

## Herramienta de Realidad Virtual Inmersiva para entrenamiento de habilidades sociales en personas con discapacidad en un entorno cuidado

Nelson Acosta<sup>1</sup>, Analía Amandi<sup>1,3</sup>, Patricia Salguero<sup>2</sup>, Santiago Faiella<sup>1</sup>, & Franco Viduzzi<sup>1</sup>

<sup>1</sup>NICE, Facultad de Ciencias Exactas, UNICEN – Tandil – Argentina

<sup>2</sup>Casabella Centro de Día – Moreto 732 – CABA – Argentina

<sup>3</sup>CONICET

Mails: [el.nelson.acosta@gmail.com](mailto:el.nelson.acosta@gmail.com), [analia.amandi@gmail.com](mailto:analia.amandi@gmail.com),  
[psalguero@centrocasabella.com.ar](mailto:psalguero@centrocasabella.com.ar), [santiagofaiella@gmail.com](mailto:santiagofaiella@gmail.com), [viduzzi franco@gmail.com](mailto:viduzzi franco@gmail.com)

### Resumen

Este artículo presenta AgileMotion, una herramienta diseñada para la generación de escenas virtuales interactivas, enfocada en el micro-aprendizaje activo y pasivo. El objetivo del desarrollo es realizar una prueba de concepto real en la rehabilitación y la inclusión social con personas con autismo y otras discapacidades del desarrollo, intelectuales y psicosociales, permitiendo la evaluación y aprendizaje de habilidades sociales, aprendizaje y entrenamiento en un entorno cuidado. Utiliza especificaciones en lenguaje natural para facilitar la creación de escenarios por parte de usuarios no expertos en desarrollo de software, como terapeutas o educadores. Ofrece una generación de escenas flexible y adaptable a las necesidades específicas de cada usuario, utilizando lenguaje natural como interfaz principal, y permitiendo que el personal médico interactúe en la escena a través de su propio avatar. Las principales herramientas utilizadas son Unity, RASA, y Photon.

**Palabras clave:** Realidad Virtual (RV), RV en discapacidad, Micro-aprendizaje usando RV, Ambientes generados por IA, Tecnología en discapacidad

### 1.-Introducción

El aprendizaje experiencial y la realidad virtual (RV) han revolucionado la educación y la terapia, permitiendo la creación de entornos inmersivos para el aprendizaje activo y la simulación de situaciones de la vida real. Sin embargo, el desarrollo de este tipo de experiencias suele requerir conocimientos técnicos avanzados, lo que limita su accesibilidad para educadores y terapeutas. Esta propuesta surge como una solución a esta problemática, ofreciendo una herramienta intuitiva y flexible para la generación de escenas virtuales a partir de lenguaje natural. El desarrollo de entornos virtuales para el aprendizaje y la terapia ha experimentado un crecimiento exponencial

en los últimos años. Plataformas como Second Life (Sec2024), OpenSim (Ope2024), Metaverso (Met2024) y VirBELA (Vir2024) han permitido la creación de mundos virtuales para la colaboración y el aprendizaje a distancia. En el ámbito de la simulación para la terapia, herramientas como Psious (Psi2024) y Limbix VR (Lim2024) ofrecen experiencias inmersivas para el tratamiento de fobias, ansiedad y otros trastornos.

También hay muchas herramientas de uso masivo, entre las cuales por mencionar algunas. Synthesia, es una de las más populares, permite crear videos a partir de texto con avatares realistas (Syn2024). RunwayML ofrece una amplia gama de herramientas de edición de video basadas en IA, incluyendo generación de imágenes, eliminación de objetos y efectos especiales (Run2024). Pictory se especializa en transformar artículos y videos largos en clips cortos y atractivos para redes sociales (Pic2024). DeepBrain AI permite crear videos a partir de texto, documentos o incluso otros videos (Dee2024). InVideo es una plataforma versátil que combina herramientas de edición tradicionales con funciones de IA para generar videos de marketing (InV2024). Sora es una plataforma creada como herramienta para la creación de mundos digitales, generando videos por IA con una increíble definición y realismo (Sor2024). Estas herramientas tienen algunas grandes ventajas, principalmente en cantidad de avatares con personificación, cubren varios idiomas, incluso algunas tienen muy buena calidad de imagen. Sin embargo hay también varias desventajas que imposibilitan su uso, precio elevado para acceder a la herramienta completa, algunos permiten videos de corta duración, no permiten que se pueda interactuar en línea en el escenario, otras requieren un proceso de desarrollo complejo.

