

## INCLUVERSO 5G

Tecnologías de realidad exteNdida y Comunicaciones para la inclUsión de personas en situación de VulnERabilidad psicoSOcial mediante redes 5G avanzadas



SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES



---

Entregable E1.2

Definición de requisitos y casos de uso de INCLUVERSO

---

# Contenidos

1	Introducción.....	5
2	Terapia.....	6
	2.1 Descripción .....	6
	2.2 Necesidades que resuelve y beneficios esperados .....	7
	2.3 Implicaciones tecnológicas.....	9
	2.3.1 Creación de escenario virtual.....	9
	2.3.2 Segmentación del cuerpo e integración con el escenario virtual.....	11
	2.3.3 Captura de bioseñales.....	12
	2.3.4 Aplicación de control para terapeuta .....	13
	2.3.5 Necesidades de conectividad.....	14
	2.4 Evaluación de la interacción y QoE.....	14
3	Telepresencia.....	16
	3.1 Descripción .....	16
	3.1.1 Necesidades que resuelve y beneficios esperados .....	17
	3.2 Implicaciones tecnológicas.....	19
	3.2.1 Calidad perceptual del sistema de telepresencia .....	19
	3.2.2 Accesibilidad de Snowl .....	19
	3.2.3 Representación del profesional de apoyo remoto (“presencia inversa”) .....	20
	3.2.4 Necesidades de conectividad.....	20
	3.3 Evaluación de la interacción y QoE.....	21
4	Teleformación.....	22
	4.1 Descripción .....	22
	4.2 Necesidades que resuelve y beneficios esperados .....	23
	4.3 Implicaciones tecnológicas.....	25
	4.3.1 Interacción persona-máquina (HCI) .....	25
	4.3.2 Captura de bioseñales y regulación emocio-cognitiva .....	26

4.3.3	Necesidades de conectividad.....	26
4.4	Evaluación de la interacción y QoE.....	26
5	Definición de requisitos de realidad extendida e interacción.....	28
5.1	Introducción a los sistemas de realidad extendida .....	28
5.2	Calidad de Experiencia en realidad extendida.....	28
5.3	Estado del arte de algoritmos de realidad extendida e interacción persona-máquina.....	29
5.3.1	Tecnologías de presencia autónoma .....	29
5.3.2	Tecnologías de interacción natural .....	30
5.3.3	Tecnologías de telepresencia.....	31
5.3.4	Tecnologías de biomarcadores digitales y regulación emocional .....	32
5.4	Requisitos de inmersividad e interacción de los casos de uso .....	32
5.4.1	Terapia.....	33
5.4.2	Telepresencia.....	36
5.4.3	Teleformación .....	37
5.5	Escenarios base.....	40
5.5.1	Escenario 1: Teleportación Virtual (Virtual Teleportation) .....	40
5.5.2	Escenario 2: Procesamiento delegado de Realidad Mixta ( <i>Mixed Reality computation offloading</i> ).....	42
5.5.3	Escenario 3: Regulación Emocional ( <i>Emotion Regulation</i> ).....	43
5.5.4	Correspondencia entre los casos de uso y los escenarios base.....	45
5.5.5	Requisitos e indicadores de rendimiento.....	46
6	Definición de requisitos de la red de comunicaciones .....	48
6.1	Introducción al 5G avanzado.....	48
6.2	Estado del arte de tecnologías de red .....	48
6.2.1	Arquitecturas de red 5G privadas .....	48
6.2.2	Productos en banda n40 .....	50
6.2.3	Emulación de la red de acceso radio.....	51
6.2.4	Realidad extendida en redes 5G avanzadas .....	52
6.3	Requisitos de conectividad y procesamiento de los escenarios base.....	54

6.3.1	Requisitos de movilidad y cobertura .....	54
6.3.2	Requisitos de tráfico .....	55
6.3.3	Análisis de 5QIs existentes .....	55
6.3.4	Requisitos de procesamiento .....	57
6.4	Especificación de los entornos de simulación de referencia.....	58
7	Acrónimos.....	59
8	Índice de tablas.....	61
9	Índice de figuras .....	62
10	Referencias.....	63

# 1 Introducción

En este entregable se recogen de manera detallada los casos de uso que guían los desarrollos del proyecto. Se consideran tres casos de uso principales basados en las necesidades de la Fundación Juan XXIII Roncalli (FJ23) y los beneficios que la tecnología Realidad eXtendida (XR) supone. Específicamente, los casos de uso son orientados a: terapia, centrada en la técnica de desensibilización sistemática, telepresencia y teleformación. Cada uno de ellos se describe en detalle, presentando las necesidades que atiende, implicaciones tecnológicas, así como otros aspectos significativos a tener en cuenta en el desarrollo del proyecto. Asimismo, se recogen las metodologías utilizadas por la Universidad Politécnica de Madrid (UPM) para la evaluación de la interacción y la calidad de experiencia de los usuarios, con especial atención a la interacción persona-máquina, captura de bioseñales y regulación emocio-cognitiva, siempre que aplique al caso de uso. Por último, en cada caso de uso la Fundación Juan XXIII Roncalli describe el tipo de actuación que lleva a cabo, así como los beneficios y resultados esperados.

A partir de las descripciones de los casos de uso, se presentan en las siguientes secciones los requisitos tanto de las tecnologías de XR como de las redes de comunicaciones abordados en el proyecto INCLUVERSO 5G. Para cada uno de los casos de uso del proyecto (terapia, telepresencia, teleformación) se concretan los detalles de implementación y se elaboran sus requisitos funcionales y no funcionales de alto nivel.

## 2 Terapia

### 2.1 Descripción

El caso de uso de terapia se centra en aplicar la tecnología XR para que pacientes puedan asistir a sus terapias y tener acceso a lugares remotos. Uno de los principales factores que se tienen en cuenta es que este tipo de tecnología crea entornos inmersivos en los que la sensación de presencia, definida como la sensación de estar en el lugar que visualiza el paciente con el *Head Mounted Display (HMD)*, es mucho más elevada que utilizando cualquier otra tecnología predecesora. En este sentido, la utilización de tecnologías XR permitiría minimizar costes y recursos permitiendo la simulación de entornos y objetos físicos, así como facilitar el acceso a la terapia a personas que no se pueden desplazar.

Como otro factor diferenciador de la tecnología XR, es que los sensores que poseen los cascos de realidad virtual permiten recoger de una manera más sencilla datos que pueden ser relevantes para estudiar los efectos de la terapia, como es el movimiento de la cabeza durante la sesión. Adicionalmente, para completar el estudio, se recogen señales fisiológicas de los pacientes durante la sesión para realizar un seguimiento de la terapia y de sus efectos en los pacientes.

El caso de terapia consiste en aplicar una de las técnicas tradicionales que se usan en psicología desde la rama cognitivo-conductual para el tratamiento de los miedos y fobias específicas, la Desensibilización Sistemática (DS). Esta técnica fue desarrollada por Wolpe [1] El objetivo consiste en lograr la reducción gradual ("*sistemática*") de la sensibilización ("*desensibilización*") a las situaciones evocadoras de ansiedad, que son imaginadas por el paciente, mientras se lleva a cabo una respuesta incompatible con la ansiedad. De esta forma se trabaja el afrontamiento frente al estímulo que genera ansiedad. La exposición ante estos estímulos temidos se produce de manera supervisada por un terapeuta y se hace progresivamente, teniendo perfectamente identificado el grado de ansiedad que produce cada uno de los estímulos. De esta forma, los/as pacientes aprenden a gestionar las emociones producidas por experiencias ansiógenas. La ansiedad generada por la experiencia ansiógena puede ser reducida a través de la emisión de una respuesta incompatible, como podría ser la respiración, con la que se pretende provocar una respuesta automática que interfiera con el malestar del estímulo aversivo.

En el Centro Ocupacional de Formación, Oportunidades e Inserción Laboral (COFOIL) de la Fundación Juan XXIII Roncalli, se detecta una situación que genera estrés en muchas personas usuarias: el miedo a subir o bajar escaleras. Esta es una situación a la que se tienen que enfrentar cada día (por ejemplo, escaleras del centro de la FJ23) y que, en ocasiones, puede llegar a suponer la no asistencia de manera autónoma al centro (por ejemplo, si a la persona le dan miedo las escaleras mecánicas del metro).

En este caso de uso se pretende aplicar la técnica de DS combinada con Realidad Virtual (RV) en personas usuarias de COFOIL que presentan miedo ante las escaleras, con el objetivo de ayudarles a enfrentarse a esta situación y que esto repercuta en una mejora de su calidad de vida.

Para ello, se establecen las siguientes fases:

- **Elaboración de una jerarquía de situaciones ansiógenas de manera individualizada:** a través de fotografías reales de los estímulos temidos (escaleras) se elaborará una jerarquía personalizada con cada persona participante, ordenando los estímulos en función del nivel de ansiedad que les provocan.
- **Entrenamiento en respiración diafragmática:** como respuesta incompatible con la ansiedad se ha elegido la relajación a través de la respiración diafragmática. Para ello, con la ayuda de las gafas de RV se visualizarán entornos 360 grados que favorezcan un estado de relajación. Además, se incorporará un audio con instrucciones guiadas para que las personas participantes puedan aprender la técnica de respiración más fácilmente.
- **Desensibilización sistemática:** se pretende asociar la situación que provoca la respuesta de ansiedad con una respuesta incompatible con ella. En este caso se presentarán los estímulos temidos (escaleras) a través de las gafas de RV y cuando el nivel de ansiedad de las personas participantes se eleve, se inducirá un estado de relajación a través de la respiración diafragmática aprendida anteriormente. Se espera que por el principio de contracondicionamiento, tras varias repeticiones, el nivel de ansiedad ante la presentación del estímulo temido vaya disminuyendo hasta alcanzar su desensibilización.
- **Generalización:** se espera que la RV contribuya a generalizar los resultados obtenidos a las situaciones de la vida real. Para ello, como fase final de la terapia, se producirá el enfrentamiento a las escaleras en la realidad.

## 2.2 Necesidades que resuelve y beneficios esperados

En los últimos años, la XR se ha empezado a incorporar como herramienta tecnológica para mejorar las intervenciones psicológicas y, precisamente, una de las áreas en las que más se ha utilizado es en los trastornos de ansiedad, especialmente en los miedos y fobias específicas [2] ,[3] . Entre los beneficios que puede proporcionar combinar las técnicas tradicionales de tratamiento de miedos y fobias (exposición y DS) con RV se encuentran los siguientes: favorece una mayor aceptación de la intervención por parte de los pacientes [4] , proporciona una mayor sensación de seguridad [5] lo que facilita la exposición a los estímulos temidos, y permite un mayor control sobre el entorno y sobre las variables a considerar del estímulo aversivo que cuando se realiza en imaginación. Con esto se evitaría algunas de las dificultades que pueden surgir aplicando

DS y que pueden reducir su eficacia clínica tales como, que el paciente no sea capaz de imaginar la situación adecuadamente o se superen los estímulos temidos en imaginación, pero no en la realidad. Por tanto, es esperable que la introducción de esta tecnología en tratamientos tradicionales produzca una mayor adhesión al tratamiento y, en consecuencia, se reduzca la tasa de abandonos terapéuticos.

Hasta el momento, se encuentran numerosas investigaciones con resultados satisfactorios, que corroboran los beneficios descritos anteriormente en los que se combinan las técnicas de intervención psicológica tradicionales con RV para el tratamiento de diferentes miedos y fobias tales como la acrofobia o miedo a las alturas [6] , [7] , [8] la fobia social [9] , [10] , la cinofobia o fobia a los perros [11] , la agorafobia [12] , la fobia a volar [13] y la fobia a los procedimientos médicos [14] . Además, hay trabajos en los que se ha demostrado que la efectividad de los tratamientos para fobias que han empleado RV como herramienta tecnológica son fácilmente extrapolables y se pueden mantener en el tiempo en la vida real de los pacientes [15]

En todas las investigaciones citadas anteriormente las personas participantes han sido sin discapacidad, por lo que no son representativas de toda la población. Resulta fundamental probar estos tratamientos con personas con discapacidad para favorecer la inclusión de este colectivo y garantizar su accesibilidad a los tratamientos [16] . Por tanto, es necesario analizar qué dificultades aparecen durante su aplicación para poder realizar las adaptaciones que sean necesarias y asegurar que todas las personas, tanto con o sin discapacidad, puedan acceder a estas intervenciones.

El beneficio principal que se espera encontrar tras esta intervención es que se reduzcan sus niveles de ansiedad y, por tanto, se mejore su calidad de vida, ya que subir y bajar escaleras es algo a lo que se tienen que enfrentar en muchas ocasiones.

La RV presenta beneficios que pueden mejorar los resultados de la intervención con la técnica de DS en personas con discapacidad intelectual por los siguientes motivos:

- Las personas con discapacidad intelectual presentan limitaciones en su capacidad de imaginación y de razonamiento abstracto. Estos factores no solo influyen en la capacidad de imaginar los escenarios temidos (escaleras en este caso concreto) sino también en la de imaginar los escenarios relajantes que se requieren en uno de los pasos de la DS. Por tanto, la RV puede facilitar este proceso ya que permite la visualización de vídeos de 360 grados para evocar esos estados.
- La RV permite ajustar las variables del entorno de manera personalizada en función de los niveles de ansiedad de la persona.
- Se reducen costes tanto económicos como temporales. Si, por ejemplo, se quiere trabajar la exposición ante las escaleras mecánicas del metro no es necesario desplazarse al lugar in situ, si no que con las gafas de RV se puede trabajar ese miedo independientemente del lugar en el que se encuentren la terapeuta y los/as pacientes.

- La intervención es más segura y controlada. Por tanto, se minimizan los riesgos físicos en la exposición y se proporciona un entorno seguro para trabajar el miedo a las escaleras.

En definitiva, la capacidad de la RV para reproducir y simular múltiples entornos contribuirá a la generalización de los resultados obtenidos con la terapia. Gracias a esto, se espera que las personas participantes puedan transferir las habilidades de afrontamiento adquiridas con esta intervención a las situaciones reales y que eso suponga una mejora en su calidad de vida.

## 2.3 Implicaciones tecnológicas

Las implicaciones tecnológicas de este caso de uso pueden ser categorizadas en los grupos: Creación de escenario virtual, Segmentación del cuerpo e integración con el escenario virtual, Captura de bioseñales, Aplicación de control para terapeuta y Necesidades de conectividad.

### 2.3.1 Creación de escenario virtual

- **Representación de escenarios neutros.** Tienen como objetivo favorecer la relajación y generar confianza en el/la paciente. Este enfoque implica la utilización de vídeos en formato 360 grados presentados a través de gafas de realidad virtual, con el objetivo de fomentar un estado mental positivo en la persona. Algunos vídeos se extraen de bases de datos públicas y simulan diversos entornos como montañas, playas y otros paisajes serenos como puede ser el salón de una casa. La variedad de opciones busca adaptarse a las preferencias individuales, creando una experiencia que se sienta personal y reconfortante para cada persona usuaria.



Figura 1: Ejemplo de una escena de un entorno sereno, un salón de una casa.

- **Representación de las capturas de las escaleras de escenarios conocidos.** Representan escaleras, en este caso, conocidas de las personas usuarias. Para ello, utilizan vídeos 360 grados que se deben capturar en las distintas escenas de

escaleras dentro de la Fundación Juan XXIII Roncalli y alrededores. Al igual que los escenarios neutros, estos vídeos se plantea utilizarlos para su visualización a través de las gafas de RV.



Figura 2: Ejemplo de unas escaleras de la Fundación Juan XXIII Roncalli en proyección equirectangular

- **Visualización de escenas de escaleras virtuales que generan ansiedad a través de las gafas.** Las escenas se desarrollan con base en la jerarquía establecida por la psicóloga y replican tanto entornos agradables con escaleras, como ambientes conocidos que generan miedo. Adicionalmente, estas escenas pueden contar con detalles que se modifiquen por la terapeuta para añadir realismo a la escena y controlar la ansiedad que produce el estímulo, como por ejemplo, poner en movimiento o parar las escaleras de metro. De esta forma, se puede proporcionar experiencias graduales y controladas que fomentan una adaptación positiva a través de la RV.

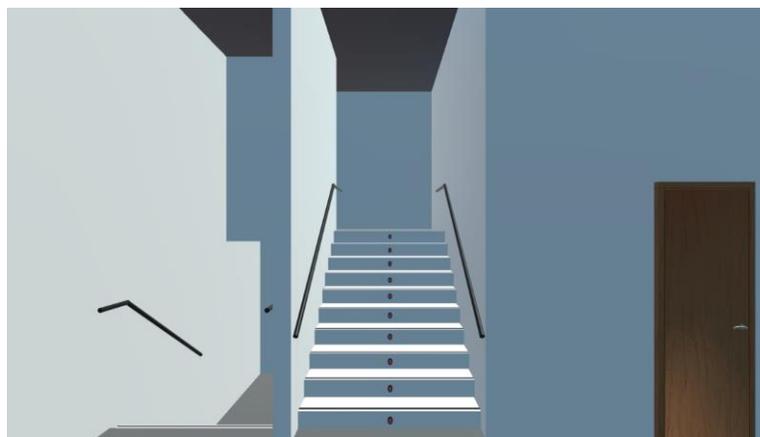


Figura 3: Ejemplo de una escena de realidad virtual que replica unas escaleras de la Fundación Juan XXIII consideradas en la jerarquía

## 2.3.2 Segmentación del cuerpo e integración con el escenario virtual

Uno de los factores que, basándonos en la literatura, es fundamental para tener una buena calidad de experiencia y asegurar el éxito de la terapia es, del inglés, *embodiment*, como la posibilidad de visualizar el cuerpo de la persona que lleva los cascos de realidad virtual en el entorno inmersivo [17]. Adicionalmente, la visualización del cuerpo tiene un valor añadido en este caso de uso ya que necesitamos que los/as pacientes lo perciban cuando se enfrentan al estímulo temido, las escaleras. Las implicaciones tecnológicas en este particular aspecto se relacionan con el algoritmo que se deben implementar para que funcione con una buena fiabilidad, ya que si no podría influir sobre los beneficios de la terapia y no hacerla efectiva. Por ello, se elige aplicar una técnica de segmentación, haciendo que los/as pacientes visualicen su propio cuerpo en lugar de una representación, posiblemente menos realista, como podría ser el cuerpo de un avatar [18]. Asimismo, se precisa que el algoritmo procese la segmentación y devuelva las máscaras con baja latencia para que el retardo no sea percibido por la persona con las gafas de RV y pueda enfrentarse a las escaleras de forma realista. La Figura 4 presenta un ejemplo de cómo vería la segmentación de las manos una persona que lleva puestas las gafas de RV y, de manera similar, la Figura 5 presenta un ejemplo de cómo vería sus piernas y pies.



