engine>. 2025.
- [17] Reis, Simão, Luís Paulo Reis, and Nuno Lau. "Vgc ai competition-a new model of meta-game balance ai competition." *2021 IEEE Conference on Games (CoG)*. IEEE, 2021.
- [18] Reis, Simão, et al. "A New Rules Balance Track for the Pokémon VGC AI Competition 2.0." *2025 IEEE Conference on Games (CoG)*. IEEE, 2025.
- [19] Bäck, Thomas, ed. *Evolutionary computation 1: Basic algorithms and operators*. CRC press, 2018.

Índice de autores

Alaguero, Mario, [107](#)

Arnedo-Moreno, Joan, [45](#)

Barrachina-Argudo, Alejandro, [127](#)

Barragán, Rocío, [163](#)

Benítez-Eyzaguirre, Lucía, [19](#)

Bermúdez Martínez, María del Pilar, [13](#)

Bermúdez, María del Pilar, [7](#)

Caballero-Hernández, Juan Antonio, [19](#), [35](#)

Cambero, Anabel, [113](#)

Cano-Iglesias, María José, [151](#)

Castellanos Silva, Matt, [81](#)

Castilla, Daniel, [23](#)

Castilla-Rodríguez, Daniel, [13](#)

Chover, Miguel, [57](#)

Contreras Puerto, Alberto, [35](#)

Del Águila Martín, Cristina, [93](#)

Franco-Mariscal, Antonio Joaquín, [151](#)

Gallego Durán, Francisco J., [139](#)

García-Sánchez, Pablo, [163](#)

Gutiérrez Vela, Francisco Luis, [93](#)

Gómez, Iñigo, [23](#), [107](#)

Hitos García, Alberto, [93](#)

Juárez-González, Pedro, [151](#)

Laplana Curtido, Oriana, [31](#)

León Miranda, Pedro, [81](#)

León, Carlos, [69](#)

Llorens-Largo, Faraón, [139](#)

Martin, Micaela Y., [57](#)

Martín, Roberto, [7](#), [23](#), [107](#)

Marín Lora, Carlos, [57](#)

Medina Medina, Nuria, [93](#)

Montaña-Miranda, Nàdia, [57](#)

Mora, Antonio, [163](#)

Muñoz Ortega, Andrés, [35](#)

Paderewski Rodríguez, Patricia, [93](#)

Palomo-Duarte, Manuel, [19](#), [35](#)

Pérez Romero, Jana María, [1](#)

Pérez, Raúl, [69](#)

Quintas, Roi, [69](#)

Ramírez, Miguel, [69](#)

Rivera, David, [69](#)

Roa-Domínguez, Juan Antonio, [19](#)

Rodríguez-Rivera, Paula, [45](#)

Romero-Hernández, Alejandro, [127](#)

Sagredo-Olivenza, Ismael, [81](#), [113](#), [127](#)

Sánchez-Coterón, Lara, [113](#)

Sánchez-Martín, Pablo, [127](#)

Villar-López, María Beatriz, [57](#)

Viudes-Carbonell, Sergio J., [139](#)

Zarzuela, Claudia, [69](#)