Este artículo presenta AgileMotion, una herramienta innovadora diseñada para la generación de escenas virtuales interactivas, enfocada en el micro-aprendizaje activo y pasivo. AgileMotion utiliza especificaciones en lenguaje natural para facilitar la creación de escenarios por parte de usuarios no expertos en desarrollo de software, como terapeutas o educadores. Ofrece una generación de escenas flexible y adaptable a las necesidades específicas de cada usuario, utilizando lenguaje natural como interfaz principal, y permitiendo que el personal médico interactúe en la escena a través de su propio avatar. La herramienta se basa en la sinergia entre Unity, un motor gráfico 3D en tiempo real, y RASA, un framework de código abierto para el desarrollo de asistentes inteligentes basados en lenguaje natural. AgileMotion permite la creación de "Agilebots", avatares inteligentes capaces de interactuar con el entorno virtual y entre sí, siguiendo instrucciones detalladas en un "Guión" similar a un guion de teatro o película.

El artículo describe la arquitectura de AgileMotion, incluyendo el rol del "Director" como núcleo central del sistema, la gestión de acciones y subacciones, y la integración del AnimationComposer para la creación de animaciones dinámicas y realistas. Además, se analiza la evolución del diseño de AgileMotion, desde su concepción inicial hasta su estado actual, destacando las decisiones de diseño clave y las soluciones implementadas para superar las limitaciones encontradas.

## 2.-Arquitectura propuesta

Este trabajo propone una herramienta que facilita la creación de escenas virtuales interactivas para el micro-aprendizaje activo y pasivo, utilizando especificaciones en lenguaje natural. Los objetivos principales buscados en esta herramienta son:

- **Facilidad de uso:** La herramienta está diseñada para ser utilizada por usuarios sin conocimientos técnicos avanzados en desarrollo de software.
- **Flexibilidad:** permite la creación de una amplia gama de escenarios y situaciones, adaptándose a las necesidades específicas de cada usuario.
- **Interactividad:** Los "Agilebots" son capaces de interactuar con el entorno virtual y entre sí, teniendo tres formas de trabajo:
  - Instrucciones detalladas en un "Guión".
  - De acuerdo al comportamiento establecido de forma autónoma.
  - Representando al personal médico que está guiando la sesión. En este caso el personal médico toma el control del Agilebot, siendo su representación directa en vivo en el ambiente virtual, e interactuando directamente con los demás usuarios.
- **Animaciones realistas:** La integración del AnimationComposer permite la creación de animaciones dinámicas y realistas para los "Agilebots".

El flujo de trabajo consiste en los siguientes pasos: el usuario escribe un guion en lenguaje natural, el director a través del GuionLoader lo traduce a una serie de acciones, estas acciones son interpretadas por CharlaTan para controlar los AgileBots, quienes interactúan entre sí y con el entorno virtual de Unity según las acciones definidas. Unity renderiza la escena en tiempo real, incluyendo animaciones, audio y video.

Si se analiza la estructura principal a nivel de procesamiento nos encontramos con los siguientes componentes básicos:

- **Director (GuionLoader):** Este componente central recibe el guion y lo traduce a una serie de "Acciones" interpretables por el sistema. Utiliza los patrones Singleton y Chain of Responsibility para la correcta ejecución del guion.
- **ComponentManager:** Crea objetos del tipo IAccion (Interfaz de acción) a partir de la lista de acciones del GuionLoader.
- **IAccion:** Define la interfaz para las acciones, con métodos como Preload, Configure, Execute, Pause, Notify, Evaluar Condición, InterceptObserver y Skip.
- **charlaTAN:** Un sistema multiagentes basado en RASA (framework de Python) que permite a los Agilebots comunicarse mediante lenguaje natural y responder a estímulos de humanos, otros agentes o el Director.