Figura 4: Ejemplo de cómo una persona con las gafas de RV puede observar sus propias manos gracias al algoritmo de segmentación

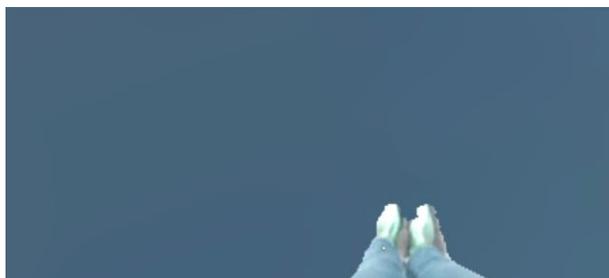


Figura 5: Ejemplo de cómo una persona con las gafas de RV puede observar sus propios pies y piernas gracias al algoritmo de segmentación

## 2.3.3 Captura de bioseñales

Varios estudios ya han demostrado la correlación entre la ansiedad y el estrés con ciertas tendencias en medidas de determinadas bioseñales. En una revisión exhaustiva de estudios sobre la detección de la ansiedad mediante bioseñales, realizada por Giannakakis [19], se encontró que las señales más destacadas para detectar la ansiedad son: la frecuencia cardíaca y su variabilidad, la respuesta galvánica de la piel y la actividad alfa del electroencefalograma (EEG). Otras señales que ofrecen percepciones relevantes son la actividad muscular en el músculo trapecio, la temperatura cutánea alrededor de los dedos y en la zona facial, los movimientos de cabeza y el comportamiento ocular.

Teniendo esto en cuenta, las necesidades de las personas usuarias y las soluciones de sensores disponibles, se seleccionan algunas de las bioseñales informativas para tener en cuenta en este caso de uso.

Mediante el HMD se pueden capturar en las sesiones:

- **Movimientos de cabeza:** Posición y rotación de la cabeza durante la visualización de los contenidos, sesión de terapia, en RV. Desde la cual se pueden obtener métricas de movimiento, patrones de comportamiento y/o trayectorias de visualización.

Mediante una pulsera que se pone en la muñeca de la persona usuaria se puede extraer en las sesiones:

- **Pulso de volumen sanguíneo (BVP)** – del inglés *Blood Volume Pulse*–: variaciones entre el nivel de presión de la sangre desde las cuales se obtiene la frecuencia cardíaca y su variabilidad.
- **Temperatura cutánea:** temperatura de la piel.
- **Respuesta galvánica de la piel:** medida del nivel de conductancia de la piel (SCL, del inglés *Skin Conductance Level*) y la respuesta de la conductancia de la piel (SCR, del inglés *Skin Conductance Response*).
- **Movimientos de la mano:** medidas de aceleración de la mano.

Por último, mediante sensores que se colocan en la zona de interés se puede obtener:

- **Electromiografía (EMG):** medida para conocer la activación de los músculos según las señales nerviosas recibidas. Aplicada para medir la activación del músculo trapecio.

El uso de bioseñales permitirá obtener realimentación en tiempo real sobre el estado psico-fisiológico de los/as pacientes. En primer lugar, esto, se utilizará como ayuda a la terapeuta para ajustar el grado de exposición en función del nivel de ansiedad de la persona participante. Por ejemplo, para cambiar entre una sesión de relajación y escenarios que causan ansiedad, determinar si es necesario parar con la terapia o añadir

complejidad a las escenas de RV de manera que sean lo más efectivas posibles para cada persona. Por otro lado, esta realimentación ayuda a comprender qué ocurre en el cuerpo de cada persona ante la situación que genera miedo y poder trabajar en cambiar la reacción para tener mayor autocontrol.

Otro uso de los biomarcadores digitales será el de evaluar el progreso de los/as pacientes. La monitorización de las sesiones permitirá evaluar el nivel de ansiedad para comparar entre sesiones o analizar el progreso a lo largo de una misma sesión.

Los datos obtenidos también se podrán utilizar para comprender su utilización para entrenar algoritmos de aprendizaje automático y para detectar estados de ansiedad y relajación. El objetivo de esto es avanzar en una propuesta de un proceso de terapia (semi-)automático que sería capaz de ajustar los estímulos (con la aprobación de un terapeuta) o cambios en las escenas al estado emocional de la persona para mejorar las sesiones de terapia.

## 2.3.4 Aplicación de control para terapeuta

Combinar la técnica de DS con RV facilita su implementación por parte de la terapeuta. Por ejemplo, ayuda a seguir adecuadamente la jerarquía de ítems elaborados con el/la paciente. Cuando la técnica DS se aplica solo con imaginación, el método tradicional, el/la terapeuta no tiene control sobre lo que la persona está imaginando, mientras que de esta forma, controla lo que está visualizando en cada momento.

Por otro lado, la tecnología puede ayudar también en la generación de la respuesta incompatible con la ansiedad al introducir escenarios relajantes en RV que induzcan dicha respuesta y que, a posteriori, la persona pueda recordar. Además, la terapeuta puede introducir audios que ayudan y facilitan el proceso de la relajación.

Para gestionar los estímulos que se presentan y monitorizar la sesión, se crea la aplicación de control para el/la terapeuta. Para su desarrollo, se han identificado las funcionalidades que debe tener:

- Control sobre el lanzamiento de la aplicación. Al lanzarla, se guarda el “ID” (identificador) del paciente junto con el número de sesión y en qué mano lleva la pulsera.
- Control sobre la posición del HMD.
- Control sobre el lanzamiento de cada una de las 4 **fases**, (con independencia absoluta, puede saltar cualquiera de ellas/parar/relanzar, etc.).
- Control sobre los audios a reproducir (música relajante, que puede reproducirse de forma paralela a los audios con instrucciones y audios con instrucciones finales).
- Control sobre escenarios de **relajación** a lanzar.
- Control sobre escenarios de **escaleras** a lanzar.

- Control sobre lo que está visualizando la persona usuaria.

Adicionalmente, esta aplicación permite **personalizar las variables** de las escenas de RV, como el movimiento de las escaleras mecánicas o el sonido (la presencia de viento o barullo). Esto permite adaptar las sesiones a la jerarquía de ansiedad que ha sido elaborada al inicio de la terapia con cada participante.

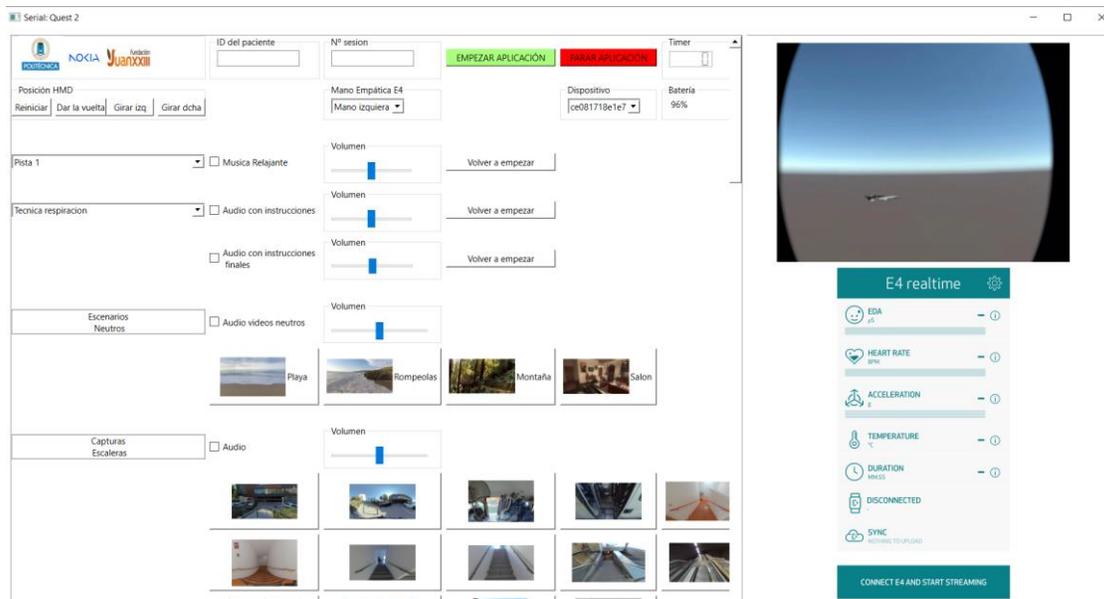


Figura 6: Ejemplo de la interfaz de la aplicación de control para la terapeuta encargada de las sesiones de terapia

### 2.3.5 Necesidades de conectividad

En este caso de uso, terapeuta y paciente se encuentran en la misma sala. La movilidad, por tanto, es limitada. Tanto el HMD como los biosensores utilizados en la sesión deben ser inalámbricos y lo menos invasivos posible. Los algoritmos, como por ejemplo, de segmentación del cuerpo debe poder ejecutarse en el borde de la red (*Edge*), por lo que se precisa de un enlace de subida de capacidad alta y con baja latencia.

### 2.4 Evaluación de la interacción y QoE

Una herramienta muy importante a la hora de evaluar nuevos entornos inmersivos es la Calidad de Experiencia (QoE). En concreto, la evaluación de la QoE nos facilita información sobre cómo fue la experiencia de una persona usuaria según distintos factores influyentes. Entre los factores más importantes en tecnologías inmersivas encontramos la calidad global percibida [20], la presencia espacial [21] y la presencia social [22]. La calidad global percibida es una opinión general sobre la experiencia completa y con este

factor obtenemos información sobre el grado de satisfacción de las personas usuarias. La evaluación de este factor es común en sistemas tradicionales [20] . En tecnologías inmersivas, sin embargo, entra en juego la sensación de estar en otros lugares (sensación de presencia) [23] . Esto es debido a que gracias a los dispositivos inmersivos como los HMDs podemos aislar la percepción de las personas usuarias y transportarlas a otro lugar. En este sentido, la presencia espacial mide la plausibilidad con la que estas personas sienten estar en otro lugar [24] .

Existen distintos cuestionarios que se han utilizado para medir la presencia espacial [25] [26] . Sin embargo, recientemente los sistemas de XR actuales incluyen una nueva variedad de métodos de interacción que influyen en la sensación de presencia. Por ejemplo, mediante el seguimiento de manos o handtracking las personas usuarias pueden interactuar con los elementos de la RV sin utilizar controladores [27] . Por último, técnicas avanzadas de segmentación, como ya se ha comentado previamente, permiten a la persona ver su propio cuerpo en la escena RV [28] , [29] . Todas estas nuevas posibilidades de interacción deben ser tenidas en cuenta a la hora de evaluar la sensación de presencia y la QoE en general. Para evaluar el impacto de este tipo de tecnologías se propone partir de cuestionarios de evaluación subjetiva ya validados en otros estudios en los que se evalúa el nivel de presencia espacial y autónoma alcanzados teniendo en cuenta cómo la persona usuaria interactúa en el mundo virtual [30] [31] [32] .

Algunos de los factores de la presencia espacial y autónoma que se evalúan en estos cuestionarios son: inmersión, adaptación al entorno RV, sensación háptica, calidad global o mareo. Mediante el uso de estos cuestionarios se puede evaluar desde distintos aspectos cómo de completa ha sido la experiencia y si el sistema cumple con los requisitos de inmersividad o presencia esperados. Para la utilización de estos cuestionarios de evaluación subjetiva con personas de la Fundación Juan XXIII se trabajará en hacerlos accesibles para que sean comprensibles y puedan utilizarse.

## 3 Telepresencia

### 3.1 Descripción

Este caso de uso se basa en el sistema de comunicaciones inmersivas desarrollado por Nokia, Owl o Snowl. Se basa en una comunicación basada en vídeo 360 grados. El vídeo se captura con una cámara de 360 grados, con al menos dos lentes de gran apertura angular, y se transmite en tiempo real. La peculiaridad de Snowl es que la persona que está en remoto atiende la conversación a través de unas gafas de RV. Los cascos de realidad virtual permiten la visualización de una parte de la escena, *viewport*, que varía con los movimientos de la cabeza, tiene tres grados de libertad (DoF, del inglés *Degrees of Freedom*). Los principales beneficios están relacionados con la inmersión que siente la persona en remoto, ya que puede tener una experiencia mucho más parecida a si estuviera en persona en el escenario en el que está situada la cámara de 360 grados.

Su aplicación tiene lugar en el piso de entrenamiento y en los pisos tutelados de la Fundación Juan XXIII Roncalli. En el piso de entrenamiento las personas que asisten a la Fundación Juan XXIII Roncalli entrenan para una vida independiente y autónoma. En los pisos tutelados, 3-4 personas conviven de manera independiente contando con el apoyo de profesionales. Cada piso tutelado y las personas del mismo tienen unas necesidades concretas. Por tanto, este caso de uso se adapta en base a las necesidades y requisitos que presenten las personas convivientes y las personas profesionales de apoyo.

Teniendo este contexto de Snowl y los pisos tutelados, el escenario básico considera al profesional en remoto con las gafas de RV y el dispositivo Snowl en el piso tutelado. De esta manera se investiga la tecnología en comunicaciones inmersivas, desde la calidad de vídeo necesaria para que el profesional en remoto pueda ejercer la supervisión con seguridad hasta herramientas adicionales que puedan mejorar la interacción y el sistema. Todas las implementaciones y modificaciones del prototipo Snowl que se añadan de manera gradual se validan este caso de uso real. Nótese que la supervisión en remoto, dado que el sistema está en pruebas, se va a realizar garantizando la seguridad de las personas convivientes en el piso tutelado y el/la profesional estará en una habitación del mismo piso para llegar a tiempo en momento de cualquier altercado que pudiera suceder.

Para planificar la supervisión, se definen una serie de tareas cotidianas, algunas divididas en subtareas, que las personas de los pisos tutelados deben practicar. Con esto lo que se consigue es poder llevar una evaluación de cómo se avanza en los pasos hasta completar la tarea, los problemas que surgen y si hay variaciones en el nivel de soporte requerido a lo largo de la práctica. Estas actividades se realizan en grupo o de manera individual dependiendo de las características de las personas, por ejemplo, la experiencia y dinámicas de los pisos tutelados. Concretamente, se definen algunas de las tareas que

se practican y que podrían realizarse bajo la supervisión del profesional de apoyo en remoto:

- Preparar el menú semanal
- Poner la lavadora.
- Tender la ropa.
- Cocinar platos sencillos.

### 3.1.1 Necesidades que resuelve y beneficios esperados

La Fundación Juan XXIII Roncalli dispone de cuatro pisos tutelados dirigidos a personas con discapacidad intelectual que presentan cierto grado de autonomía pero que, requieren apoyo para llevar a cabo las diferentes actividades de la vida cotidiana.

Los apoyos se distribuyen en las viviendas de la siguiente manera:

- Piso BERLÍN: lunes a viernes. Acompañados por las tardes y monitor por las noches (compartido con piso Kenia).
- Piso KENIA: lunes a domingo. Acompañados por las tardes y monitor por las noches (compartido con piso Berlín).
- Piso HAWAI: uso de lunes a viernes con apoyo continuo.
- Piso BOMBAY: usuarios con apoyos puntuales (2-3 h/día) y fines de semana.

Estos pisos son espacios diseñados para ofrecer un contexto en el que las personas residentes puedan adquirir autonomía y aprender a relacionarse con el entorno comunitario de una forma segura. Para ello, los/as educadores/as se encargan de proporcionar ayuda para el desempeño de las diferentes actividades del día a día y fomentar un clima de convivencia y cooperación.

Es importante tener en cuenta que no todos las personas residentes tienen las mismas necesidades, por lo que se trata de ofrecer un apoyo personalizado en función de las capacidades, preferencias y rutinas de cada participante. A pesar de estas diferencias individuales, un día habitual en el piso tutelado suele incluir los siguientes tipos de actividades:

- Rutina matutina: incluye tareas de aseo, preparación del desayuno y planificación del día.
- Actividades diarias: pueden incluir actividades educativas, recreativas y de habilidades sociales. Algunos ejemplos son aprender a manejar el dinero y/o utilizar el transporte público.
- Apoyo en tareas del hogar: se les proporciona ayuda en tareas cotidianas tales como la limpieza, la colada, la planificación del menú, la elaboración de la lista de

la compra, la preparación de recetas sencillas, etc.

- Apoyo a la salud: implica ayuda en tareas de control de la medicación y/o acompañamiento en las visitas a los profesionales de salud y revisiones médicas.
- Tiempo libre y actividades recreativas: en esta área se persigue fomentar la iniciativa, la participación activa en la toma de decisiones y la capacidad de llegar a acuerdos. Esto puede incluir juegos de mesa, ver películas, ir al cine o al teatro, acudir a algún centro deportivo, etc.
- Apoyo emocional y social: los educadores también realizan un seguimiento individual de los residentes, mediación de conflictos, y fomento de la interacción social para favorecer el bienestar y un buen clima de convivencia.
- Rutina nocturna: incluye actividades como hacer la cena y prepararse para dormir.

La instalación de una supervisión remota a través del sistema Owl o Snowl se espera que pueda proporcionar múltiples beneficios entre los que se encuentran:

- Mayor flexibilidad: los educadores podrán supervisar las tareas que están llevando a cabo los convivientes desde cualquier ubicación física, pudiendo dar soporte a más pisos en el mismo tiempo.
- Intervención inmediata en situaciones de emergencia en las que no hay profesional en el piso: la accesibilidad en tiempo real a través del dispositivo permite a los educadores responder ante situaciones imprevistas que puedan producirse en el piso. De esta forma, podrán proporcionar ayuda a las personas residentes sin que tengan que esperar a que la persona profesional de apoyo pueda acceder a la vivienda. Esta posibilidad de actuar rápidamente contribuye a garantizar la seguridad y el bienestar de los convivientes.
- Reducción de costes: la supervisión en remoto permite reducir recursos personales y económicos al ahorrar desplazamientos a las personas profesionales.
- Mayor autonomía para los residentes: se espera que este sistema permita fomentar la seguridad y confianza en sí mismos al tratarse de una supervisión menos intrusiva. El hecho de que el profesional no esté presente físicamente contribuye a la promoción de la independencia y el desarrollo de habilidades de autogestión.

En definitiva, se espera que el sistema de comunicaciones inmersivas ofrezca una serie de ventajas significativas que mejoren la calidad de vida de los residentes y la eficiencia en la atención y apoyos que los profesionales proporcionan.

## 3.2 Implicaciones tecnológicas

En este caso de uso las implicaciones tecnológicas pueden dividirse en tres retos principales: conseguir una calidad perceptual suficiente del sistema de telepresencia para que la persona de apoyo pueda desempeñar sus funciones exitosamente, Owl o Snowl debe ser un sistema accesible para el uso de las personas de los pisos tutelados y se debe tener en cuenta la representación del profesional de apoyo en el entorno local, puesto que va a ser con lo que interactúen las personas convivientes.

### 3.2.1 Calidad perceptual del sistema de telepresencia

La sesión de telepresencia debe mantener suficiente calidad perceptual para la ejecución del caso de uso: calidad visual, calidad auditiva y retardo en la comunicación.

La calidad visual viene determinada principalmente por la resolución del sensor de captura, así como la tasa binaria con la que se codifica y transmite el vídeo inmersivo [33]. Por un lado, la resolución debe ser suficiente como para ejercer una supervisión adecuada y observar los detalles de la escena. Por otro, el sistema debe ser capaz de funcionar en distintas condiciones de saturación de la red y transmitir incluso a tasas binarias bajas. Finalmente, la sensación de presencia espacial que proporciona el vídeo inmersivo puede compensar las limitaciones de calidad visual [32]. Se deberá estudiar el compromiso entre estos tres factores para determinar en qué condiciones es posible desarrollar el caso de uso de manera adecuada.

La calidad auditiva debe ser suficiente para mantener la conversación en todo momento. En particular, es importante mantener la continuidad de la señal de voz, para que no haya interrupción en la comunicación [34] [35].

