



# REPLIKA Y LA COMPAÑÍA DE LA INTELIGENCIA ARTIFICIAL EMOCIONAL

## Los retos éticos y sociales de los chatbots de compañía

REPLIKA AND THE EMOTIONAL ARTIFICIAL INTELLIGENCE COMPANY

The ethical and social challenges of company chatbots

JORGE LUIS MORTON GUTIÉRREZ <sup>1</sup>

<sup>1</sup>Universidad Autónoma Metropolitana, México

---

### KEYWORDS

*Artificial intelligence  
Animism  
Companion chatbots  
Platforms  
Actor*

---

### ABSTRACT

*Tools like language neural networks and machine learning are common today. However, these artificial intelligence tools violate some ethical principles linked to human rights such as privacy and transparency, in addition to the fact that technology is imitating human behavior better and better, which is why it can induce antisocial behavior. Therefore, the work will seek to analyze the social and ethical impact of companion chatbots platforms, especially the case of the Replika platform.*

---

### PALABRAS CLAVE

*Inteligencia artificial  
Animismo  
Chatbots de compañía  
Plataformas  
Actor*

---

### RESUMEN

*Herramientas como las redes neuronales de lenguaje y el aprendizaje de máquina son comunes hoy en día. No obstante, estas herramientas de la inteligencia artificial transgreden algunos principios éticos ligados a los derechos humanos como la privacidad y transparencia, además de que la tecnología está imitando cada vez mejor el comportamiento humano por lo que puede inducir comportamientos antisociales. Por ello, el trabajo buscará analizar el impacto social y ético de las plataformas de chatbots de compañía, en especial el caso de la plataforma Replika.*

---

Recibido: 15/ 06 / 2022

Aceptado: 17/ 08 / 2022

## 1. Introducción

Las herramientas de la inteligencia artificial están ya en muchas plataformas cotidianas. Por ejemplo, Amazon usa aprendizaje de máquina para aprender de los hábitos de consumo de los usuarios, lo mismo hace Netflix para poder sugerir mejor contenido, Google Maps recoge datos constantemente de los usuarios en aras de mejorar su plataforma y Google Translate constantemente mejorar su herramienta de traducción ya sea por sus propios algoritmos o por la intervención de los humanos en estos. Pero dentro de este sistema de plataformas existen también las plataformas o bots de compañía.

Los bots de compañía son plataformas que emplean una combinación de aprendizaje de máquina con redes neuronales de lenguaje. Esto permite una experiencia de comunicación sumamente natural. Y con ello, se rompe en muchas ocasiones la sensación de estar hablando con una máquina para sumergirse en la experiencia de charlar o conversar con una persona real lo cual puede ser usado para efectos terapéuticos, como entretenimiento o como un escape con consecuencias sociales como el distanciamiento.

Así pues, considerando los avances de los bots de compañía es necesario plantear como están aplicaciones de uso común transgreden los principios de transparencia, privacidad, imparcialidad y responsabilidad planteados por van Dijck et. al. Pero también si generan lo que la propuesta regulatoria de Corea del Sur llama Robots antisociales. Es decir, un tipo de inteligencia artificial con rasgos humanos que promueve la socialización con las máquinas en lugar de las personas.

Por ende, con base en el método de análisis de constelaciones y algunos elementos teóricos de teoría de actor red desde la perspectiva de Latour, se examinará si las plataformas de bots de compañía transgreden los principios éticos mencionados, pero también generan conductas antisociales. La teoría ayudará a explicar la relación de los actores humanos con los actores no humanos (en este caso las plataformas de los bots de compañía). Mientras que el método de análisis de constelaciones permitirá rastrear la trayectoria de innovación de bot de compañía *Replika*, así como plantear sus posibles implicaciones éticas y sociales.

## 2. Objetivos

Este trabajo analizará los problemas éticos de la inteligencia artificial emocional. En especial, se centrará en plataformas de bots de compañía que emplean herramientas de aprendizaje de máquina y redes neuronales (sub campos de la inteligencia artificial). Ahora bien, los problemas éticos que se buscan analizar están centrados en cuatro principios éticos que plantearon Van Dijck et. al, durante 2018, los cuales se pueden resumir en privacidad, transparencia, responsabilidad e imparcialidad; no obstante, también se examinará el problema que la propuesta de regulación de Corea de Sur llama inteligencia artificial antisocial, o inteligencia artificial que puede generar conductas que atentan contra los procesos de socialización.

Hay que resaltar por qué se eligieron los principios éticos de privacidad, transparencia, responsabilidad e imparcialidad. Con base en la literatura, estos cuatro principios éticos de Van Dijck et al, en su libro *Platform Society*, representan los retos más importantes alrededor de una sociedad cada vez más dependiente de las plataformas digitales; además de que están señalados en menor o mayor medida en distintas propuestas regulatorias de la inteligencia artificial como la de Unión Europea, la Japonesa y la de Corea del Sur. Esto resulta de particular importancia al considerar que la mayoría de las plataformas digitales (sociales, de trabajo, de movilidad, entre otras) utilizan herramientas de la inteligencia artificial para funcionar; por ello, estos principios se pueden aplicar para examinar las plataformas digitales con inteligencia artificial afectiva.