Desde el enfoque del nivel de ejecución tenemos los siguientes componentes:

- Agilebots: Avatares inteligentes que representan a los usuarios en el mundo virtual. Estos pueden ser controlados directamente o actuar de forma autónoma según las acciones del guion.
- Unity: Motor gráfico que renderiza el mundo virtual en 3D en tiempo real.
- Photon Unity Networking (PUN): Framework que permite la sincronización en tiempo real de avatares y la creación de escenarios mediante RPC (llamadas a procedimiento remoto).
- Agora: Framework para la transmisión de video y audio, permitiendo la comunicación entre usuarios en el mundo virtual.
- Animation Composer: Herramienta que permite crear y ejecutar animaciones complejas y dinámicas para los Agilebots.

En líneas generales, AgileMotion se compone de los siguientes módulos principales:

- Director: Actúa como el núcleo central del sistema, gestionando la carga, ejecución y reinicio de los "Guiones".
- Guion: Define la secuencia de acciones que se llevarán a cabo en la escena virtual, utilizando un formato similar a un guion de teatro o película.
- Acciones: Representan las diferentes acciones que los "Agilebots" pueden realizar, como moverse, interactuar con objetos o hablar.
- AnimationComposer: Permite la creación y gestión de animaciones dinámicas y realistas para los "Agilebots".
- Interfaz de usuario: Proporciona una interfaz gráfica intuitiva para la creación y edición de "Guiones", así como para la configuración de las acciones y animaciones.

### 3.-Desarrollo

Este proyecto nace como una exploración de las capacidades del framework AgileTalk (con su núcleo desarrollado en 2018), con un objetivo específico: determinar la viabilidad de utilizar lenguaje natural para poder convertir un conjunto de oraciones en una secuencia de acciones que los Agilebots pudieran ejecutar. Lo estrecho del objetivo llevó a una rápida prueba de concepto donde un Agilebot puede moverse, agarrar y dejar cosas, e interactuar con su entorno de manera limitada. El agente se basa principalmente en tres tecnologías: RASA para el procesamiento del lenguaje natural y comportamiento básico del agente, Unity para la inmersión en un ambiente virtual, y Photon para sincronizar todos los participantes del mundo virtual.

La definición de El\_Director como centro de la arquitectura de todo el sistema tuvo el objetivo principal de ampliar las posibilidades de interacción de los Agilebots, de manera más compleja y abstracta, tanto con su entorno, como también entre sí. Entre las habilidades consideradas está la de colocar objetos sobre muebles del entorno, seguir a otro Agilebot en su camino a una ubicación, realizar animaciones de manera sincronizada o con cierto desfase temporal respecto a

otros Agilebots, y comunicarse utilizando la voz. En esencia, se aspira a dotar a los Agilebots de habilidades que les permitieran comportarse de forma muy parecida a lo que podría hacer una persona en el mundo real.