Finalmente, el retardo debe ser suficientemente bajo como para permitir una comunicación fluida entre la persona de apoyo y las personas usuarias del piso tutelado [29] [31].

### 3.2.2 Accesibilidad de Snowl

Un aspecto básico del dispositivo de telepresencia Snowl es que sea sencillo de manejar por las personas que viven en el piso tutelado. Es necesario que se sientan cómodas con el dispositivo y puedan utilizarlo con facilidad. Algunas operaciones básicas que se deberán realizar con él son:

- Desplazarlo por el piso, para lo que sería conveniente disponer de un trípode con ruedas o similar.
- Iniciar y terminar las llamadas con facilidad, por ejemplo: una interfaz gráfico sencillo o el uso de un botón.
- Mantenimiento básico: reiniciar el sistema en caso de error, cargar las baterías...

Todas estas operaciones se deberán realizar de forma rutinaria y repetitiva por lo que, además de ser accesible, el sistema debe ser robusto.

### 3.2.3 Representación del profesional de apoyo remoto (“presencia inversa”)

Una característica de Snowl es que provee una comunicación altamente asimétrica. Mientras que las personas conectadas en remoto con las gafas de RV pueden visualizar toda la escena, el vídeo 360 grados capturado por la cámara, las personas que se encuentran alrededor de la misma no pueden ver a la persona remoto. Esto, por otro lado, tampoco tendría sentido ya que las gafas de RV le cubren una gran área de la cara. Como solución alternativa, se utilizan avatares que representan a la persona conectada en remoto. A continuación, se presentan algunos ejemplos de avatares que pueden utilizarse, siendo de izquierda a derecha menos realistas o fieles a la persona conectada en remoto con las gafas de RV [36]

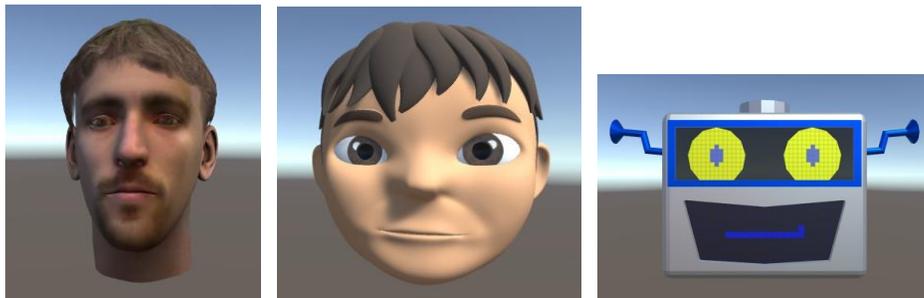


Figura 7: Ejemplos de avatares con mayor a menor realismo de izquierda a derecha

El avatar o la representación que se utilice influirá en la presencia inversa, término utilizado para referirnos a la sensación de presencia social percibida por las personas alrededor de la cámara 360 grados. Concretamente, la sensación de que las personas en remoto, están mirando a las locales e interactuando con ellas.

Adicionalmente, a los avatares se les pueden añadir elementos como movimiento de los labios sincronizados con el audio o movimiento de la cabeza sincronizados con los movimientos de la persona en remoto, incrementando el realismo y, previsiblemente, la presencia inversa [37] .

### 3.2.4 Necesidades de conectividad

En el piso tutelado debe desplegarse un dispositivo de telepresencia. Este necesita, como ya se ha comentado, disponer de movilidad dentro del piso, así como de conectividad bidireccional con la red de comunicaciones. Dado que el flujo de datos principal es el vídeo inmersivo que se captura y transmite, es prioritaria la capacidad del enlace ascendente.

El tutor remoto debe poder conectarse desde cualquier lugar, para poder garantizar su disponibilidad en todo momento. Aunque la conexión es bidireccional, en este caso debe garantizarse la capacidad del enlace de bajada.

### 3.3 Evaluación de la interacción y QoE

En términos de evaluación de la interacción, además de lo descrito en la Sección 2.4, en este caso de uso otro factor que toma relevancia es la presencia social. Este factor mide cómo las personas usuarias se sienten en un entorno compartido. En sistemas no inmersivos, este factor ya ha sido evaluado mediante el uso de cuestionarios de evaluación subjetiva [38]. Sin embargo, el uso de tecnologías inmersivas incrementa las posibilidades de interacción en entornos multiusuarios. Por ejemplo, hay entornos inmersivos que permiten a las personas usuarias desplazarse por el entorno virtual, pudiendo interactuar con otras desde distintos puntos del espacio [39]. Otras, permiten que las personas compartan el mismo entorno virtual pudiendo interactuar de forma conjunta sobre él. Estas nuevas formas de interacción deben ser tenidas en cuenta a la hora de evaluar el grado en el que las personas se sienten con otras personas en un espacio compartido y si consiguen que esa interacción sea más parecida a una en persona. Para ello, en este caso de uso en el que se contempla una comunicación entre las personas convivientes del piso tutelado y el profesional de apoyo remoto, se considerará el uso de cuestionarios que permitan la evaluación de la presencia social como los propuestos en [32] [40] [41].

Además de los aspectos relacionados con presencia espacial y social, se evaluarán aspectos relacionados con la calidad de experiencia que pueden verse afectados por factores técnicos del sistema de comunicación. En concreto, se realizarán algunas preguntas acerca de la calidad visual del vídeo visualizado a través del HMD, así como la calidad del audio y el retardo en la comunicación, tanto por parte del profesional de apoyo remoto como por los usuarios en el piso tutelado. Además, también se evaluará la QoE global de los usuarios. Para evaluar estos factores, se utilizarán escalas tipo Likert [42]. Finalmente, también se realizarán preguntas relacionadas con posible mareo o malestar causado durante la experiencia [42]. De la misma manera que en el caso de uso de terapia, los cuestionarios de evaluación subjetiva tendrán que ser accesibles y comprensibles para las personas que los tienen que contestar.

Las evaluaciones se incluirán en el calendario para realizarlos después de la realización de las tareas como las presentadas en la Sección 3.1.

## 4 Teleformación

### 4.1 Descripción

La Fundación Juan XXIII Roncalli cuenta con cursos de formación que otorgan a sus estudiantes certificados de profesionalidad homologados. El acceso a esta formación y al certificado de profesionalidad facilitan la inserción laboral de las personas con discapacidad que lo obtienen. Sin embargo, existen algunas dificultades asociadas al desarrollo de estos cursos. En primer lugar, el número de plazas es limitado, y algunas personas que querrían tener acceso a esta formación no pueden hacerlo. Esto es particularmente importante en personas usuarias del COFOIL, que tienen más dificultades o necesidades de apoyo. En segundo lugar, cuando el alumnado de los cursos no puede asistir presencialmente a las clases, bien por motivos médicos o en épocas vacacionales como el verano, le cuesta recordar las lecciones aprendidas y coger el ritmo para seguir aprendiendo, o incluso pierden una parte significativa del curso y eso les impide obtener el certificado de profesionalidad. El objetivo de este caso de uso es utilizar la tecnología XR para paliar estos problemas.

La sensación de presencia e inmersividad que provee la tecnología XR son factores que motivan su utilización en sesiones de tele-formación. Además, la telepresencia inmersiva permite seguir las lecciones de cocina desde casa, con un nivel de seguimiento y evaluación similar al que se tendría si se accediera de forma presencial. Asimismo, la compatibilidad de la monitorización de los movimientos, dadas por las propias gafas de RV o la interacción con un mundo y objetos virtuales, añade valor a esta investigación. En definitiva, el caso de uso permite explorar las tecnologías XR en una aplicación real con unos objetivos de formación predefinidos y con personas con discapacidad.

Las personas participantes serán del área de inserción socio-laboral de COFOIL, ya que tienen menos oportunidades para acceder a los cursos de formación y precisan mayores de necesidades de apoyo. El objetivo del caso de uso es mejorar sus posibilidades de inserción laboral mediante la combinación de dos experiencias: acceder a cursos de formación profesional y entrenar habilidades cognitivas específicas.

En primer lugar, se desarrollará un programa piloto para poder seguir de forma remota, empleando tecnologías de XR, algunas sesiones del curso de Operaciones básicas de Pastelería (RD 1179/2008, de 11 de julio). Las competencias generales que este certificado aporta a las personas que lo poseen es: preelaborar, elaborar y presentar elaboraciones sencillas de pastelería y asistir siguiendo instrucciones en la preparación de otras más complejas, realizando operaciones básicas de aprovisionamiento interno y aplicando técnicas y normas básicas de manipulación, preparación y conservación de alimentos.

Para complementar el apoyo que estas personas pueden necesitar en su proceso de inserción laboral, se propone trabajar adicionalmente la estimulación cognitiva, enfocada a mejorar las habilidades cognitivas con ejercicios de atención, inhibición y velocidad de procesamiento. Una motivación adicional de la estimulación cognitiva para los/as participantes en el área de inserción laboral es que mejoren los procesos atencionales para favorecer sus capacidades de control de respuesta y el rendimiento y eficiencia en sus puestos de trabajo.

La estimulación cognitiva se define como el conjunto de habilidades y técnicas aplicadas para mejorar las funciones cognitivas de las personas [43]. Está demostrado que realizar este tipo de actividades contribuye a crear nuevas conexiones neuronales gracias a la plasticidad cerebral, entendida como la capacidad del cerebro para modificar su estructura y funcionamiento con la finalidad de poder adaptarse a las demandas del entorno [44]. En el caso de personas con discapacidad intelectual es especialmente importante implementar programas de estimulación cognitiva con el objetivo de fortalecer las capacidades existentes, desarrollar nuevas capacidades y crear estrategias que puedan compensar aquellas que no pueden llegar a desarrollar. El uso de la XR se ha ido incrementando en el campo de la neurorehabilitación y la estimulación cognitiva como alternativa a las actividades clásicas de lápiz y papel [45].

La idea es que las sesiones de estimulación cognitiva tengan un formato de juego en el que hay continuidad, donde los/las participantes vayan ascendiendo de nivel conforme mejoran sus habilidades o retrocediendo en caso de cometer un número de errores. La motivación de esto es incentivarles a superarse mientras que, si no superan un nivel, desciendan uno para evitar frustraciones. La realidad extendida nos proporciona versatilidad para definir una dinámica de actividad similar pero que podemos particularizar con objetos específicos de distintos campos de aplicación, vinculados con sus intereses personales y laborales.

## 4.2 Necesidades que resuelve y beneficios esperados

Entre los beneficios de utilizar la RV para llevar a cabo programas de estimulación cognitiva frente a los métodos tradicionales se encuentran los siguientes: proporciona realimentación en tiempo real [46] permitiendo a las personas tener información sobre su rendimiento de forma inmediata. Por otro lado, fomenta un aprendizaje activo e incrementa la motivación de las personas usuarias [47]. Además, permite incrementar gradualmente el nivel de dificultad de las tareas, lo que facilita adaptarla a las características de la persona. El beneficio principal que se espera encontrar después de esta intervención es que mejore su capacidad atencional, se reduzca su velocidad de procesamiento y aumente su capacidad de inhibición de respuesta.

Otra de las ventajas que puede proporcionar la RX frente a los programas de estimulación cognitiva clásica es que las actividades presentan una mayor accesibilidad al no requerir de habilidades de lectoescritura como suelen demandar las clásicas actividades de lápiz y papel. Por tanto, todas las personas podrán participar en las actividades independientemente de su nivel educativo y competencia lectoescritora.

Además, las actividades que se plantean requieren que la persona se mueva por el espacio durante su realización. Esto no solo contribuye a trabajar aspectos cognitivos como la coordinación, orientación y praxias, sino que ayuda a consolidar sus aprendizajes de manera más efectiva [48].

Por último, se ha encontrado que los programas de realidad virtual facilitan la transferencia de las habilidades adquiridas al entorno natural en el que se desenvuelve la persona [49]. Por tanto, se espera que los objetivos conseguidos con este programa se puedan generalizar en su día a día y, en consecuencia, repercuta de forma positiva en sus posibilidades de inserción socio-laboral.

Como necesidad principal que resuelve se espera que el curso de teleformación en cocina proporcione a las personas de COFOIL la oportunidad de realizar una formación a la que normalmente tienen dificultades de acceso. Asimismo, se incrementan sus oportunidades de formación y se espera que la particularidad de realizarlo a través de la tecnología tenga los siguientes beneficios:

- Favorecer un aprendizaje práctico: la RV permite a las personas con discapacidad intelectual practicar habilidades culinarias en un entorno seguro y controlado.
- Promover la participación activa y la motivación: la posibilidad de realizar la formación incorporando la RV puede resultar especialmente atractivo y motivador para los participantes, ya que la integración de elementos visuales, auditivos y táctiles pueden estimular su participación y compromiso con el contenido del curso.
- Facilitar la adaptabilidad y personalización: la RV permite adaptar y personalizar la formación teniendo en cuenta las necesidades individuales de los participantes. Esto permite ajustar el ritmo de aprendizaje, los niveles de dificultad y los contenidos del curso de tal forma que, se garantice una experiencia de aprendizaje significativa y efectiva.
- Mejorar las habilidades de autonomía: a través de la formación, las personas pueden aprender y practicar habilidades de cocina, lo que puede ser una forma efectiva de fomentar su autonomía, favorecer su independencia en la vida diaria e incrementar sus posibilidades de acceso a otras formaciones y/o laborales.

## 4.3 Implicaciones tecnológicas

### 4.3.1 Interacción persona-máquina (HCI)

La teleformación en operaciones básicas de pastelería combina algunos de los aspectos tecnológicos desarrollados en los casos de uso anteriores: telepresencia, reproducción de contenidos inmersivos, visualización del propio cuerpo hasta interacción natural con objetos.

Las lecciones del curso de cocina son grabadas en vídeo 360 grados en las inmediaciones de la Fundación Juan XXIII Roncalli. En cuanto a la visualización de las clases, esta vez es el alumnado el que lleva las gafas de RV. Durante estas visualizaciones el alumnado podrá verse las manos y ciertas zonas de interés del entorno físico que sean útiles para afianzar los conceptos que se trabajan en el vídeo. Una vez vistas las lecciones, el alumnado debe ejecutarlas. Aquí es cuando recibe apoyo por parte de la profesora, quien de manera remota por medio de las gafas de RV sigue la ejecución y resuelve dudas. Es importante garantizar baja latencia para gestionar y anticiparse a cualquier problema que pueda surgir.

Se propone complementar en paralelo esta formación con sesiones de terapia cognitiva que involucran juegos diseñados para entrenar sus habilidades cognitivas. Los juegos están inspirados en algunas actividades definidas en las técnicas de estimulación cognitiva clásicas. A partir de escenarios básicos, el juego inmersivo compuesto por múltiples niveles, podrá utilizarse hasta un máximo de 12 horas. Estos juegos están diseñados específicamente para entrenar la atención selectiva, sostenida y alternante, la velocidad de procesamiento y la inhibición, áreas en las que los sujetos tienen dificultades. De esta forma, las actividades clásicas se trasladan a un entorno de XR, haciéndolas así inmersivas y requiriendo que las personas se muevan e interactúen con el mundo virtual para resolver los niveles. Es posible que la RV permita a los participantes concentrarse mejor en las actividades y evitar distracciones externas gracias al uso del HMD [50]. Además, la experiencia que proporciona es multisensorial, lo que puede favorecer el aprendizaje [51].

El juego se diseña para seguir la progresión de cada sujeto, volviéndose más difícil o fácil según su nivel. Se sube de nivel al completar un número de tareas exitosas consecutivas, mientras que después de un número diferente de tareas fallidas consecutivas, el nivel disminuye. Adicionalmente, las actividades se modifican para adaptarse a los intereses laborales específicos de las personas de COFOIL participantes en las sesiones. Algunos escenarios de ejemplo podrían incluir: limpieza, servicio de catering, supermercados, puericultura, animales, tiendas de ropa.

La dinámica de juego y la XR permiten que la persona que está participando se haga responsable de su propio progreso al recibir una realimentación inmediata, aumentando

su motivación y autonomía. Durante la sesión, siempre guiada por la terapeuta, se mide la respuesta de los/as participantes y así se puede supervisar y monitorizar su progreso.

## 4.3.2 Captura de bioseñales y regulación emocio-cognitiva

Al igual que en el caso de uso de terapia, para abordar la regulación emocio-cognitiva de las personas usuarias contemplada en las sesiones de terapia cognitiva, se lleva a cabo la captura de las siguientes bioseñales mediante el HMD y los sensores utilizados: movimientos de cabeza, pulso de volumen sanguíneo, temperatura cutánea, respuesta galvánica de la piel y electromiografía.

En este caso, también se utiliza como biomarcador el movimiento ocular. En este sentido, las características oculares más importantes que permiten obtener información sobre los estados emocionales y cognitivos de las personas usuarias son el diámetro pupilar y la frecuencia de parpadeo, aunque el patrón general de movimiento ocular (en términos de fijaciones y sacadas) sirve como señal complementaria [52]. Al igual que en el caso de terapia, la captura de estas bioseñales permite monitorizar en tiempo real el estado emocio-cognitivo de las personas, siempre bajo la supervisión de la terapeuta.

Finalmente, las medidas capturadas también son datos muy útiles para continuar el desarrollo de algoritmos de aprendizaje automático que permitan detectar estados emocio-cognitivos. En este sentido, se parte de los desarrollos llevados a cabo en el caso de uso de terapia para ser analizados en este caso de uso. Los resultados de estas pruebas permiten conocer cómo utilizar las señales capturadas en este caso de uso para realizar los ajustes necesarios sobre los mismos, re-entrenarlos y validarlos.

## 4.3.3 Necesidades de conectividad

Este caso de uso requiere de conectividad tanto en el centro de formación como en el lugar en el que se encuentra el alumnado. En el centro de formación es necesario disponer de dispositivos de telepresencia para poder capturar y transmitir las clases y conectarse para su visualización. También es preciso que la profesora disponga de dispositivos de realidad virtual (HMDs). Este caso de uso requiere de suficiente capacidad de transmisión tanto en el enlace ascendente como el descendente.

Para el seguimiento de las clases, así como del entrenamiento específico en atención, se debe disponer de un HMD y biosensores (solo para la terapia cognitiva), todo ello inalámbrico y con conectividad suficiente con el borde de la red.

## 4.4 Evaluación de la interacción y QoE

Para el escenario de estimulación, las personas usuarias deben interactuar con un entorno virtual para desarrollar ejercicios que mejoren la actividad cognitiva. En dichos

ejercicios, las personas realizan tareas de búsqueda y selección de elementos. En la tarea de búsqueda precisan utilizar el equipo de realidad virtual para localizar distintos elementos en el entorno.

Para la tarea de selección en realidad virtual, se plantean dos tipos de interacción, basadas en controladores o basadas en seguimiento de las manos. El uso de los controladores está muy extendido y se caracteriza por su precisión. Estos controladores permiten funcionalidades de distintos niveles y, según la complejidad, requieren de un proceso de aprendizaje diferente [53]. El seguimiento de manos se concibe como una interfaz natural de interacción, lo que significa que imita las interacciones en el mundo físico [27] . Sin embargo, se pierde precisión. Para no causar frustración o problemas durante el juego, se decide utilizar los controladores en funcionalidades sencillas, como por ejemplo, coger un objeto virtual presionando un solo botón.