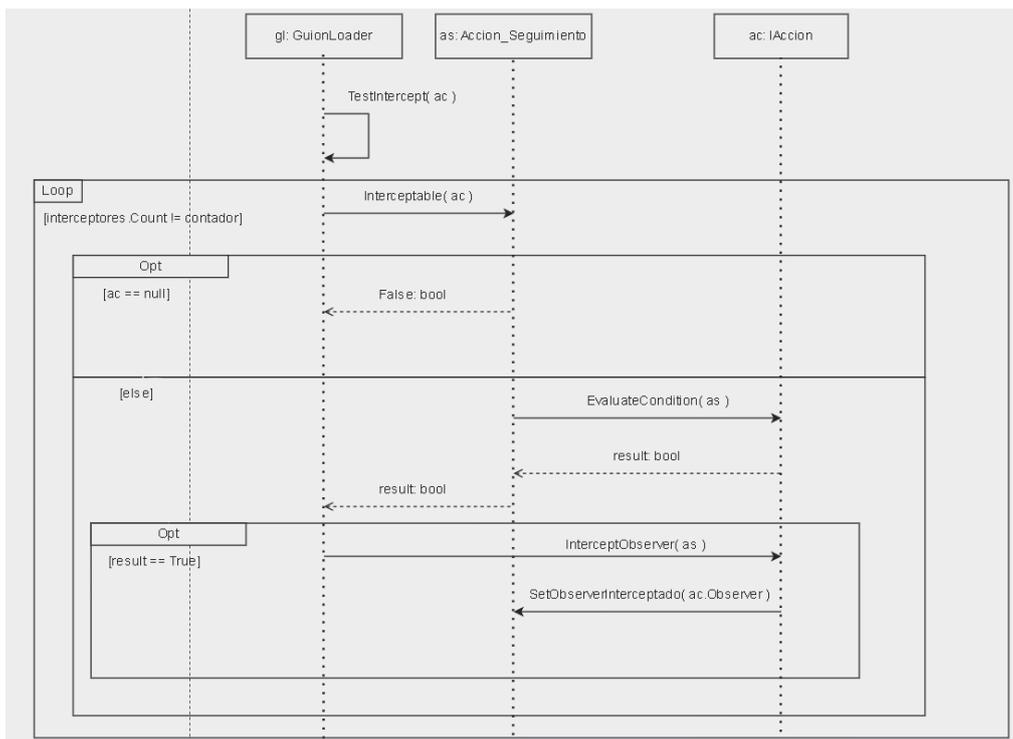


Fig. 1.-Diagrama de secuencia del funcionamiento de la intercepción de acciones

La incorporación del AnimationComposer es fundamental para facilitar la creación de animaciones. Anteriormente las animaciones debían ser una sola secuencia que moviera todo el avatar, encargándose de mover cada parte del cuerpo. Esta pieza de la arquitectura permite fragmentar las animaciones en partes más pequeñas y fáciles de manejar, dando origen a las animaciones “Atómicas”. Al atomizarlas cada animación puede hacerse mucho más detallada, sin necesidad de prestar atención a lo que debían hacer otras partes del avatar. Pero el gran aporte de esta herramienta es que ahora estas animaciones “Atómicas” podrían además combinarse para formar animaciones “Compuestas”, es decir, animaciones más grandes y complejas creadas a partir de la combinación de las piezas atomizadas. Esta flexibilidad permite generar una gran variedad de movimientos y acciones, sin tener que crear cada animación desde cero. La capacidad de atomizar y componer animaciones con el AnimationComposer es clave para el dinamismo de AgileMotion, ahora es posible animar los Agilebots con movimientos más naturales y sofisticados, permitiendo crear y expandir bibliotecas de animaciones.

La Fig. 1 muestra un diagrama de una parte del funcionamiento de la intercepción de acciones, esquema que permite la interacción básica entre varios chatbots que controlan los diferentes avatares en el entorno virtual.

#### **4.-Visualización.**

Una parte significativa de las acciones desarrolladas incluye un parámetro para seleccionar el modo de la cámara (Fig. 2) con un componente de la clase *Cameras* que sintetiza toda la funcionalidad de control de cámaras. De esta manera, al ejecutar una acción que cuente con dicho parámetro, la cámara se posiciona sobre el avatar según el modo elegido. Algunos de los modos de cámara disponibles son inspirados en el género literario donde tenemos la primera persona (modo 1, se ve el mundo con el avatar como protagonista), segunda persona (modo 2, se ve como partícipes de la escena pero no el protagonista), tercera persona (modo 3, permite ver más claramente el avatar y sus alrededores), la cámara de frente para enfocar en un plano medio al avatar (modo 4), la cámara trasera para ver lo que está detrás del avatar (modo 5), o simplemente no utilizar la cámara directamente (modo 0, si no es necesaria la cámara para esa acción).

Se le ha dado mucha importancia a la funcionalidad que permite visualizar las escenas, tratando de realizar composiciones (gráficas) más elaboradas. Para esto, se emplea un avatar llamado *Camarógrafo* (que por decisión de diseño es invisible) en una *Accion\_Conjunta*. Este avatar utilizará la *Accion\_Seguimiento* con la condición establecida por la acción para la cual se desea una toma más compleja. Además, mediante la *Accion\_Posicionar* o la *Accion\_MoverAvatar*, se puede desplazar al camarógrafo para obtener tomas aún más sofisticadas y/o complejas como lo son paneos (movimientos de cámara, sean de rotación en cualquiera de los tres ejes o acercamiento/alejamiento) o cambios de planos a voluntad (primer plano enfocando la cara, plano medio enfocando los hombros y cara, etc) y con la gradualidad que se desee (controlando la velocidad de movimiento del camarógrafo o directamente teletransportarlo a la siguiente posición).

#### **5.-Análisis**

*AgileMotion* ha sido probado en algunos diferentes escenarios de micro-aprendizaje, demostrando su eficacia para la creación de experiencias interactivas y personalizadas.