Para la evaluación de la interacción se propone recoger las siguientes variables:

- Número de aciertos
- Número de errores
- Número de omisiones
- Tiempo de respuesta

Y, como bioseñales:

- Movimiento de la cabeza (3DoF)
- Movimiento ocular y medida pupilar
- Respuesta galvánica de piel
- Frecuencia cardíaca
- Activación del musculo trapecio

Para llevar a cabo la evaluación de interacción y QoE en el escenario de teleformación en cocina, se parte de los mismos procedimientos y cuestionarios descritos en el caso de uso de telepresencia.

## 5 Definición de requisitos de realidad extendida e interacción

### 5.1 Introducción a los sistemas de realidad extendida

Se espera que en los próximos años los sistemas de XR se utilicen en mayor medida, de una manera similar a los sistemas convencionales que utilizamos en la actualidad. XR es un término paraguas que se utiliza para nombrar a aquellas tecnologías que ofrecen una visión alternativa de la realidad [54]. La amplia variedad y combinación de plataformas puede interpretarse sobre una escala continua, denominada Realidad-Virtualidad, que abarca desde un mundo real hasta un mundo puramente sintético. Siguiendo la figura presentada a continuación, la XR puede ser una combinación de Realidad Aumentada (AR), Virtualidad Aumentada (AV) y/o Realidad Virtual (RV). La AR permite la integración de objetos virtuales en la realidad. La AV permite la integración de objetos reales en el entorno virtual. La RV crea un entorno inmersivo alrededor de los usuarios con los cascos de realidad virtual, permitiendo la interacción con los objetos que contiene.



Figura 8: Reality–Virtuality Continuum [56]

Como objetivo común de los sistemas XR, es que facilitan la comunicación, interacción con objetos reales y virtuales, así como, la colaboración entre personas. Las experiencias creadas por los sistemas de XR están en auge y su alcance es tal que encuentran campos de aplicación desde la educación, entretenimiento, asistencia o industria. Adicionalmente, en las últimas décadas las nuevas tecnologías han aparecido como herramienta para complementar este tipo de programas y han contribuido a mejorar la inclusión de las personas con discapacidad intelectual [55].

### 5.2 Calidad de Experiencia en realidad extendida

Desde un punto de vista técnico, algunos de los desafíos que presenta la evaluación de QoE en XR son comunes a las tecnologías anteriores y están relacionados con las restricciones de ancho de banda, ya que afecta la calidad de los contenidos y flujos transmitidos en tiempo real. En consecuencia, estas restricciones afectan a la QoE, cuyos factores influyentes se dividen en tres categorías: humana (visión y audición,

malestar/mareo, inmersión y expectativas y experiencia), sistema (contenido, red condiciones y parámetros de codificación) y contexto (contexto físico, temporal, social y de tarea).

Con la evolución de la tecnología, la evaluación subjetiva de la QoE ha sido abordada por las recomendaciones de la Unión Internacional de Telecomunicaciones (UIT), desde la ITU-T P.800 (voz), P.913 (video 2D), ITU-R BT.500 (televisión) hasta la ITU-T P.919 (vídeo de 360 grados). Estas recomendaciones definen un procedimiento detallado para ejecutar la prueba subjetiva, incluyendo la caracterización del contenido fuente y las degradaciones, la estructura de las sesiones de prueba, escalas de evaluación o métodos de análisis estadístico de los resultados. Este nivel de detalle proporciona un marco común para realizar experimentos de una manera similar, necesario para comparar resultados entre experimentos. Generalmente, estas recomendaciones son seguidas por la comunidad investigadora que han estudiado ingeniería y trabajan de ello. La principal limitación es que cada recomendación se adapta a condiciones de prueba específicas, centradas principalmente en la evaluación de la calidad audiovisual perceptiva, y la generalización fuera de ellas no es sencilla.

En XR, sin embargo, la calidad audiovisual no se puede analizar de forma aislada. Actualmente, la calidad visual lograda por los sistemas de comunicación de vídeo de 360 grados es de hecho inferior a la de la transmisión de vídeo convencional o a otras experiencias de RV. Sin embargo, la posibilidad de proveer una comunicación entre personas en tiempo real en un entorno inmersivo hace que la tecnología XR siga teniendo un gran potencial. En concreto, este potencial está relacionado con los beneficios que se ofrecen en términos de aspectos humanos y de contexto. La evaluación de la QoE en XR bajo restricciones de comunicación deberá considerar no solo propiedades perceptuales, como la calidad del vídeo, sino también aspectos socioemocionales relacionados con los efectos psicológicos inducidos en los usuarios (por ejemplo, la presencia espacial o social). Aunque el UIT-T está introduciendo estos aspectos en su planificación (por ejemplo, en la Rec. P.1320 "Evaluación de la QoE de las reuniones de realidad extendida (XR)"), todavía es un trabajo en vías de desarrollo.

## 5.3 Estado del arte de algoritmos de realidad extendida e interacción persona-máquina

### 5.3.1 Tecnologías de presencia autónoma

Tradicionalmente, las personas usuarias han sido representados por avatares generados por ordenador que pueden variar desde manos en 3D, dibujos animados o humanos hasta representaciones más realistas y personalizadas [57]. También existe una nueva tendencia a utilizar los propios cuerpos de los/as usuarios/as mediante captura y segmentación de vídeo egocéntrico como su auto-avatar, ya que ahorran la necesidad

de costosos dispositivos de seguimiento o escaneos corporales para proporcionar una representación realista de la persona. Esto se ha hecho en la literatura mediante el uso de información de profundidad [58], [59], color [60], [61] o algoritmos de aprendizaje profundo [62].

La Tabla 1 enumera, en orden cronológico, estudios de la literatura relacionados con el impacto de diferentes formas de representación de avatares y realismo los mismos en factores como la representación de uno/a mismo/a. En general, todos los estudios indican el número de participantes y en la mayoría de los casos el tamaño de la muestra es suficiente para extraer conclusiones estadísticamente significativas. Si bien esto no se tuvo en cuenta desde los primeros estudios, donde en ocasiones ni siquiera se reportaba información personal como el género o la edad, parece que actualmente la comunidad investigadora está teniéndolo en cuenta en sus estudios con usuarios/as. Sin embargo, en muchos casos todavía falta una representación diversa de la sociedad. Otros aspectos que pueden afectar las evaluaciones y la implicación con la tecnología, como la experiencia con XR, las discapacidades intelectuales o el nivel de educación, no se recopilan ni reportan.

Tabla 1: Estado del arte de estudios relacionados con la representación de avatares

Work	Year	Participants	Gender	Age Group	Skin Color	Other Factors
Garau <i>et al.</i> [6]	2003	48	24 female, 24 male	university students	<i>n/a</i>	<i>n/a</i>
Lok <i>et al.</i> [13]	2003	40	3 female, 33 male	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>
Ries <i>et al.</i> [23]	2008	6	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	university students
Mohler <i>et al.</i> [15]	2010	48	<i>n/a</i>	19 – 43 (M=29.9)	<i>n/a</i>	university community
Gutekunst <i>et al.</i> [10]	2014	16	8 female, 8 male	(M=29.1)	<i>n/a</i>	<i>n/a</i>
Lugrin <i>et al.</i> [14]	2015	30	8 female, 22 male	(M=21.60, SD=2.43)	<i>n/a</i>	<i>n/a</i>
Waltemate <i>et al.</i> [27]	2018	29	15 female, 14 male	19 – 33 (M=24)	<i>n/a</i>	<i>n/a</i>
Ogawa <i>et al.</i> [18]	2018	39	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>
Wang <i>et al.</i> [28]	2019	10	1 female, 9 male	(M=23.6, SD=1.26)	<i>n/a</i>	<i>n/a</i>
Yoon <i>et al.</i> Study 1 [30]	2019	24	12 female, 12 male	18 – 41 (M=27.13, SD=4.74)	<i>n/a</i>	<i>n/a</i>
Yoon <i>et al.</i> Study 2 [30]	2019	24	12 female, 12 male	18 – 31 (M=22, SD=3.19)	<i>n/a</i>	<i>n/a</i>
Ogawa <i>et al.</i> [18]	2020	17	3 female, 14 male	(M=24.53, SD=4.42)	<i>n/a</i>	height (M=169.94,SD=7.77 cm)
Fribourg <i>et al.</i> [5]	2020	40	20 female, 20 male	18 – 31 (M=32.5, SD=10.1)	<i>n/a</i>	<i>n/a</i>
Previous Study [4] - University	2023	54	26 female, 28 male	18 – 29 (M=20.28, SD=2.74)	1 black skin	university population
This Study - Local Population	2024	34	16 female, 17 male	20 – 66 (M=35.40, SD=12.91)	6 black skin color	local population

### 5.3.2 Tecnologías de interacción natural

Una de las ventajas de utilizar tecnologías inmersivas es la transformación de las formas de interacción. En concreto, la XR nos permite interactuar con nuestro entorno físico mezclado con el mundo virtual. Uno de los ejemplos más sencillos es la navegación. Habitualmente, la navegación en mundos virtuales se realizaba mediante teclado y ratón. En cambio, la navegación inmersiva se realiza mediante desplazamientos en el mundo físico, que luego se transfiere al mundo virtual. Este tipo de interfaces que mapean las interacciones tal y como suceden en el mundo físico y las trasladan al mundo virtual se denominan Interfaces Naturales de Usuario (NUI).

Entre las tecnologías existentes para generar interacción natural de personas usuarias en XR están: comandos de voz para interactuar con agentes virtuales [39], seguimiento de manos para pulsar botones [28], manipulación de objetos mediante segmentación

egocéntrica [29] o el uso de la posición del cuerpo para visualizar distintos puntos de vista en una teleconferencia XR [40] . Gracias a este tipo de interfaces las personas usuarias pueden realizar acciones en el mundo XR sin necesidad de un aprendizaje extra, lo que supondría una barrera a la hora de introducirse en esta tecnología. Además, si existen requisitos que demanden las manos y la manipulación de objetos, hacen imposible el uso de controladores externos, dejando como única opción la interacción natural.

En el caso concreto de este proyecto, por ejemplo, para caso de uso de terapia mediante desensibilización sistemática se requiere que las personas puedan verse los pies para poder enfrentarse a los estímulos ansiógenos (las escaleras). Por lo tanto, es requisito que puedan interactuar con el entorno de XR de forma natural, es decir, utilizando su cuerpo.

### 5.3.3 Tecnologías de telepresencia

El término de telepresencia se utiliza para denominar a aquellas tecnologías que permiten hacer sentir a la persona que está en remoto presente en el entorno en el que transcurre la reunión o evento. Las tecnologías de telepresencia emergentes, como los telerobots fijos y móviles (por ejemplo, robots de telepresencia, brazos robóticos y hologramas) tienen el potencial de aliviar el problema de la interacción social de las personas que no puedan asistir a eventos presencialmente, atenuando la limitación de inaccesibilidad y creando un ambiente más inclusivo [63], [64].

Algunos sistemas permiten a los usuarios que se encuentran a distancia explorar el entorno físico del evento y desplazarse alrededor de objetos para verlos desde distintos ángulos, teniendo una interacción más realista [65]. Algunas de las limitaciones de estos sistemas móviles es la manera en la que se produce el desplazamiento, no siendo siempre útil para los objetivos que se plantean ni accesible para las personas que hacen uso de ello.

Otros sistemas de telepresencia, como el Owl o Snowl, se basan en la captura y transmisión de vídeo 360 grados, es decir, toda la escena que se tiene alrededor. En este caso la telepresencia se consigue creando el efecto de que la persona que visualiza la escena 360 se ha teletransportado a ese entorno físico en el que está colocado el sistema. Aquí, en cambio, la persona no puede moverse con 6DoF, es decir, no hay profundidad pero, sin embargo, sí tiene 3DoF con los que la parte de la escena que visualiza cambia con los movimientos de rotación de su cabeza. Estas tecnologías permiten el intercambio de señales no verbales, como el lenguaje corporal o la mirada, entre participantes que influyen en gran medida en la efectividad de las comunicaciones en persona [66]. Adicionalmente, estudios de la literatura afirman que los usuarios/as

que se conectan en remoto pueden aumentar la sensación de presencia espacial, impactando positivamente en la empatía y atención [67], [68], [32] .

## 5.3.4 Tecnologías de biomarcadores digitales y regulación emocional

Como se mencionó previamente, varios estudios han demostrado cómo las señales biológicas llevan información útil para comprender el estado emocio-cognitivo [19] , [70], [69]. La información contenida en estas señales permite el reconocimiento de estados emocionales, de estrés y de relajación. En este contexto, algunas señales de particular relevancia son: Actividad ElectroDérmica (EDA), frecuencia cardíaca (HR), diámetro pupilar y activación de músculos [19] . En el pasado la captura de estas señales implicaba instrumentación invasiva, y limitaba los movimientos, hoy hay muchos sensores no invasivos y portátiles diseñados para medir señales fisiológicas. La disponibilidad y accesibilidad de estas tecnologías abren las puertas a su uso en aplicaciones innovadoras, siendo de particular relevancia las posibles aplicaciones en el ámbito médico.

En el contexto de este proyecto es de fundamental importancia la no invasividad de los sensores. Las pulseras inteligentes permiten recopilar señales con dispositivos con forma de reloj totalmente no invasivos. Ejemplos de este tipo de tecnologías son el Empatica E4 y el Empatica Embrace Plus, que se han utilizado para la detección de emociones y estados de ansiedad [73], [69], [74]. Ambos dispositivos miden EDA, HR, aceleración y temperatura.

Como se mencionó previamente, el músculo trapecio se contrae en situaciones de estrés y ansiedad y mediante sensores EMG se puede detectar la actividad muscular, permitiendo reconocer el estado psicofísico de los sujetos [19] , [72]. Algunos sensores adecuados para este propósito, especialmente por la ausencia de cables y la apariencia discreta, son los sensores Noraxon mini y los Cometa pico blue.

Por último, obtener información sobre el diámetro pupilar y el número de parpadeos puede complementar los datos extraídos por otros sensores y ayudar a detectar cambios en el estado psicológico [73], [19] , [72]. Además, estos datos siempre van acompañados de la dirección de la mirada, lo que es útil para obtener una comprensión más profunda de la visión del mundo de sujetos con discapacidad intelectual. Estos datos pueden obtenerse fácilmente mediante el uso de gafas HMD específicas, como Varjo Aero o HTC Vive XR Elite (complementado con un accesorio externo).

## 5.4 Requisitos de inmersividad e interacción de los casos de uso

En este apartado se presentan los requisitos de inmersividad e interacción de cada uno de los casos de uso de INCLUVERSO 5G. Para ello, las siguientes tablas recogen la siguiente información: título del caso de uso con la tecnología implicada, descripción,

caracterización del caso de uso desde el punto de vista de personas usuarias y caracterización del caso de uso desde el punto de vista tecnológico.

## 5.4.1 Terapia

<b>Título</b>
Terapia: desensibilización sistemática
<b>Descripción</b>
<p>El objetivo de esta terapia es tratar las conductas de evitación o comportamientos ansiosos producidos por estímulos a los que las personas tienen fobia. En concreto, se trabaja el miedo a las escaleras, una fobia compartida por personas con discapacidad de la FJ23 que afecta a su autonomía en el día a día e imposibilita el acceso a entornos con escaleras. Se adapta la técnica de la desensibilización sistemática, altamente testada para el tratamiento de fobias, por medio de la tecnología facilitando, por ejemplo, la práctica en imaginación o la exposición a diferentes estímulos, y mejorando así la efectividad de la terapia. La desensibilización sistemática está dividida en cuatro fases. Sin embargo, no se precisa de un entorno inmersivo, gracias a unos cascos de RV, hasta las dos últimas fases: práctica en imaginación y desensibilización. En la práctica en imaginación, el paciente visualiza entornos considerados neutros que fomentan la imaginación. En la etapa de desensibilización el paciente visualiza escenas que generan ansiedad. En esta fase se contempla la opción 3DoF, en la que se visualiza el entorno real en 360 grados y la opción 6DoF en la que el paciente se mueve por el entorno, acercándose a las escaleras y visualizando su cuerpo.</p> <p>Esta experiencia se puede utilizar en un entorno totalmente automatizado que incluye un algoritmo de segmentación de aprendizaje profundo que se ejecuta en el servidor de computación en el borde.</p>

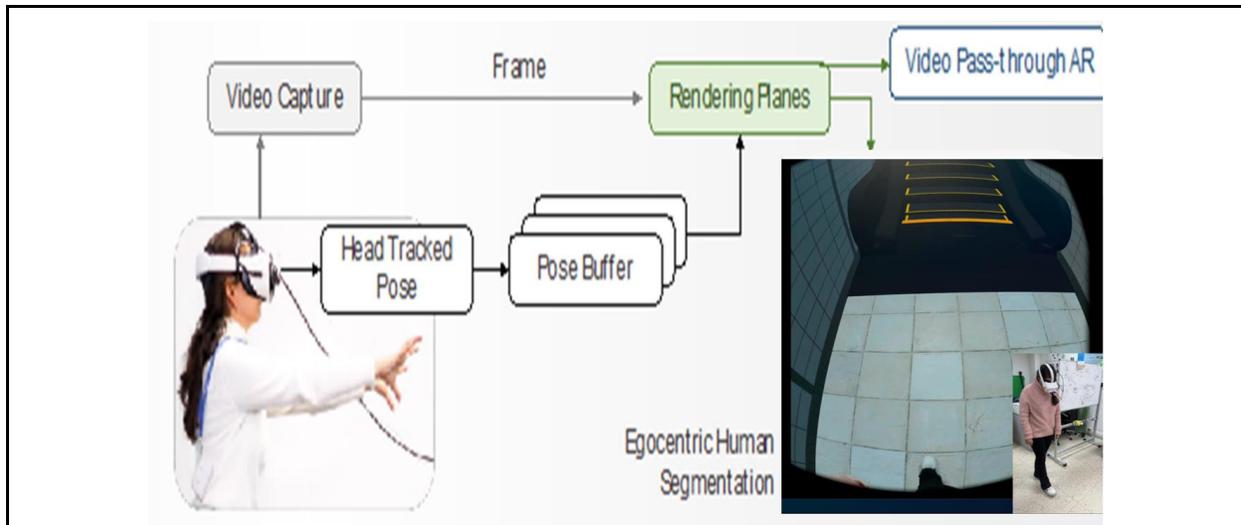


Figura 9: Ejemplo de paciente visualizando la segmentación de su cuerpo en un entorno virtual que genera ansiedad.

Caracterización del caso de uso desde el punto de vista de personas usuarias:

La aplicación es para una persona, con interacción en tiempo real con el entorno virtual y aplicando computación delegada, como por ejemplo, para obtener la máscara de segmentación obtenida en el servidor de computación en el borde, que permite que los/as pacientes puedan visualizar su cuerpo. Adicionalmente, el/la paciente está monitorizado/a con biosensores que capturan señales fisiológicas para transformarlas en biomarcadores digitales que provean de información a la terapeuta acerca del estado emocional del paciente y se mantenga una monitorización a lo largo de la sesión.

La desensibilización sistemática está supervisada por la terapeuta que, a través de una aplicación de control, guía la sesión eligiendo los estímulos a presentar, activando funcionalidades como el sonido y modificando los elementos virtuales que se visualizan, por ejemplo, las escaleras mecánicas con movimiento o sin él.