El uso de lenguaje natural como interfaz principal ha demostrado ser efectivo para facilitar la creación de escenarios por parte de usuarios no expertos en desarrollo de software.

La integración del *AnimationComposer* ha permitido la creación de animaciones más realistas y dinámicas, mejorando la calidad general de las experiencias.

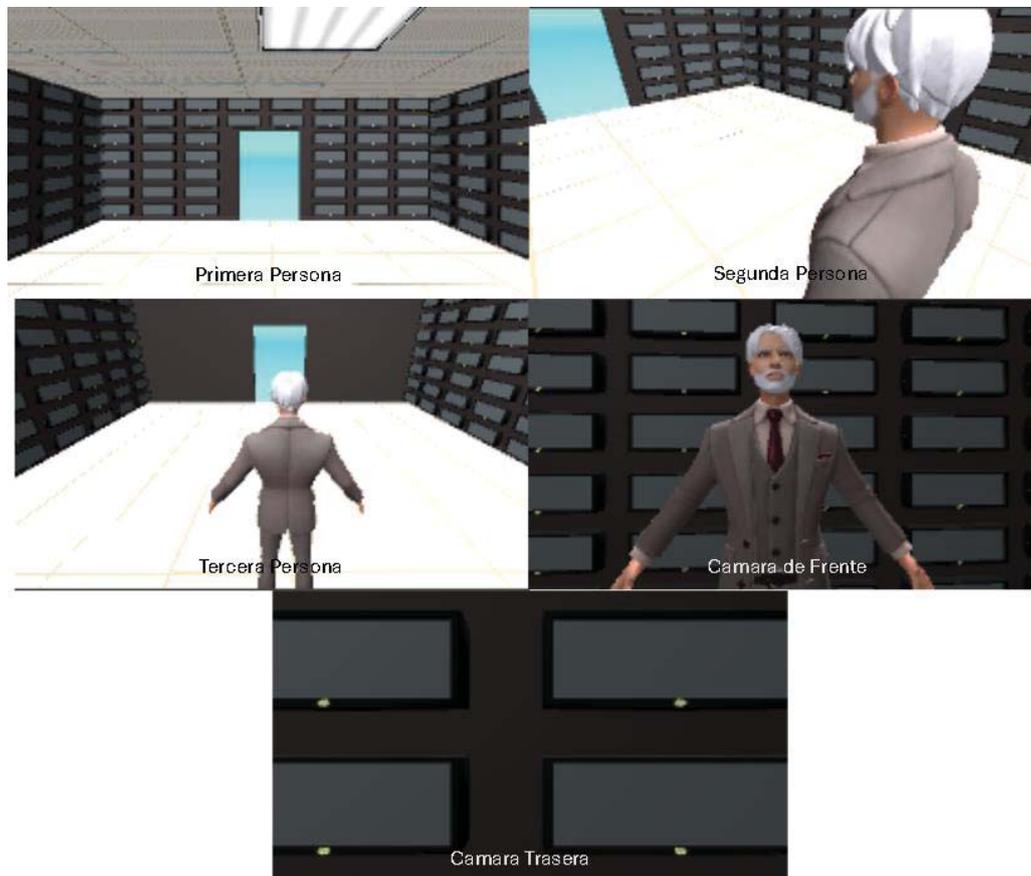


Fig. 2.-Diferentes posiciones de la cámara según el modo en que se utilice

Las limitaciones presupuestarias son la base de varias restricciones del sistema:

- Hay un gran mercado de voces, de contar con ese material sería una poderosa característica para potenciar la empatía de los participantes. Hay actualmente herramientas que permiten entrenar una IA para generar una determinada voz en particular, esto potenciaría si se puede hacer que el avatar que represente a una persona tenga su propia voz.
- Con los personajes, actores o avatares sucede exactamente lo mismo que con las voces. También hay formas en las que se puede generar un avatar basado en una persona determinada, esto permitiría que las personas se vinculen en mayor medida con el ambiente.