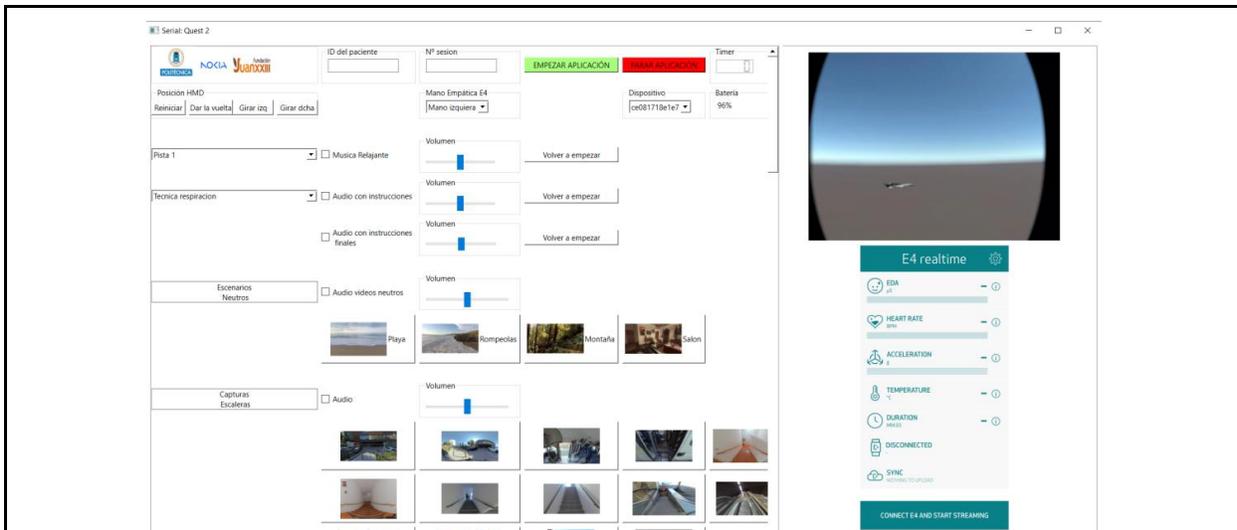


Figura 10: Ejemplo de la interfaz de la aplicación de control que la terapeuta utiliza en las sesiones

#### Actores:

- Paciente (persona que se introduce en el entorno virtual).
- Terapeuta (persona que utiliza la aplicación de control).

#### Sistema:

- Cascos de realidad virtual.
- Red de comunicación inalámbrica.
- Biosensores.
- Servidores de computación en el borde, donde se ejecutan los algoritmos de aprendizaje profundo para realizar la segmentación egocéntrica y alojar la aplicación inmersiva.

#### Caracterización del caso de uso desde el punto de vista técnico:

##### Recopilación de información:

La información local (cámara, movimiento) se transmite al servidor de computación en el borde para ejecutar los algoritmos de aprendizaje profundo (segmentación egocéntrica).

La segmentación generada por el algoritmo de aprendizaje profundo se transmite al paciente.

Bioseñales recogidas por los biosensores que llevan los/as pacientes.

Referencia al escenario de desarrollo:

Computación Delegada en MR + Regulación Emocional.

## 5.4.2 Telepresencia

Título

Telepresencia inmersiva

Descripción

El objetivo de la telepresencia inmersiva es proveer a los pisos tutelados de la FJ23 de una supervisión remota por parte de los profesionales de apoyo. De esta forma, las personas convivientes desempeñan las tareas cotidianas o resuelven emergencias con una supervisión remota, menos intrusiva. Como consecuencia, las personas ganan confianza en si mismas, mejorando su independencia y calidad de vida. Adicionalmente, se reducen los recursos económicos y de tiempo.

El profesional lleva unos cascos de realidad virtual. Las personas que conviven en el piso tutelado disponen del sistema, Owl o Snowl, que tiene integrada una cámara 360 grados, altavoz y una jetson con la que controlar el sistema y visualizar al profesional remoto en forma de avatar y, gracias al soporte con ruedas, pueden desplazar el sistema por el piso.

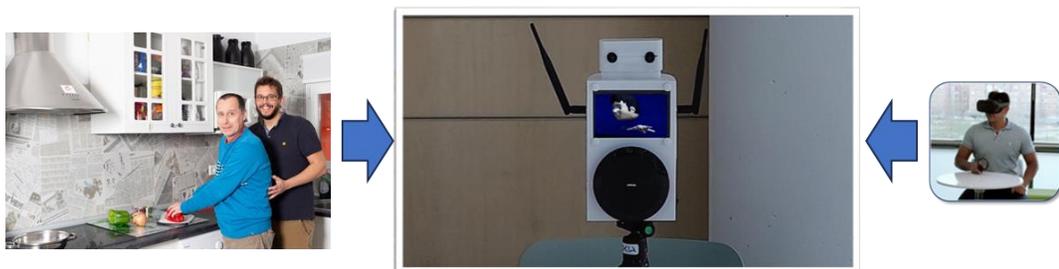


Figura 11: Ejemplo de telepresencia inmersiva. En la imagen central se observa Snowl

Caracterización del caso de uso desde el punto de vista de las personas usuarias:

Aplicación multiusuario con usuario remoto, interacción A/V en tiempo real, avatar y personas usuarias/puntos finales conectadas.

<p>Actores:</p> <ul style="list-style-type: none"> <li>• Locales, personas convivientes en el piso.</li> <li>• Profesional en remoto.</li> <li>• Sistema:             <ul style="list-style-type: none"> <li>• Snowl en el piso tutelado con cámara de adquisición de 360 grados, jetson, altavoz y soporte con ruedas.</li> <li>• Redes de comunicación inalámbrica tanto en los pisos tutelados como en la ubicación del profesional en remoto.</li> <li>• Cascos de realidad virtual.</li> <li>•</li> </ul> </li> <li>• Interactividad:             <ul style="list-style-type: none"> <li>• Comunicación oral bidireccional entre usuarios locales y profesional en remoto.</li> <li>• Representación del avatar con movimiento de la cabeza y boca sincronizada con el profesional en remoto.</li> </ul> </li> </ul>
<p>Caracterización del caso de uso desde el punto de vista técnico:</p>
<p>Recopilación de información:</p> <p>La información local (A/V) es transmitida al profesional en remoto.</p> <p>Avatares (movimientos recogidos por los cascos de realidad virtual) y audio desde el profesional remoto al piso tutelado.</p>
<p>Referencia al escenario de desarrollo:</p>
<p>Teleportación Virtual.</p>

### 5.4.3 Teleformación

<p><b>Título</b></p>
<p>Teleformación</p>
<p><b>Descripción</b></p>
<p>La teleformación está enfocada a las personas usuarias del COFOIL de la FJ23. Como objetivos principales se consideran: mejorar las habilidades de atención, control</p>

inhibitorio y velocidad de procesamiento, como habilidades necesarias para el desarrollo de una actividad laboral, y asistencia a teleformación de actividades específicas, considerando visualización de vídeos 360 grados y tutorías con la profesora por medio de tecnología inmersiva, utilizando a Snowl. El objetivo de esta teleformación es reforzar la interiorización de las actividades laborales del alumnado, complementando la formación tradicional con esta y mejorando el entrenamiento cognitivo. Todo esto puede llevarse a cabo de manera remota, haciendo posible que el alumnado no disminuya la curva de aprendizaje durante los meses de verano o semanas que no pueda asistir presencialmente al centro de la FJ23.

Para las actividades de entrenamiento de atención, control inhibitorio y velocidad de procesamiento, se utilizan biosensores que permiten monitorizar el desarrollo de las actividades y evaluar la mejora en la realización de actividades de formación profesional.

Durante la visualización de los vídeos 360 grados, el alumno/a manipula utensilios de cocina virtuales o reales. Típicamente, el alumnado puede visualizar:

- Objetos en tiempo real.
- Segmentación de manos en tiempo real.



Figura 12: Ejemplo de una alumna visualizando una clase de cocina en el entorno virtual mientras puede visualizar sus manos y manipular utensilios de cocina necesarios para llevar a cabo la receta

Caracterización del caso de uso desde el punto de vista de personas usuarias:

Se precisa de interacción en tiempo real con el entorno virtual y objetos por lo que se aplican los resultados de computación en el borde.

Recopilación de información:

La información local (cámara, pose) se transmite al servidor de computación en el borde para ejecutar los algoritmos de aprendizaje profundo (segmentación egocéntrica).

La segmentación generada por el algoritmo de aprendizaje profundo se transmite al alumno/a.

Bioseñales recogidas por los biosensores que llevan los/as alumnos.

Actores:

- Alumnado.
- Profesora.

Sistema:

- Cascos de realidad virtual.
- Biosensores.
- Red de comunicación inalámbrica.
- Servidores de computación en el borde, donde se ejecutan los algoritmos de aprendizaje profundo para realizar la segmentación egocéntrica y alojar la aplicación inmersiva.

Interactividad:

- Manipulación de elementos virtuales o reales.
- Segmentación (manos, seguimiento de objetos...).
- Comunicación oral bidireccional entre estudiantes locales y profesora en remoto.
- Representación del avatar con movimiento de la cabeza y boca sincronizada con la profesora en remoto.

Caracterización del caso de uso desde el punto de vista técnico:

Recopilación de información:

La información local (cámara egocéntrica, pose) se transmite al servidor de computación en el borde para ejecutar los algoritmos de aprendizaje profundo (segmentación egocéntrica o segmentación de objetos reales).

La segmentación generada por el algoritmo de aprendizaje profundo se transmite al paciente.

La información local (A/V) es transmitida a la profesora que guía la tutoría en remoto. Avatares (movimientos recogidos por los cascos de realidad virtual) y audio desde la profesora en remoto al alumno/a.

Referencia al escenario de desarrollo:

Teleportación Virtual + Computación Delegada en MR + Regulación Emocional

## 5.5 Escenarios base

Dado que cada caso de uso se puede implementar de distintas maneras, en esta sección se describen los escenarios base en los que se pueden desplegar los distintos casos de uso. Para ello se emplea una aproximación funcional, centrada en las distintas arquitecturas básicas necesarias para implementar los casos de uso. Los escenarios base, por tanto, están estructurados en función del tipo y capacidades de dispositivos que se deben emplear, y de los principales flujos de datos implicados. Esta aproximación es la misma que se ha empleado en 3GPP TR 26.928 (“XR in 5G”) y en el entregable “D1: Network Requirements and Capabilities” de Metaverse Standard Forum.

Las arquitecturas descritas son simplificadas e ilustrativas, y consideran fundamentalmente la interacción entre dispositivos XR y clientes XR, posiblemente a través de una red 5G o 5G avanzada. Se indica también el elemento o elementos para los que se considera movilidad (es decir, conectividad a través de red móvil) en cada uno de los escenarios base.

### 5.5.1 Escenario 1: Teleportación Virtual (Virtual Teleportation)

Este escenario está relacionado con la arquitectura de realidad extendida conversacional (“XR Conversational”) definida en 3GPP TR 26.928. Se trata de un esquema híbrido compuesto por un dispositivo de telepresencia, que captura en tiempo real el entorno local mediante vídeo 360; un servidor de telepresencia, que gestiona las sesiones y los flujos; y uno o varios clientes de XR, que asisten remotamente a la reunión inmersiva. La Figura 13 representa una arquitectura de alto nivel de este escenario, mostrando sus principales componentes e interfaces.

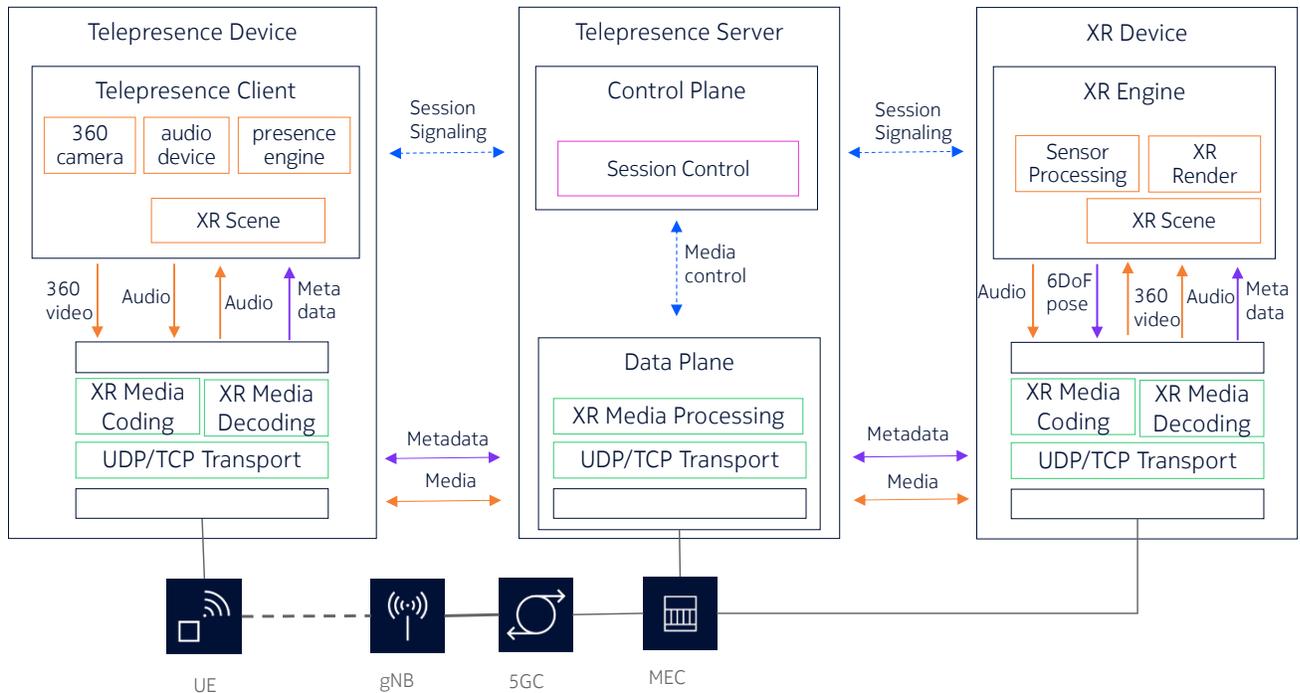


Figura 13: Arquitectura de alto nivel del escenario teleportación virtual

El dispositivo de telepresencia tiene las siguientes capacidades:

- Captura y procesado de la escena inmersiva mediante una cámara 360.
- Comunicación de audio bidireccional (captura y reproducción).
- Representación de presencia de los usuarios remotos en la escena local.
- Comunicación por 5G.

El servidor de XR tiene las siguientes capacidades:

- Plano de control: gestión de la señalización y de las sesiones de telepresencia.
- Procesado de los flujos de datos multimedia.
- Comunicación por 5G.

El cliente de XR tiene las siguientes capacidades:

- Captura y procesado de sensores.
- Generación y *render* de la escena XR.
- Codificación, decodificación y transporte de flujos multimedia.
- Comunicación por 5G u otras redes

## 5.5.2 Escenario 2: Procesamiento delegado de Realidad Mixta (*Mixed Reality computation offloading*)

Este escenario está relacionado con la arquitectura de computación distribuida para realidad extendida (“XR Distributed Computing”) definida en 3GPP TR 26.928. El dispositivo de XR delega parte del procesamiento a un servidor externo en el borde de la red (*Edge*), para poder ejecutar algoritmos de inteligencia artificial que no pueden implementarse en el propio dispositivo de XR por capacidad de proceso o consumo de energía. La Figura 14 representa una arquitectura de alto nivel de este escenario, mostrando sus principales componentes e interfaces.

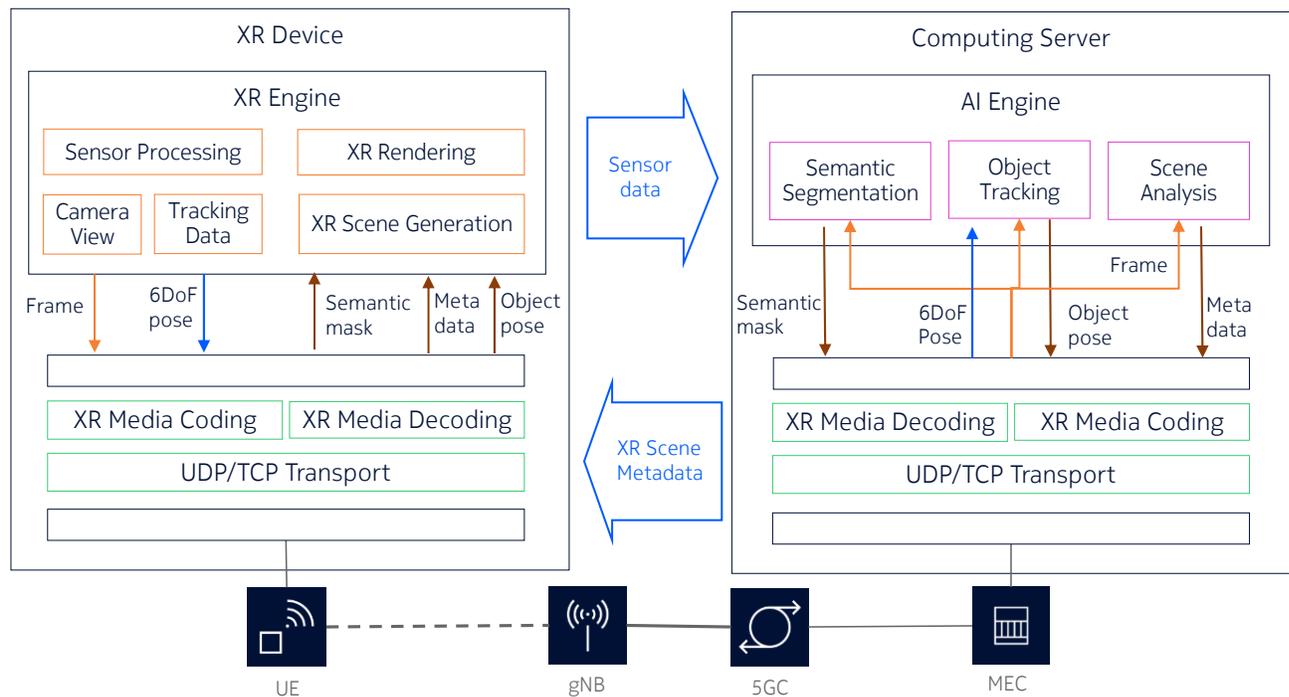


Figura 14: Arquitectura de alto nivel del escenario de procesamiento delegado de Realidad Mixta

El cliente de XR tiene las siguientes capacidades:

- Captura y procesamiento de sensores, incluyendo:
  - Cámara egocéntrica.
  - Seguimiento y posicionamiento del dispositivo en 3 dimensiones.
  - Generación y *render* de la escena XR.
  - Codificación, decodificación y transporte de flujos multimedia.
  - Comunicación por 5G.

El servidor de XR tiene las siguientes capacidades:

- Implementación de algoritmos de inteligencia artificial.

- Codificación, decodificación y transporte de flujos multimedia.
- Comunicación por 5G.

El cliente XR se conecta a la red y accede a un servidor de procesamiento en el borde de la red. Ambos dispositivos negocian los formatos de codificación y transporte que van a utilizar para intercambiar información.