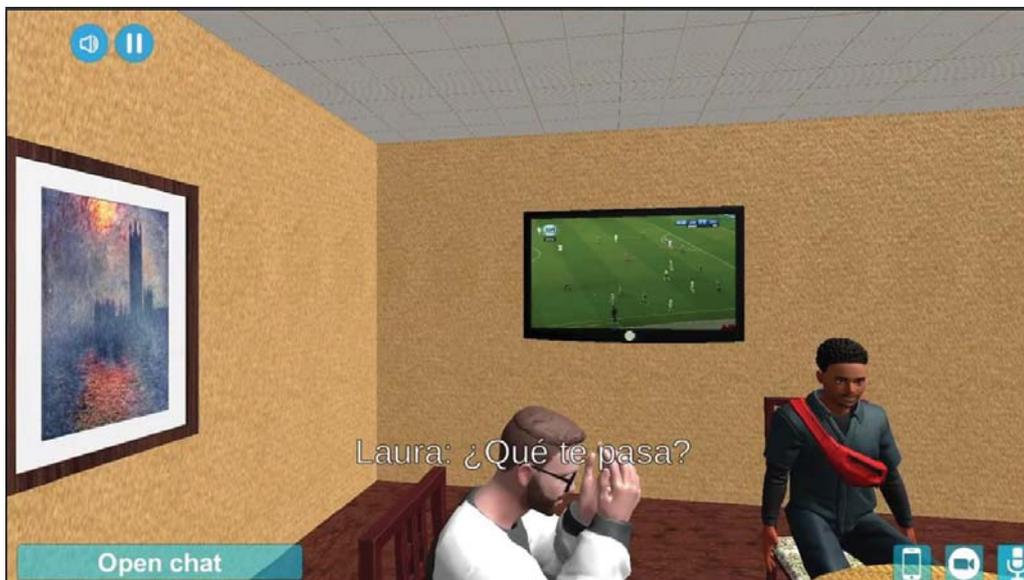


Fig. 3.- (Captura)... Personas sentada tapándose la cara y persona sentada, en una confitería con mucho ruido

Con los edificios y áreas (supermercado, hotel, escuela, etc) donde los personajes deben interactuar sucede lo mismo, son más horas dedicadas de diseñadores o adquiriendo el material de proveedores externos. En las pruebas se hicieron mundos virtuales de forma amateur, pero se podría representar el barrio de las personas involucradas, de tal forma que la inmersión sea mucho más real.

Se muestran varias capturas del mundo virtual donde los personajes interactúan. La Fig. 3 muestra una escena en una confitería con personas sentadas tapándose la cara molesta por el ruido ambiente.

Mientras que la Fig. 4 muestra una animación en objetos en el parque (calesita y otros personajes en movimiento) y se trabaja lo peligroso de tratar de subir o bajar con la calesita en movimiento y el saber que debe esperar su turno para acceder a los juegos. La Fig. 5 muestra una comunicación con el celular en ambiente virtual, permite la comunicación real en el ambiente virtual. La Fig. 6 muestra la ejecución de aplicaciones en el celular del mundo virtual.

Centro Casabella es un Centro de Rehabilitación Especializado para niños, niñas y adolescentes que presentan dificultades en la comunicación, en las relaciones sociales, las emociones, la conducta, la motricidad y otras problemáticas. Con su accionar diario promueve la igualdad de todos los derechos y libertades para los niños y jóvenes con discapacidad, según se establece en la Convención Internacional sobre los Derechos de las Personas con Discapacidad y otros tratados internacionales. Así mismo apoya el valor transformador del conocimiento y de la tecnología al

servicio del desarrollo de la calidad de vida de las personas y de las familias y la cooperación y el trabajo conjunto de profesionales, familia, personas cercanas y de la comunidad, como la forma de gestión, a fin de conseguir resultados eficaces.



Fig. 4.- (Captura)... Animación en objetos en el parque (calesita y otros personajes en movimiento)