El cliente XR recoge datos de varios sensores, tales como dispositivos de seguimiento y captura, incluyendo una cámara egocéntrica. Estos datos se pueden procesar en el mismo dispositivo o en el servidor de procesamiento en el borde de la red. Ya sea la información recogida directamente de los sensores o la procesada localmente (como la posición actual en la escena inmersiva) se envía al servidor de procesamiento.

El servidor ejecuta uno o varios algoritmos de inteligencia artificial para generar información aumentada que se envía de vuelta al cliente de XR y se integra en la escena inmersiva, bien como medio 2D o 3D, bien como metadatos. Finalmente, el cliente XR crea la escena XR combinando los medios y metadatos recibidos o generados localmente, y muestra la escena XR a través de su dispositivo de visualización, como un casco de realidad virtual (HMD) o gafas de realidad aumentada.

Ejemplos de algoritmos considerados en este escenario son:

- Segmentación semántica de la cámara egocéntrica para la integración de la visión propia dentro de la escena inmersiva.
- Detección y seguimiento de objetos en el entorno local para integrarlos en la escena XR e interactuar con ellos.
- Análisis del entorno local (p. ej. identificación de objetos o de escenas).

### 5.5.3 Escenario 3: Regulación Emocional (*Emotion Regulation*)

Este escenario está también relacionado con la arquitectura de computación distribuida para realidad extendida (“XR Distributed Computing”) definida en 3GPP TR 26.928. En este caso, el servidor de computación en el borde de la red analiza los datos capturados a los usuarios para generar información de biomarcadores digitales del usuario, como su estado emocional, que se envía de vuelta a la aplicación de XR. La Figura 15 representa una arquitectura de alto nivel de este escenario, mostrando sus principales componentes e interfaces.

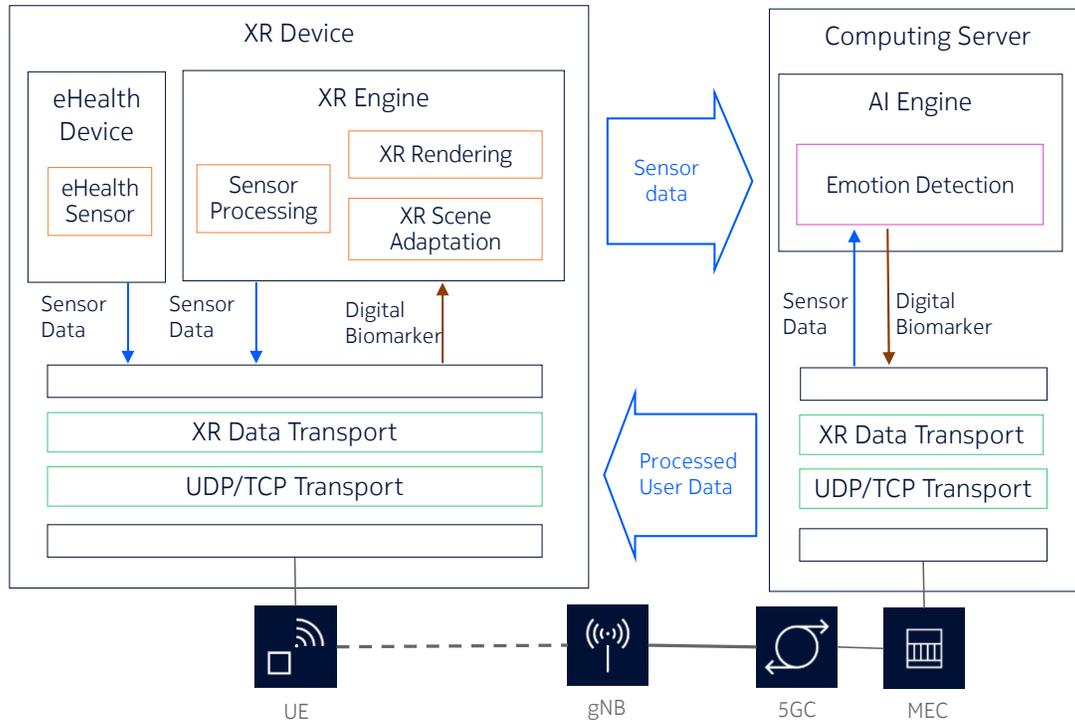


Figura 15: Arquitectura de alto nivel del escenario de regulación emocional

El cliente de XR tiene las siguientes capacidades:

- Captura y procesamiento de sensores.
- Integración de sensores biomédicos externos (p. ej. medida de señales de pulso o conductividad de la piel).
- Generación y *render* de la escena XR.
- Representación y transporte de los datos de los sensores.
- Comunicación por 5G.

El servidor de XR tiene las siguientes capacidades:

- Implementación de algoritmos de inteligencia artificial.
- Representación y transporte de los datos de los sensores.
- Comunicación por 5G.

El cliente XR recoge información de los sensores, tanto propios del dispositivo de XR como de los dispositivos biomédicos adicionales, y la envía al servidor de computación. Allí se ejecuta un algoritmo que permite modelar el estado emocional o cognitivo del usuario, identificando por ejemplo ansiedad o atención. Esta información se envía de nuevo al dispositivo de XR y se utiliza para adaptar la escena XR que se muestra al usuario.

### 5.5.4 Correspondencia entre los casos de uso y los escenarios base

Los tres casos de uso definidos en el proyecto comparten algunos puntos comunes. Para posibilitar su despliegue, así como para facilitar la investigación en las tecnologías que los permiten, se han definido distintos escenarios base de despliegue. Estos escenarios ofrecen soluciones potenciales para la implementación de los distintos casos de uso, y describen para ello una arquitectura de alto nivel que se puede analizar para obtener los requisitos técnicos. Como se ha mencionado anteriormente, este análisis está alineado con el descrito en el informe técnico 3GPP TR 26.928 (“XR in 5G”) y en el entregable “D1: Network Requirements and Capabilities” de Metaverse Standard Forum. La asociación entre los casos de uso y los escenarios base se muestra en la siguiente tabla:

Tabla 2: Correspondencia entre los casos de uso y los escenarios base

	Terapia	Telepresencia	Teleformación
Teleportación Virtual		X	X
Computación Delegada en MR	X		X
Regulación Emocional	X		X

Como se comprueba en la Tabla 2, le ejecución de algunos casos de uso extremo a extremo puede implicar combinar varios de los escenarios base, bien simultáneamente, bien de forma secuencial.

La definición de los escenarios base facilita también la investigación en tecnología dentro del proyecto. La Tabla 3 muestra la relación entre las tareas del paquete de trabajo P3 y los escenarios base. De las tecnologías investigadas en este paquete de trabajo, se seleccionará y adaptará el conjunto adecuado para la ejecución de los casos de uso que se aborda en el paquete de trabajo P4.

Tabla 3: relación entre las tareas del paquete de trabajo P3 y los escenarios base

		Teleportación Virtual	Computación Delegada en MR	Regulación Emocional
A3.1	Presencia autónoma		X	
A3.2	Interacción natural		X	
A3.3	Telepresencia inmersiva	X		
A3.4	Biomarcadores digitales			X
A3.5	Sistema XR	X	X	X
A3.6	Co-diseño	X	X	X

### 5.5.5 Requisitos e indicadores de rendimiento

La Tabla 4 resume los indicadores de rendimiento (KPIs, Key Performance Indicators) de los distintos escenarios base.

Tabla 4: KPIs de los distintos escenarios base

KPIs		Teleportación Virtual	Computación Delegada en MR	Regulación Emocional
Sesión	Núm. Usuarios	[2-10]	[1-10]	[1-100]
	Velocidad Usuario	Caminando	Estático	Cualquiera
	Distancia Usuarios	Cualquier lugar	Local (área MEC)	Cualquier lugar
Tasa de datos	Vídeo	5-20 Mbps UL	10-100 Mbps UL 1-10 Mbs DL	-
	Audio	0.1-0.5 Mbps UL/DL	-	-
	Sensores	-	0.1 Mbps UL	0.1 Mbps UL
	Datos	0.2 Mbps DL	0.2 Mbps DL	0.1 Mbps DL
Latencias	Latencia round-trip	< 600 ms	< 100 ms	< 2 s
	Tasa refresco XR	30 fps	60 fps	-
	Frecuencia IA	-	15 infer/s	1 infer/s
QoE	MOS perceptual	> 3.5	> 4.0	-

	Presencia espacial	Alta	Alta	-
	Presencia social	Muy alta	-	-
	Presencia autónoma	Media	Muy alta	
	Cibermareo	Mínimo (VSR > 4.5)	Bajo (VSR > 4.0)	-

Notas:

UL: Uplink, DL: Downlink

VSR: Vertigo Score Rating (ITU-T P.919). Escala desde 1 (insoportable) a 5 (sin problemas).

## 6 Definición de requisitos de la red de comunicaciones

### 6.1 Introducción al 5G avanzado

El 5G Avanzado, conocido como 5G-Advanced, representa una evolución significativa en la tecnología de telecomunicaciones, introduciendo varias mejoras y nuevas funcionalidades sobre la base del estándar 5G actual. Según la versión 18 del 3GPP, que marca el inicio de la 5G-Avanzada, esta tecnología se especificará en las versiones 18, 19 y 20 del 3GPP. Se espera que el 6G llegue al mercado alrededor de 2030, y 5G-Advanced servirá como un puente entre 5G y 6G, introduciendo mejoras tanto a nivel de sistema en el estándar 5G NR como en capacidades específicas para industrias y aplicaciones verticales.

El 5G Avanzado introducirá mejoras necesarias para aplicaciones más exigentes, como la realidad extendida móvil y aplicaciones hápticas. Inyectará más inteligencia en la red, utilizando el aprendizaje automático para adaptarse al entorno, mejorar el rendimiento y gestionar optimizaciones complejas. También llevará la banda ancha móvil a nuevas clases de dispositivos (por ejemplo, realidad extendida, dispositivos de capacidad reducida, nuevas innovaciones de enlace lateral, operación de ancho de banda reducido) y abrirá 5G a nuevos sectores. La eficiencia energética será central en 5G Avanzado, tanto en maximizar la vida útil de la batería del dispositivo como en reducir el consumo de energía de la red.

Se espera que para el año 2030, el 5G Avanzado esté implementado en una gran parte de la red móvil, preparando el terreno para la adopción de tecnologías avanzadas como la inteligencia artificial, el aprendizaje automático y la realidad extendida. Estas tecnologías tendrán un impacto significativo en varios sectores, desde el empresarial hasta el doméstico, facilitando servicios de alta demanda de banda ancha y funciones avanzadas.

### 6.2 Estado del arte de tecnologías de red

#### 6.2.1 Arquitecturas de red 5G privadas

El avance tecnológico y la creciente necesidad de conectividad de alto rendimiento han impulsado la expansión global de las redes privadas 5G. Estas redes están diseñadas para satisfacer demandas específicas de diversos sectores, ofreciendo mayores anchos de banda, movilidad mejorada, latencias reducidas, seguridad avanzada y una densa conectividad de dispositivos IoT. Este salto tecnológico es evidente en la transición

gradual desde tecnologías más antiguas (TETRA, P25, Digital Mobile Radio, GSM-R, WiFi, WiMax, Sigfox, LoRa, 4G LTE...) hacia el 5G, observado en sectores como manufactura, educación, minería, energía y en la administración de ciudades inteligentes [75].

Se estima que más de 1.148 empresas y organizaciones en 74 países ya implementan redes privadas 4G LTE o 5G. Los países líderes en este despliegue incluyen a Estados Unidos, Alemania, China, Reino Unido y Japón, con España mostrando un crecimiento más lento pero significativo [76]. Este auge se ve limitado por desafíos como el alto coste y la complejidad técnica. Sin embargo, la colaboración con operadoras móviles y la adopción de servicios gestionados están facilitando la superación de estas barreras, permitiendo a las empresas diseñar, implementar y mantener redes 5G eficientes y seguras. De hecho, según un informe de Straits Research, se proyecta que el mercado de estas redes experimente un crecimiento exponencial, pasando de 1.610 millones de dólares en 2022 a 54.890 millones de dólares en 2031 [76].

Las redes privadas 5G se diseñan específicamente para organizaciones únicas, especialmente industriales, proporcionando servicios de red dentro de sus recintos a dispositivos como robots móviles y sensores. Para implementar estas redes, se deben considerar aspectos como el espectro, los propietarios, operadores y la relación entre redes privadas y públicas, además de la disponibilidad de componentes y la viabilidad económica [80]. Según el informe técnico 3GPP TR 22.804 [77], existen dos formas de redes privadas 5G: autónomas e integradas en redes públicas.

Una red privada 5G autónoma se despliega como un sistema independiente y aislado de la red pública. Esta red cuenta con un identificador único (PLMN), utiliza espectro privado y aloja un sistema 5G completo (RAN y núcleo de red) dentro de su propio perímetro lógico. A pesar de su autonomía, los dispositivos de la red privada pueden necesitar acceder a servicios de la red pública, para lo cual se puede establecer una conexión opcional entre ambas redes mediante un firewall [80], [79].

El despliegue integrado con la red pública de una red privada 5G se puede clasificar en tres casos: 1) RAN Compartida, donde las redes privada y pública comparten parte del RAN pero mantienen separadas otras funciones de red; 2) RAN Compartida y Plano de Control, donde además del RAN compartido, la red pública controla las tareas de red, pero el tráfico de la red privada permanece en áreas definidas; y 3) Alojada por la Red Pública, donde la red privada está completamente alojada por la red pública, utilizando virtualización y *slices* de red para garantizar independencia y aislamiento [79]. En general, este tipo de despliegue ofrece menor personalización, autocontrol y seguridad en comparación con un despliegue autónomo.

En 3GPP TR 23.734 [78], se proponen dos soluciones para la continuidad del servicio entre redes privadas y públicas. La primera se basa en capacidades de radio dual. Un usuario puede estar simultáneamente registrado y conectado en ambas redes, privada y pública. La segunda solución se basa en una función de interconexión no 3GPP (N3IWF).

Para acceder a los servicios de la red pública a través de una red privada, un usuario obtiene conectividad IP a través de la red privada, descubre un N3IWF proporcionado por la red pública y establece conectividad con la red pública a través del N3IWF.

## 6.2.2 Productos en banda n40

La banda n40 de 5G, que opera en la frecuencia de 2300 MHz (también conocida como la banda de 2.3 GHz en 5G), es una banda de frecuencia TDD (Time Division Duplexing), que requiere solo una banda de frecuencia para tanto el enlace ascendente como el descendente. En España, se ha aprobado una actualización en el marco regulatorio del espectro radioeléctrico que incluye la habilitación de 20 MHz en la banda de 2300-2400 MHz para su uso en redes de comunicaciones móviles de banda ancha en régimen de autoprestación. Esto permite a las empresas y entidades disponer de espectro para desarrollar y gestionar sus propias aplicaciones y redes, incluyendo aplicaciones industriales. De los 20 MHz disponibles, 10 MHz se asignan con preferencia para redes de servicios públicos de distribución de electricidad, gas o agua.

Esta medida forma parte de un esfuerzo por actualizar las condiciones de uso de las bandas de frecuencia y adaptarlas a los estándares y acuerdos internacionales. La Comisión Nacional de los Mercados y la Competencia (CNMC) ha evaluado estos cambios y considera adecuada la reserva de espectro para autoprestación, siempre que se cumplan las condiciones y procedimientos que aseguren un uso eficiente y efectivo del espectro. Este enfoque representa una oportunidad para que las empresas y entidades creen redes personalizadas que se ajusten a sus necesidades específicas de conectividad.

Este cambio normativo no solo implica una mayor flexibilidad en la gestión del espectro, sino que también pone de relieve la importancia de contar con dispositivos y terminales móviles compatibles con esta banda específica. En este contexto, es relevante conocer qué modelos de smartphones y dispositivos móviles en el mercado actual están habilitados para operar en la banda n40. Para cubrir distintas opciones de despliegue y factor de forma, consideraremos dispositivos de diversas características: smartphone (iPhone 15), punto de acceso móvil (Inseego MIFI M2000), y módem de exteriores (Askey NUQ3000M).

El iPhone 15 ofrece soporte para la conectividad 5G, incluyendo una amplia gama de bandas 5G NR sub-6 GHz y 5G NR mmWave. Este dispositivo también cuenta con tecnologías de conectividad avanzadas como Gigabit LTE con 4x4 MIMO y LAA, Wi-Fi 6 (802.11ax) con 2x2 MIMO, y Bluetooth 5.3. La batería del iPhone 15 soporta hasta 20 horas de reproducción de video y hasta 80 horas de reproducción de audio, ofreciendo una carga rápida capaz de alcanzar el 50% en aproximadamente 30 minutos con un adaptador de 20W o superior.

El Inseego MIFI M2000 es un hotspot móvil que integra la tecnología 5G, ofreciendo una opción de conectividad para usuarios que requieren acceso a internet de alta velocidad en movimiento. Este dispositivo utiliza el chipset Qualcomm Snapdragon SDX55, permitiendo velocidades de descarga que pueden alcanzar hasta 2.7 Gbps, dependiendo de las condiciones de la red y la cobertura. Es compatible con una amplia gama de bandas de frecuencia 5G y LTE, lo que le confiere una buena adaptabilidad a diferentes redes y regiones. En términos de conectividad, el M2000 puede conectar hasta 30 dispositivos Wi-Fi simultáneamente, y emplea Wi-Fi 6, que proporciona una mejor gestión del tráfico de datos y eficiencia en entornos con muchos usuarios. En cuanto a la seguridad, ofrece características como WPA3 y soporte para VPN, lo que ayuda a proteger la información transmitida. El dispositivo también cuenta con una pantalla táctil de 2.4 pulgadas para la navegación y gestión del sistema, y ofrece opciones de carga rápida para su batería, que varía en capacidad dependiendo del modelo específico.

El Askey NUQ3000M es un módem 5G NR ODU (Outdoor Unit) diseñado para proporcionar soluciones de conectividad en exteriores. Utiliza el procesador Qualcomm SDX65 y es compatible con la especificación 3GPP Release16. Este dispositivo es capaz de soportar tanto las bandas de onda milimétrica (mmWave) como Sub-6 GHz del 5G, brindando una cobertura flexible en diferentes entornos de red. Para la banda Sub-6 GHz, ofrece modos de operación dual NSA y SA, mientras que para mmWave, soporta el modo NSA. En cuanto a sus características físicas, el NUQ3000M incluye indicadores LED para alimentación y señal, un botón de reinicio y una ranura para tarjeta SIM tipo Nano, con opción para eSIM. Además, cuenta con un conector Ethernet RJ45 resistente al agua, que soporta velocidades de hasta 5Gbps y Power over Ethernet (PoE) PD. Su diseño compacto y robusto lo hace adecuado para instalaciones exteriores en diversos entornos.

### 6.2.3 Emulación de la red de acceso radio

Hay muchos simuladores Red de Acceso Radio (RAN) en el estado del arte, algunos de ellos ampliamente utilizados en la investigación y la industria. El Simulador Viena y ns-3 son probablemente dos de los simuladores más conocidos [90], [91]. Ambos fueron inicialmente diseñados para LTE, pero han incorporado capacidades de simulación 5G en los últimos años. Otro ejemplo interesante es el Matlab Toolbox 5G [92], muy útil para ingenieros e investigadores, ya que la flexibilidad proporcionada por Matlab les permite probar rápidamente sus propios modelos o algoritmos.

Los emuladores de RAN a nivel de sistema son particularmente interesantes porque pueden filtrar y manejar tráfico IP real, simulando con precisión y en tiempo real una tecnología o configuración RAN específica. Hay múltiples emuladores de RAN en el estado del arte, bien incluidos como funcionalidades en los simuladores mencionados anteriormente o bien desarrollos nuevos como SimuLTE [93] o Simu5G [94]. Además,

otros emuladores de la industria también son relevantes, como NetTest de Polaris Networks [95] o 5G Network Emulation Solutions de Keysight [96]. Todos estos emuladores están diseñados para estudiar el comportamiento de la red. Sin embargo, presentan limitaciones de rendimiento o de exceso de complejidad a la hora de usarlos para el desarrollo de aplicaciones.

Existen algunos emuladores diseñados específicamente para investigación en la capa de aplicación, como eXP-RAN [97] o FikoRE. El primero implementa un modelo de RAN abstracto y carece de una representación precisa de algunas funcionalidades relevantes. FikoRE, desarrollado por Nokia, proporciona un compromiso razonable entre capacidad de emulación de las principales características de 5G y simplicidad de diseño y manejo, lo que se traduce en un alto rendimiento, pudiendo procesar Gbps de tráfico con el hardware adecuado.

## 6.2.4 Realidad extendida en redes 5G avanzadas

La combinación de tasas binarias altas y baja latencia requerida por los servicios de XR plantea nuevos desafíos al acceso inalámbrico. Por eso en 3GPP se están introduciendo mecanismos específicos para el soporte de servicios XR en 5G avanzado a partir de Release 18.

Una experiencia XR aceptable requiere tasas de refresco de al menos 60 fps y una resolución de 2K por ojo, mientras que una verdaderamente inmersiva requiere 90 o incluso 120 fps con resoluciones de hasta 8K por ojo. Esto implica el uso de motores XR potentes que, en ocasiones, no es posible o deseable ejecutar en el dispositivo de usuario, por ejemplo, debido a la disipación de calor y las limitaciones de batería. Por lo tanto, el primer escenario base planteado por 3GPP para Release 18 es del renderizado asistido o dividido a través de la red (*split rendering*): el equipo de usuario (UE) envía datos de sensores en tiempo real por el enlace ascendente (*uplink*) a la nube, que realiza el renderizado y produce los datos multimedia que se envían de vuelta a los dispositivos del usuario por el enlace descendente (*downlink*).

Para apoyar la implementación de este caso de uso, en 5G avanzado se están definiendo las siguientes funcionalidades:

- Técnicas de marcado y procesado de los datos XR en la red (*XR Awareness*)
- Técnicas de ahorro de energía para gestionar datos en ráfagas con períodos no enteros (16.67, 11.11 y 8.33 ms para tasas de refresco de 60, 90 y 120 fps).
- Optimizaciones de la capa física para mejorar la capacidad de la celda.

Si bien estas técnicas se han diseñado sobre el escenario base de *split rendering* definido en 3GPP TR 26.928, las optimizaciones para tráfico XR, y en particular el *XR Awareness*, son de interés para los escenarios base considerados en el proyecto, especialmente para los de teleportación virtual y computación delegada de MR.

Tanto en enlace ascendente como descendente, el *XR Awareness* contribuye a las optimizaciones de recursos de radio y se basa en las nociones de *PDU Set* y *Data Burst* definidos por en la arquitectura de sistemas de 5G (3GPP TR 23.501). El *PDU Set* permite a la RAN identificar los PDU que transportan contenido que la aplicación procesa como una unidad única (por ejemplo, una porción de una imagen o un fotograma de audio) y el *Data Burst* identifica la duración de una transmisión de datos, como se muestra en la Figura 16.

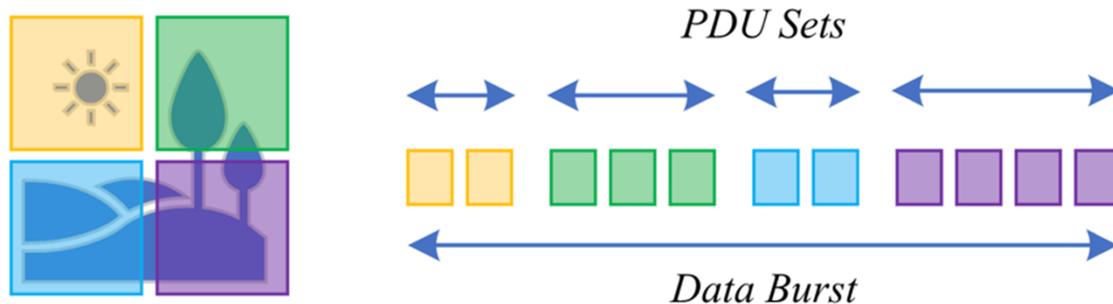


Figura 16: Representación de alto nivel de las funcionalidades de PDU Set y Data Burst

El 5G avanzado define la posibilidad de gestionar la Calidad de Servicio (QoS) de los flujos de XR basada en los PDU Set. Para ello se incorpora la siguiente información a los datos transmitidos por la red:

- Número de secuencia de PDU set
- Indicador de PDU final dentro del PDU Set.
- Número de secuencia de cada PDU dentro del PDU Set.
- Tamaño de la ráfaga de datos en bytes.
- PDU Set Importance, que identifica la importancia relativa de un PDU Set dentro de su flujo de datos (QoS Flow).

Con la Indicación de Manejo Integrado de Conjunto de PDU (PSIHI, por sus siglas en inglés), la RAN sabe si todos los PDU del PDU Set son necesarios para la capa de aplicación. Por un lado, la RAN puede dejar de transmitir un PDU Set en cuanto detecte que uno de los PDU del conjunto se ha perdido (ya que los otros PDU serían inútiles de todos modos). Por otro, la RAN puede usar el valor de *PDU Set Importance* en caso de congestión para descartar los PDU Set de menor importancia.

El marcado de los distintos paquetes que pertenecen a un mismo PDU Set, así como la información necesaria para procesarlo, puede hacerse en capa de aplicación empleando cabeceras RTP/SRTP. Este mecanismo se define en la nueva norma 3GPP TS 26.522 (“5G Real-time Media Transport Protocol Configurations”).

## 6.3 Requisitos de conectividad y procesamiento de los escenarios base

Los tres escenarios base descritos en el capítulo anterior se usarán como referencia para establecer los requisitos que debe cumplir la red móvil para proporcionar los servicios considerados en el proyecto.

### 6.3.1 Requisitos de movilidad y cobertura

A la hora de establecer los requisitos, se considerará cada uno de los escenarios base de forma independiente. De esta manera, se podrán explorar distintos escenarios en función de sus características técnicas. Los casos de uso del proyecto se corresponderán con una particularización concreta de cada uno de estos escenarios.

En el escenario de teleportación virtual, se asume que el equipo que está conectado a la red móvil como UE es el sistema de telepresencia (Snowl). No existe una restricción específica de la localización del dispositivo, que podrá estar tanto en interiores como en exteriores. Se considera que la sesión podrá durar hasta una hora aproximadamente, en la que el dispositivo se puede cambiar de posición de forma esporádica.

En los escenarios de computación delegada de MR, el equipo de usuario (UE) es el cliente de XR (gafas de VR/AR). Se asume que estará en interiores y en un espacio relativamente reducido, sin apenas movilidad (en relación a la celda móvil). Las sesiones tendrán duración de algunos minutos, normalmente.

El escenario de regulación emocional puede desarrollarse en conjunto con alguno de los otros, o de forma independiente. En este último caso, no hay restricción de ningún tipo acerca de la movilidad del usuario. El sistema puede estar funcionando de forma continua durante varias horas.

La Tabla 5 resume los requisitos de movilidad, cobertura y densidad de celda de los escenarios:

Tabla 5: Requisitos de movilidad, cobertura y densidad de celda de los escenarios base

	Cobertura	Movilidad	Densidad
Teleportación Virtual	Interior / Exterior	Baja	Microcelda
Computación Delegada en MR	Interior	Estático	Micro/Picocelda
Regulación Emocional	Interior / Exterior	Media	Macro/Microcelda

### 6.3.2 Requisitos de tráfico

La Tabla 6 resume las principales características del tráfico para los escenarios base. Los parámetros se refieren al modelo de calidad de servicio de una sesión de PDU (5G PDU session QoS), siguiendo el esquema de 3GPP TR 26.928.

Tabla 6: Principales características del tráfico para los escenarios base

	Teleportación Virtual	Computación Delegada en MR	Regulación Emocional
DL Rate	0.3-0.8 Mbps	1-20 Mbps	0.1 Mbps
UL Rate	5-20 Mbps	10-200 Mbps	0.1 Mbps
DL PDB	100 ms	20 ms	300 ms
UL PDB	100 ms	20 ms	300 ms
DL PER	FFS	FFS	FFS
UL PER	FFS	FFS	FFS
Tipo Tráfico	Constante (por frame)	Constante (por frame)	A ráfagas

Notas:

- DL: Downlink
- UL: Uplink
- PDB: Packet delay budge
- PER: Packet Error Rate
- FFS: *For further study*. 3GPP no ha definido aún las tasas de error adecuadas para proporcionar servicios de XR en tiempo real.

### 6.3.3 Análisis de 5QIs existentes

Tal como se describe en 3GPP TR 26.928, es posible hasta cierto punto utilizar los 5QIs existentes para proporcionar servicios de XR sobre 5G avanzado. Para arquitecturas basadas en *streaming* adaptativo sobre HTTP, se pueden emplear los 5QIs ya considerados en 3GPP. Para otro tipo de tráfico, el punto de partida se establece en el informe técnico ITU-T GSTR-5GQoE (“Quality of experience (QoE) requirements for real-time multimedia services over 5G networks”), donde se analizan potenciales 5QIs que podrían usarse en servicios interactivos sobre redes 5G.

El escenario base de computación delegada en MR es muy similar al escenario homónimo en ITU-T GSTR-5GQoE, para el que se proponen los siguientes 5QIs en la Tabla 7.

Tabla 7: 5QIs del escenario base de computación delegada en MR

5QI Value	Resource Type	Default Priority Level	Packet Delay Budget	Packet Error Rate	Default Maximum Data Burst Volume	Default Averaging Window	Example Services
80	Non-GBR	68	10 ms	$10^{-6}$	N/A	N/A	Low Latency eMBB applications Augmented Reality
87	Delay-critical GBR	25	5 ms	$10^{-3}$	500 bytes	2000 ms	Interactive Service - Motion tracking data
88		25	10 ms	$10^{-3}$	1125 bytes	2000 ms	Interactive Service - Motion tracking data
89		25	15 ms	$10^{-4}$	17000 bytes	2000 ms	Visual content for cloud/edge/split rendering
90		25	20 ms	$10^{-4}$	63000 bytes	2000 ms	Visual content for cloud/edge/split rendering

Los escenarios base de teleportación virtual y regulación emocional podrían asimilarse parcialmente al escenario de emergencias médicas descrito en ITU-T GSTR-5GQoE, para el que se proponen los siguientes 5QIs en la Tabla 8.

Tabla 8: 5QIs de los escenarios base de teleportación virtual y regulación emocional

5QI Value	Resource Type	Default Priority Level	Packet Delay Budget	Packet Error Rate	Default Maximum Data Burst Volume	Default Averaging Window	Example Services
1	GBR	20	100 ms	$10^{-2}$	N/A	2000 ms	Conversational Voice
2		40	150 ms	$10^{-3}$	N/A	2000 ms	Conversational Video (Live Streaming)
65		7	75 ms	$10^{-2}$	N/A	2000 ms	Mission Critical user plane Push To Talk voice (e.g. MCPTT)
66		20	100 ms	$10^{-2}$	N/A	2000 ms	Non-Mission-Critical user plane Push To Talk voice

67		15	100 ms	$10^{-3}$	N/A	2000 ms	Mission Critical Video user plane
7	Non-GBR	70	100 ms	$10^{-3}$	N/A	N/A	Voice, Video (Live Streaming)
8		80	300 ms	$10^{-6}$	N/A	N/A	Video (Buffered Streaming)
9		90					
69		5	60 ms	$10^{-6}$	N/A	N/A	Mission Critical delay sensitive signalling (e.g. MC-PTT signalling)
70		55	200 ms	$10^{-6}$	N/A	N/A	Mission Critical Data (e.g. example services are the same as 5QI 6/8/9)

Estos 5QIs pueden emplearse como punto de partida para el análisis y diseño de QoS que se desarrollará en el proyecto. Es importante considerar que, en la práctica, las configuraciones de QoS que son posibles en las redes 5G y 5G avanzadas actuales son limitadas y que apenas hay experiencia práctica de utilizar distintos 5QIs en pruebas de campo, por lo que es probable que la configuración de QoS finalmente empleada en el proyecto utilice 5QIs diferentes a los analizados por ITU-T.

### 6.3.4 Requisitos de procesamiento

La estimación inicial de requisitos de procesamiento para los escenarios virtuales se realiza a partir de los datos existentes en el estado del arte. Estas estimaciones presentadas en la Tabla 9 se usan como referencia inicial para el diseño de los algoritmos de IA y procesado en el proyecto, pero deberán ser verificadas a lo largo de su desarrollo.

Tabla 9: Estimación inicial de requisitos de procesamiento para los escenarios base planteados

	Teleportación Virtual	Computación Delegada en MR	Regulación Emocional
Tasa de cuadro	30 fps	15 - 60 fps	10 muestras / s
Resolución de cuadro	3840 x 2160	1280 x 480	N/A
Tasa binaria E/S	~10 Mbps	~100 Mbps	~100 kbps

Tiempo de inferencia	-	10-20 ms	0.1 – 1 s
Memoria GPU estimada	-	< 4GB	< 4GB
Referencia	Kachach 2021 [98]	González Morín 2023 [99]	

## 6.4 Especificación de los entornos de simulación de referencia

Finalmente, en base al análisis previo, se especifican unos escenarios de referencia para la configuración del emulador de red en las pruebas del proyecto presentados en la Tabla 10.

Tabla 10: Parámetros basados en los escenarios de referencia para la configuración del emulador

	Teleportación Virtual	Computación Delegada en MR	Regulación Emocional
Tipo de celda	Microcelda Urbana	Picocelda urbana / Hostpot	Macrocelda Urbana
Banda n40	20 MHz TDD	20 MHz TDD	20 MHz TDD
Banda n78	100 MHz TDD	100 MHz TDD	-
Banda n256	-	800 MHz TDD	-
Distancia a la antena	150 m	50 m	500 m
Usuarios concurrentes	1-10	1-10	10-50
Distancia media de los usuarios	200 m	50 m	1000 m
Distancia máxima de los usuarios	500 m	200 m	5000 m

## 7 Acrónimos

AR	Realidad Aumentada
AV	Virtualidad Aumentada
BVP	Pulso de volumen sanguíneo
COFOIL	Centro Ocupacional de Formación, Oportunidades e Inserción Laboral
CNMC	Comisión Nacional de los Mercados y la Competencia
DL	Downlink
DoF	Grados de Libertad
DS	Desensibilización Sistemática
EDA	Actividad ElectroDérmica
EEG	Electroencefalograma
EMG	Electromiografía
FJ23	Fundación Juan XXIII Roncalli
HR	Frecuencia Cardíaca
HMD	Head Mounted Display
KPI	Indicadores de Rendimiento
NUI	Interfaces Naturales de Usuario
ODU	Outdoor Unit
PoE	Power over Ethernet
PSIHI	Manejo Integrado de Conjunto de PDU
QoE	Calidad de Experiencia
QoS	Calidad de Servicio
RAN	Red de Acceso Radio
RV	Realidad Virtual
SCL	Skin Conductance Level
SCR	Skin Conductance Response
TDD	Time Division Duplexing



UIT	Unión Internacional de Telecomunicaciones
UL	Uplink
UPM	Universidad Politécnica de Madrid
USA	Escala de Unidades Subjetivas de Ansiedad
VSR	Vertigo Score Rating
XR	EXtended Reality/ Realidad eXtendida

## 8 Índice de tablas

Tabla 1: Estado del arte de estudios relacionados con la representación de avatares .....	30
Tabla 2: Correspondencia entre los casos de uso y los escenarios base .....	45
Tabla 3: relación entre las tareas del paquete de trabajo P3 y los escenarios base ....	46
Tabla 4: KPIs de los distintos escenarios base .....	46
Tabla 5: Requisitos de movilidad, cobertura y densidad de celda de los escenarios base .....	54
Tabla 6: Principales características del tráfico para los escenarios base .....	55
Tabla 7: 5QIs del escenario base de computación delegada en MR.....	56
Tabla 8: 5QIs de los escenarios base de teleportación virtual y regulación emocional .....	56
Tabla 9: Estimación inicial de requisitos de procesamiento para los escenarios base planteados .....	57
Tabla 10: Parámetros basados en los escenarios de referencia para la configuración del emulador .....	58

## 9 Índice de figuras

Figura 1: Ejemplo de una escena de un entorno sereno, un salón de una casa.....	9
Figura 2: Ejemplo de unas escaleras de la Fundación Juan XXIII Roncalli en proyección equirectangular .....	10
Figura 3: Ejemplo de una escena de realidad virtual que replica unas escaleras de la Fundación Juan XXIII consideradas en la jerarquía .....	10
Figura 4: Ejemplo de cómo una persona con las gafas de RV puede observar sus propias manos gracias al algoritmo de segmentación .....	11
Figura 5: Ejemplo de cómo una persona con las gafas de RV puede observar sus propios pies y piernas gracias al algoritmo de segmentación.....	11
Figura 6: Ejemplo de la interfaz de la aplicación de control para la terapeuta encargada de las sesiones de terapia .....	14
Figura 7: Ejemplos de avatares con mayor a menor realismo de izquierda a derecha	20
Figura 8: Reality–Virtuality Continuum [56] .....	28
Figura 9: Ejemplo de paciente visualizando la segmentación de su cuerpo en un entorno virtual que genera ansiedad.....	34
Figura 10: Ejemplo de la interfaz de la aplicación de control que la terapeuta utiliza en las sesiones .....	35
Figura 11: Ejemplo de telepresencia inmersiva. En la imagen central se observa Snowl .....	36
Figura 12: Ejemplo de una alumna visualizando una clase de cocina en el entorno virtual mientras puede visualizar sus manos y manipular utensilios de cocina necesarios para llevar a cabo la receta.....	38
Figura 13: Arquitectura de alto nivel del escenario teleportación virtual .....	41
Figura 14: Arquitectura de alto nivel del escenario de procesamiento delegado de Realidad Mixta.....	42
Figura 15: Arquitectura de alto nivel del escenario de regulación emocional .....	44
Figura 16: Representación de alto nivel de las funcionalidades de PDU Set y Data Burst .....	53

## 10 Referencias

- [1] Wolpe, J. (1968). Psychotherapy by reciprocal inhibition. *Conditional reflex: a Pavlovian journal of research & therapy*, 3, 234-240.
- [2] Botella, C., Fernández-Álvarez, J., Guillén, V., García-Palacios, A., & Baños, R. (2017). Recent progress in virtual reality exposure therapy for phobias: a systematic review. *Current psychiatry reports*, 19, 1-13.
- [3] Brito, H., & Vicente, B. (2018). Realidad virtual y sus aplicaciones en trastornos mentales: una revisión. *Revista chilena de neuro-psiquiatría*, 56(2), 127-135.
- [4] Garcia-Palacios, A., Botella, C., Hoffman, H., & Fabregat, S. (2007). Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by patients with specific phobias. *Cyberpsychology & behavior*, 10(5), 722-724.
- [5] Soto Triana, J. S., & Gómez Villamizar, C. (2018). Avances en realidad virtual e intervenciones en psicología clínica.
- [6] Ramírez, A., Benítez, S. J., & Gómez, M. A. (2017). La acrofobia y su tratamiento a través de realidad virtual. *Revista Ingeniería, Matemáticas y Ciencias de la Información*, 4(7).
- [7] Donker, T., Cornelisz, I., Van Klaveren, C., Van Straten, A., Carlbring, P., Cuijpers, P., & Van Gelder, J. L. (2019). Effectiveness of self-guided app-based virtual reality cognitive behavior therapy for acrophobia: a randomized clinical trial. *JAMA psychiatry*, 76(7), 682-690.
- [8] Alvarez Urueña, F. I., & Aguayo, L. V. (2023). La realidad virtual para la acrofobia: Evaluación de la usabilidad de una VRBox comercial. *Revista Espanola de Comunicacion en Salud*, 14(1).
- [9] Anderson, P., Rothbaum, B. O., & Hodges, L. F. (2003). Virtual reality exposure in the treatment of social anxiety. *Cognitive and Behavioral Practice*, 10(3), 240-247.
- [10] Reyna, D., Caraza, R., Gonzalez, M., Ayala, A., Martinez, P., Loreda, A., ... & Reyes, P. Realidad Virtual en el Tratamiento de Fobia Social.
- [11] Benavides, C. (2015, September). Virtual reality in the treatment of cynophobia. In *2015 10th Computing Colombian Conference (10CCC)* (pp. 499-503). IEEE.
- [12] Botella, C., Villa, H., García-Palacios, A., Baños, R. M., Quero, S., & Alcañiz, M. Virtual reality exposure in the treatment of panic disorder and agoraphobia: a comparison with in vivo exposure. *Clinical Psychology and Psychotherapy*.
- [13] Mor, S., Botella, C., Campos, D., Carlbring, P., Tur, C., & Quero, S. (2022). An internet-based treatment for flying phobia using 360° images: A feasibility pilot study. *Internet Interventions*, 28, 100510.
- [14] Kılıç, A., Brown, A., Aras, I., Hui, R., Hare, J., Hughes, L. D., & McCracken, L. M. (2021). Using virtual technology for fear of medical procedures: a systematic review of the effectiveness of virtual reality-based interventions. *Annals of Behavioral Medicine*, 55(11), 1062-1079.