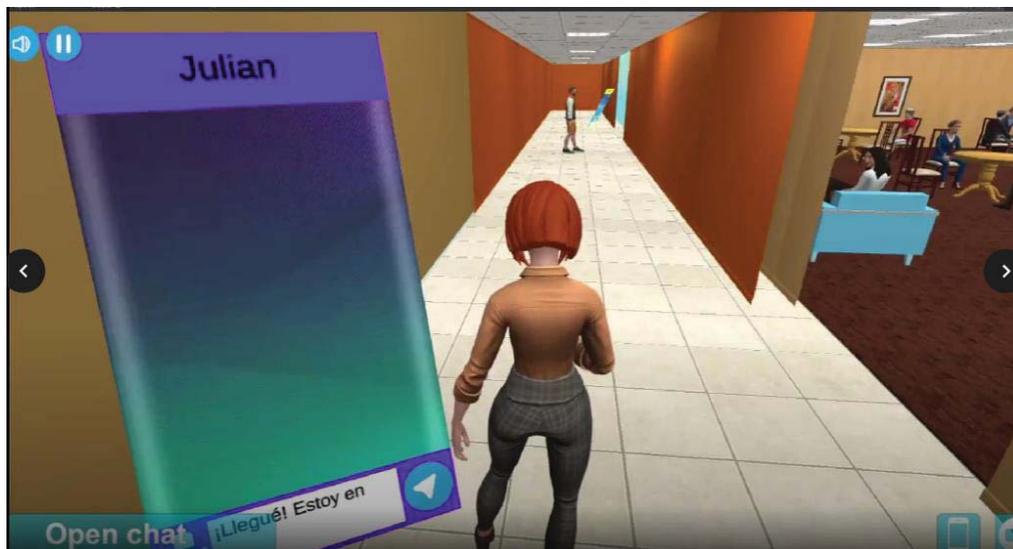


Fig. 5.- (Captura)... comunicación con el celular en ambiente virtual

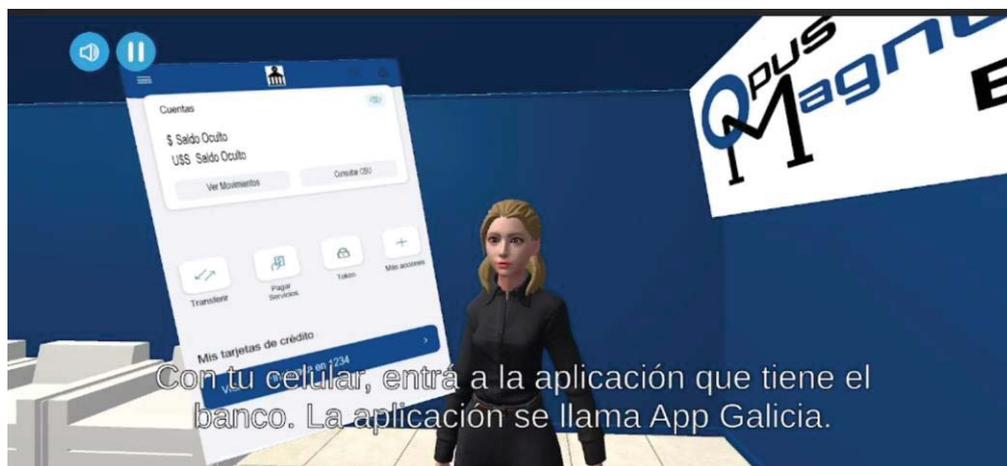


Fig. 6.- (Captura)... Ejecución de aplicaciones en el celular del mundo virtual

Es con estas bases que, en el marco de este proyecto "Casabella" aportó el equipo de trabajo especializado en TEA, psicólogos y personal especializado en el área, que se encargó de definir los primeros entornos de simulación de escenas de entrenamiento y también de realizar las pruebas con personas con TEA en el espacio institucional.

De esta forma, el trabajo de campo se dividió en dos etapas claramente definidas.

En una primera etapa el grupo de especialistas conformado por psicólogos especializados en TEA confeccionaron 4 guiones y la descripción del entorno, en formato microlerning sobre conductas de habilidades sociales. Para ello se tuvo en cuenta que el proceso del entrenamiento en Habilidades Sociales debe implicar cuatro elementos de forma estructurada:

1. Entrenamiento en habilidades, donde se enseñen conductas específicas y, se practican y se integran en el repertorio conductual del sujeto.
2. Reducción de la ansiedad en situaciones sociales problemáticas.
3. Reestructuración cognitiva, en donde se intentan modificar valores, creencias, cogniciones y/o actitudes del sujeto.
4. Entrenamiento en solución de problemas, donde se enseña a los sujetos a percibir correctamente los valores de todos los parámetros situacionales relevantes, a procesarlos, a seleccionar una respuesta y enviarla.