- [15] Morina, N., Ijntema, H., Meyerbröker, K., & Emmelkamp, P. M. (2015). Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments. *Behaviour research and therapy*, 74, 18-24.
- [16] Bircanin, F., Brereton, M., Sitbon, L., Ploderer, B., Azaabanye Bayor, A., & Koplick, S. (2021, May). Including adults with severe intellectual disabilities in co-design through active support. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).
- [17] Ziemke, T. (2013). What's that thing called embodiment?. In *Proceedings of the 25th Annual Cognitive Science Society* (pp. 1305-1310). Psychology Press.
- [18] Argelaguet, F., Hoyet, L., Trico, M., & Lécuyer, A. (2016, March). The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE virtual reality (VR)* (pp. 3-10). IEEE.
- [19] Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2019). Review on psychological stress detection using biosignals. *IEEE transactions on affective computing*, 13(1), 440-460.
- [20] Brunnström, K., Beker, S. A., De Moor, K., Dooms, A., Egger, S., Garcia, M. N., ... & Zgank, A. (2013). Qualinet white paper on definitions of quality of experience.
- [21] Raake, A., & Egger, S. (2014). Quality and quality of experience. In *Quality of Experience: Advanced concepts, applications and methods* (pp. 11-33). Cham: Springer International Publishing.
- [22] Lowenthal, P. R. (2009). Social presence. In *Encyclopedia of Distance Learning, Second Edition* (pp. 1900-1906). IGI Global.
- [23] Coelho, C., Tichon, J., Hine, T. J., Wallis, G., & Riva, G. (2006). Media presence and inner presence: the sense of presence in virtual reality technologies. *From communication to presence: Cognition, emotions and culture towards the ultimate communicative experience*, 11, 25-45.
- [24] Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3549-3557.
- [25] Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3), 225-240.
- [26] Schwind, V., Knierim, P., Haas, N., & Henze, N. (2019, May). Using presence questionnaires in virtual reality. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-12).
- [27] Voigt-Antons, J. N., Kojic, T., Ali, D., & Möller, S. (2020, May). Influence of hand tracking as a way of interaction in virtual reality on user experience. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)* (pp. 1-4). IEEE.
- [28] Cortés, C., Rubio, M., Pérez, P., Sánchez, B., & García, N. (2022, March). Qoe study of natural interaction in extended reality environment for immersive training. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 363-368). IEEE.

- [29] Villegas, A., Perez, P., Kachach, R., Pereira, F., & Gonzalez-Sosa, E. (2020, March). Realistic training in VR using physical manipulation. In *2020 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)* (pp. 109-118). IEEE.
- [30] Pérez, P., González-Sosa, E., Kachach, R., Pereira, F., & Villegas, A. (2021, November). Ecological validity through gamification: an experiment with a mixed reality escape room. In *2021 IEEE international conference on artificial intelligence and virtual reality (AIVR)* (pp. 179-183). IEEE.
- [31] Cortés, C., Gutiérrez, J., Pérez, P., Viola, I., César, P., & García, N. (2022, October). Impact of self-view latency on quality of experience: Analysis of natural interaction in xr environments. In *2022 IEEE International Conference on Image Processing (ICIP)* (pp. 3131-3135). IEEE.
- [32] Orduna, M., Pérez, P., Gutiérrez, J., & García, N. (2022). Methodology to assess quality, presence, empathy, attitude, and attention in 360-degree videos for immersive communications. *IEEE Transactions on Affective Computing*, 14(3), 2375-2388.
- [33] Orduna, M., Díaz, C., Muñoz, L., Pérez, P., Benito, I., & García, N. (2019). Video multimethod assessment fusion (VMAF) on 360VR contents. *IEEE Transactions on Consumer Electronics*, 66(1), 22-31.
- [34] Simiscuka, A. A., Togou, M. A., Verma, R., Zorrilla, M., O'Connor, N. E., & Muntean, G. M. (2022, June). An evaluation of 360 video and audio quality in an artistic-oriented platform. In *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)* (pp. 1-5). IEEE.
- [35] Hirway, A., Qiao, Y., & Murray, N. (2024). A Quality of Experience and Visual Attention Evaluation for 360 videos with non-spatial and spatial audio. *ACM Transactions on Multimedia Computing, Communications and Applications*.
- [36] Latoschik, M. E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., & Botsch, M. (2017, November). The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM symposium on virtual reality software and technology* (pp. 1-10).
- [37] Kokkinara, E., & McDonnell, R. (2015, November). Animation realism affects perceived character appeal of a self-virtual face. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games* (pp. 221-226).
- [38] Hauber, J., Regenbrecht, H., Hills, A., Cockburn, A., & Billinghurst, M. (2005). Social presence in two-and three-dimensional videoconferencing.
- [39] Ferracani, A., Faustino, M., Giannini, G. X., Landucci, L., & Del Bimbo, A. (2017, October). Natural experiences in museums through virtual reality and voice commands. In *Proceedings of the 25th ACM international conference on Multimedia* (pp. 1233-1234).
- [40] Gupta, K., Lee, G. A., & Billinghurst, M. (2016). Do you see what I see? The effect of gaze tracking on task space remote collaboration. *IEEE transactions on visualization and computer graphics*, 22(11), 2413-2422.

- [41] Wang, P., Bai, X., Billingham, M., Zhang, S., He, W., Han, D., ... & Han, S. (2020). Using a head pointer or eye gaze: The effect of gaze on spatial AR remote collaboration for physical tasks. *Interacting with Computers*, 32(2), 153-169.
- [42] Recommendation, ITU - T. (2020). Subjective test methodologies for 360° video on head-mounted displays. *International Telecommunication Union*, 919.
- [43] Espert Tortajada, R., & Villalba Agustín, M. D. R. (2014). Estimulación cognitiva: una revisión neuropsicológica. *Terapeia*, (6), 73-93.
- [44] Guadamuz Delgado, J. ., Miranda Saavedra, . M., & Mora Miranda, N. (2022). Actualización sobre neuroplasticidad cerebral. *Revista Médica Sinergia*, 7(6), e829.
- [45] Galvis Restrepo, A. Y., y Lopera Murcia, A. M. (2019). Recursos informáticos y discapacidad intelectual: Aplicaciones en el contexto escolar. *Revista Iberoamericana de Psicología*, 11(3), 73-83.
- [46] Schiza, E., Matsangidou, M., Neokleous, K., & Pattichis, C. S. (2019). Virtual reality applications for neurological disease: a review. *Frontiers in Robotics and AI*, 6, 100.
- [47] Alashram, A. R., Annino, G., Padua, E., Romagnoli, C., & Mercuri, N. B. (2019). Cognitive rehabilitation post traumatic brain injury: A systematic review for emerging use of virtual reality technology. *Journal of Clinical Neuroscience*, 66, 209-219.
- [48] Cesari, V., Galgani, B., Gemignani, A., & Menicucci, D. (2021). Enhancing qualities of consciousness during online learning via multisensory interactions. *Behavioral Sciences*, 11(5), 57.
- [49] Fernández, R., y Canals, G. (2018). Estimulación Cognitiva en personas con síndrome de Down Programa NeuronUp en Aura Fundación. Síndrome de Down Vida Adulta, 29.
- [50] Zhang, C. (2020). The why, what, and how of immersive experience. *IEEE Access*, 8, 90878-90888.
- [51] Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, 110(6), 785.
- [52] Skaramagkas, V., Giannakakis, G., Ktistakis, E., Manousos, D., Karatzanis, I., Tachos, N. S., ... & Tsiknakis, M. (2021). Review of eye tracking metrics involved in emotional and cognitive processes. *IEEE Reviews in Biomedical Engineering*, 16, 260-277.
- [53] Rantamaa, H. R., Kangas, J., Kumar, S. K., Mehtonen, H., Järnstedt, J., & Raisamo, R. (2023). Comparison of a vr stylus with a controller, hand tracking, and a mouse for object manipulation and medical marking tasks in virtual reality. *Applied Sciences*, 13(4), 2251.
- [54] Neely, L., Carnett, A., Quarles, J., MacNaul, H., Park, S. W., Oyama, S., ... & Najafirad, P. (2023). The case for integrated advanced technology in applied behavior analysis. *Advances in Neurodevelopmental Disorders*, 7(3), 415-425.
- [55] Sánchez, B. H., & Cedeño, G. G. (2021, June). Uso y presencia de las tecnologías en personas con discapacidad. In *Congreso Nacional de Ciencia y Tecnología-APANAC* (pp. 329-335).

- [56] Skarbez, R., Smith, M., & Whitton, M. C. (2021). Revisiting Milgram and Kishino's reality-virtuality continuum. *Frontiers in Virtual Reality*, 2, 647997.
- [57] Waltemate, T., Gall, D., Roth, D., Botsch, M., & Latoschik, M. E. (2018). The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE transactions on visualization and computer graphics*, 24(4), 1643-1652.
- [58] Alaei, G., Deasi, A. P., Pena-Castillo, L., Brown, E., & Meruvia-Pastor, O. (2018, March). A user study on augmented virtuality using depth sensing cameras for near-range awareness in immersive vr. In *IEEE VR's 4th workshop on everyday virtual reality (WEVR 2018)* (Vol. 10, p. 3).
- [59] Lee, E., & Hannafin, M. J. (2016). A design framework for enhancing engagement in student-centered learning: Own it, learn it, and share it. *Educational technology research and development*, 64, 707-734.
- [60] Gutekunst, M., Geuss, M., Rauhoeft, G., Stefanucci, J., Kloos, U., & Mohler, B. (2014). A video self-avatar influences the perception of heights in an augmented reality oculus rift. In *24th International Conference on Artificial Reality and Telexistence: 19th Eurographics Symposium on Virtual Environments; Bremen, Germany, December 8th-10th, 2014* (pp. 9-12). Eurographics Association.
- [61] Perez, P., Gonzalez-Sosa, E., Kachach, R., Ruiz, J., Benito, I., Pereira, F., & Villegas, A. (2019, March). Immersive gastronomic experience with distributed reality. In *2019 IEEE 5th Workshop on Everyday Virtual Reality (WEVR)* (pp. 1-6). IEEE.
- [62] Gruosso, M., Capece, N., & Erra, U. (2021). Exploring upper limb segmentation with deep learning for augmented virtuality.
- [63] Boudouraki, A., Elmimouni, H., Orduna, M., Pérez, P., Gonzalez-Sosa, E., Cesar, P., ... & Fischer, J. E. (2023, October). Emerging Telepresence Technologies for Hybrid Meetings: an Interactive Workshop. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (pp. 547-552).
- [64] Elmimouni, H., Marcin, J., Rae, I., Hansen, J. P., Orduna, M., Read, J. C., ... & Sabanovic, S. (2022). Emerging Telepresence Technologies in Hybrid Learning.
- [65] Rae, I., Takayama, L., & Mutlu, B. (2013, April). In-body experiences: embodiment, control, and trust in robot-mediated communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1921-1930).
- [66] Grondin, F., Lomanowska, A. M., & Jackson, P. L. (2019). Empathy in computer-mediated interactions: a conceptual framework for research and clinical practice. *Clinical Psychology: Science and Practice*, 26(4), e12298.
- [67] Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 23, 1515-1529.
- [68] Salminen, M., Järvelä, S., Ruonala, A., Harjunen, V. J., Hamari, J., Jacucci, G., & Ravaja, N. (2019). Evoking physiological synchrony and empathy using social VR with

biofeedback. *IEEE Transactions on Affective Computing*, 13(2), 746-755.

- [69] Zhao, B., Wang, Z., Yu, Z., & Guo, B. (2018, October). EmotionSense: Emotion recognition based on wearable wristband. In 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI) (pp. 346-355). IEEE.
- [70] Jang, E. H., Park, B. J., Kim, S. H., Eum, Y., & Sohn, J. H. (2014, May). A study on analysis of bio-signals for basic emotions classification: recognition using machine learning algorithms. In *2014 International Conference on Information Science & Applications (ICISA)* (pp. 1-4). IEEE.
- [71] Blechert, J., Lajtman, M., Michael, T., Margraf, J., & Wilhelm, F. H. (2006). Identifying anxiety states using broad sampling and advanced processing of peripheral physiological information. *Biomedical sciences instrumentation*, 42, 136-141.
- [72] Schleifer, L. M., Spalding, T. W., Kerick, S. E., Cram, J. R., Ley, R., & Hatfield, B. D. (2008). Mental stress and trapezius muscle activation under psychomotor challenge: A focus on EMG gaps during computer work. *Psychophysiology*, 45(3), 356-365.
- [73] Angelini, L., Bompard, C., Abou Khaled, O., & Mugellini, E. (2023, July). Detecting Stress in VR 360° Immersive Experiences for Older Adults Through Eye Tracking and Psychophysiological Signals. In *International Conference on Human-Computer Interaction* (pp. 511-521). Cham: Springer Nature Switzerland.
- [74] Campanella, S., Altaieb, A., Belli, A., Pierleoni, P., & Palma, L. (2023). A method for stress detection using empatica E4 bracelet and machine-learning techniques. *Sensors*, 23(7), 3565.
- [75] Ramón Millán. “El brillante futuro de las redes privadas 5G”. *Revista BIT*, nº 229, 2023.
- [76] Straits Research, “Private 5G Network Market Forecast, Size, Report to 2031”, report SRTE1709DR.
- [77] 3GPP, TR 22.804, “Technical specification group services and system aspects; Study on communication for automation in vertical domains (release 16),” Jul. 2020.
- [78] 3GPP TR 23.734, “Study on enhancement of 5G system (5GS) for vertical and local area network (LAN) services (Release 16)”, Jun. 2019.
- [79] Wen, M., Li, Q., Kim, K. J., López-Pérez, D., Dobre, O. A., Poor, H. V., ... & Tsiftsis, T. A. (2021). Private 5G networks: Concepts, architectures, and research landscape. *IEEE Journal of Selected Topics in Signal Processing*, 16(1), 7-25.
- [80] Aijaz, A. (2020). Private 5G: The future of industrial wireless. *IEEE Industrial Electronics Magazine*, 14(4), 136-145.
- [90] Müller, M. K., Ademaj, F., Dittrich, T., Fastenbauer, A., Ramos Elbal, B., Nabavi, A., ... & Rupp, M. (2018). Flexible multi-node simulation of cellular mobile communications: the Vienna 5G System Level Simulator. *EURASIP Journal on Wireless Communications and Networking*, 2018, 1-17.

- [91] Wehrle, K., Günes, M., & Gross, J. (Eds.). (2010). *Modeling and tools for network simulation*. Springer Science & Business Media.
- [92] Matlab, "5G Toolbox" 5G Toolbox, Available: <https://www.mathworks.com/products/5g.html>. [Accessed: 13/08/2024].
- [93] Viridis, A., Stea, G., & Nardini, G. (2015). Simulating lte/lte-advanced networks with simulte. In *Simulation and Modeling Methodologies, Technologies and Applications: International Conference, SIMULTECH 2014 Vienna, Austria, August 28-30, 2014 Revised Selected Papers* (pp. 83-105). Springer International Publishing.
- [94] Nardini, G., Sabella, D., Stea, G., Thakkar, P., & Viridis, A. (2020). Simu5g—an omnet++ library for end-to-end performance evaluation of 5g networks. *IEEE Access*, 8, 181176-181191.
- [95] Polaris Networks, "NetTest 5G Network Emulator", NetTest 5G Network Emulator, <https://www.polarisnetworks.net/5g-network-emulators.html> [Accessed: 13/08/2024].
- [96] Keysight, "5G Network Emulation Solutions", 5G Network Emulation Solutions, Available: <https://www.keysight.com/zz/en/solutions/5g/5g-network-emulation-solutions.html> [Accessed: 13/08/2024].
- [97] Esper, J. P., Abdallah, A. S., Clayman, S., Moreira, W., Oliveira, A., Correa, S. L., & Cardoso, K. V. (2020). eXP-RAN—An emulator for gaining experience with radio access networks, edge computing, and slicing. *IEEE Access*, 8, 152975-152989.
- [98] Kachach, R., Orduna, M., Rodríguez, J., Pérez, P., Villegas, Á., Cabrera, J., & García, N. (2021, July). Immersive telepresence in remote education. In *Proceedings of the International Workshop on Immersive Mixed and Virtual Environment Systems (MMVE'21)* (pp. 21-24).
- [99] Gonzalez Morin, D., Gonzalez-Sosa, E., Perez, P., & Villegas, A. (2023). Full body video-based self-avatars for mixed reality: from E2E system to user study. *Virtual Reality*, 27(3), 2129-2147.