Este entrenamiento suele encontrarse de forma implícita, en los entrenamientos en habilidades sociales. De los cuatro guiones se seleccionaron dos (La calesita y el bar) sobre los cuales trabajó el equipo de profesionales especializado en informática

En la segunda etapa, el grupo de especialistas que confeccionó los guiones evaluó las primeras versiones del entorno virtual solicitando algunos ajustes y modificaciones hasta que se llegó a un

demo factible de ser presentado a un grupo de adolescentes de entre 11 y 13 años con Trastorno del Espectro Autista concurrentes al centro y se consignaron los resultados. Los mismos tendían a la mejora de los entornos y los avatares para ajustarse a los entornos reales.

## 6.-Conclusiones

En este artículo se ha presentado una herramienta prometedora para la generación de escenas virtuales interactivas para el micro-aprendizaje activo y pasivo. Su facilidad de uso, flexibilidad y capacidad para crear animaciones realistas la convierten en una solución ideal para educadores, terapeutas y otros profesionales que buscan crear experiencias de aprendizaje inmersivas y personalizadas. Se ha investigado y comparado varias herramientas que usan lenguaje natural para la creación de escenarios virtuales, y se ha analizado las tendencias actuales en el desarrollo de RV para el aprendizaje y la terapia.

Durante el desarrollo de las pruebas en campo, con personas que se están tratando en Casabella, se han encontrado varios problemas. Algunas de las personas, a las que está orientada la herramienta, se les dificultan el uso de un sistema de realidad virtual HMD. Al trabajar con avatares y voces estándares (gratuitos) en lugar de los que representan a los pacientes, a veces les cuesta identificarse. En casos donde el paciente se le dificulta controlar su movimiento en el ambiente, se han generado videos para que pueda vivir de forma pasiva la experiencia.

La herramienta desarrollada cumple con el objetivo de realizar una prueba de concepto real que permita cubrir un área de vacancia en la rehabilitación y la inclusión social con personas con autismo y otras discapacidades del desarrollo, intelectuales y psicosociales porque permite la evaluación de las habilidades sociales, el aprendizaje y el entrenamiento de ellas en un entorno cuidado.

### Agradecimientos:

- Fundación SADOSKY, que financió parte del desarrollo del proyecto durante el año 2023 a través de la “Convocatoria de Soluciones Innovadoras para Desafíos de Software” (<https://fundacionsadosky.org.ar/soluciones-innovadoras/>)
- Centro de Día Casabella (<http://centrocasabella.com.ar/>), con sede en Moreto 732, Ciudad de Buenos Aires, Argentina. Tel: 11 4672-3643 / 11 2183-2018

### Bibliografía

- (Ago2024) Agora Unity: <https://www.agora.io/en/unity/>
- (Cam2022) Campo, Marcelo & Amandi, Analía & Conte, Argentina & Acosta, Nelson & Mazza, Argentina & Guerrero, Matias & Berthelot, Matias & Romano, Argentina. (2022). Let's see... Let's talk: Towards an Intelligent VR Platform for Training Professionals in Disabilities. (Cas2023) Centro Casabella: <http://centrocasabella.com.ar/>
- (Dee2024) <https://deepbrain.io>. 2024.
- (InV2024) <https://invideo.io/make/add-text-to-video-online/>. 2024.

- (Lia2021) SAFe PI Planning. A Step-By-Step Guide to PI Planning Using the Scaled Agile Framework. Liam Kane.
- (Lim2024) <https://ismguide.com/resources/limbix-vr/>. 2024.
- (Met2024) Metaverso de Meta Inc. <https://about.meta.com/ltam/metaverse/>. 2024.
- (Ope2024) <https://opensim.stanford.edu/>. 2024.
- (Pho2024) Photon Unity Networking: <https://www.photonengine.com/>
- (Pic2024) <https://pictory.ai/?el=2000&htrafficsource=pictoryblog>. 2024.
- (Psi2024) <https://espacio.fundaciontelefonica.com/empresa/psious/>. 2024.
- (Ras2024) Rasa conversational AI: <https://rasa.com/>
- (Run2024) <https://runwayml.com/>. 2024.
- (Sad2023) Fundacion Sadosky: <https://fundacionsadosky.org.ar/>
- (Sec2024) <https://id.secondlife.com/openid/login?>. 2024.
- (Sor2024) <https://openai.com/index/sora/>. 2024.
- (Syn2024) <https://synthesia.io>. 2024.
- (Vir2024) <https://www.virbela.com/why-virbela/what-is-virbela>. 2024.