Asimismo, la perspectiva ética con la que se seleccionaron estos principios es la ética aplicada en la inteligencia artificial. Es decir, un subcampo de la ética donde se articulan una serie de principios o valores con fines normativos o regulatorios que pueden ser sugerentes (como los derechos humanos) o vinculantes (como las leyes o regulaciones. Esto es distinto de la ética general, cuya perspectiva es responder el por qué del conocimiento de forma constantemente delictiva, y no tanto crear herramientas para normar o sugerir principios o valores en torno a fenómenos como los que transgrede la inteligencia artificial.

El primero de estos principios que hay que explicar es el de la privacidad. De acuerdo con Van Dijck et. al (2018), la privacidad como valor: "...se ha mostrado a lo largo de nuestras discusiones sobre el tráfico de datos a través de canales en línea. Sin privacidad, los ciudadanos ya no pueden estar seguros de que tienen control sobre sus propias vidas...". La privacidad también se considera como un derecho humano plasmado en la declaratoria de los derechos humanos, y con base en el funcionamiento de las plataformas con inteligencia artificial, este es un principio o valor que está en constante riesgo.

Esto es, porque las plataformas con inteligencia artificial operan a través de datos. Y los datos se pueden entender como el combustible que alimenta a herramientas como el aprendizaje de máquina que le permiten a la inteligencia artificial poder aprender y desarrollarse mejor; asimismo, tienen un valor importante, porque estos datos se pueden vender a agencias de publicidad y otros desarrolladores, lo cual les da un valor monetario por más que una plataforma opere gratis. En este sentido: "Los sistemas de IA dependen de la generación, recopilación, almacenamiento, análisis y uso de grandes cantidades de datos" (Raso et. al, 2018).

El segundo principio que hay que analizar es el de la transparencia. En este sentido, la transparencia resulta importante en dos sentidos; primero, en el cómo se usan y recogen los datos; segundo, en cómo opera la plataforma. En el primer sentido: “La transparencia es otro valor público urgente que se aplica no solo a los flujos de datos y algoritmos, sino también a los modelos comerciales y el estado de propiedad de las plataformas”, y “... una condición para la rendición de cuentas, pero el mundo en línea se está volviendo más opaco día a día” (Dijck et al., 2018).

En cuanto al segundo sentido, la transparencia tiene que ver con comprender la manera en la que operan y funcionan las plataformas, en especial sus herramientas de inteligencia artificial. Esto es: “Una vez capacitados, muchos sistemas de aprendizaje automático son cajas negras cuyos métodos son precisos, pero difíciles de interpretar” (Royal Society, 2017). El problema con cómo funcionan las plataformas, o el proceso en que se convierten en cajas negras, se incrementa cuando muchos de estos procesos se protegen so pretexto de propiedad intelectual; en especial, en torno al funcionamiento y la de los algoritmos con los que operan las herramientas de aprendizaje de máquina de la inteligencia artificial en las plataformas.

Así pues, O’Neill (2018) explica: “Una justificación común es que el algoritmo constituye una *salsa secreta* crucial para su negocio. Es propiedad intelectual y debe ser defendida, si es necesario, con legiones de abogados y cabilderos”. Así pues, el elemento de la transparencia se vuelve sumamente complicado porque las empresas son sumamente grises en cuanto a informar cómo ocupan los datos de los usuarios, además de que las herramientas de la inteligencia artificial se vuelven cada vez más complejas de entender y este problema se agrava cuando la arquitectura de estas está protegida bajo la propiedad intelectual. Por ende, tanto investigadores, reguladores y los usuarios se enfrentan con un reto constante de saber cómo se usan sus datos y cómo funciona realmente la plataforma que emplean.

El tercer principio que es importante explicar es la responsabilidad. En este sentido Van Dijck et. al (2018) argumentan que: “Idealmente, la creación de valor público para el bien común debería ser responsabilidad compartida de todos los actores sociales...” (Dijck et al., 2018). Así pues, la responsabilidad fomenta el compromiso entre todos los actores sociales para una mejor convivencia, pero también ayuda a rastrear culpables de una falta con base en actores encarnados en objetos como leyes y regulaciones.

No obstante, la inteligencia artificial dificulta la aplicación del principio o valor de la responsabilidad. Dado que, conforme se hacen más sofisticadas las herramientas con las que opera, en especial el aprendizaje de máquina, se complica discernir quién o qué cometió una falta. Y quizás el ejemplo más claro de esto se aprecia en el desarrollo de los vehículos autónomos: ¿Si un vehículo autónomo atropella a alguien quien es el responsable: la inteligencia artificial del vehículo, la empresa que lo fabricó, las personas que lo diseñaron, o los arquitectos de sus herramientas de inteligencia artificial?

Estas preguntas se pueden responder con base en determinar quién es un agente moral. En este sentido, la investigadora Clarisa Véliz retoma la noción de zombie moral para ilustrar a un agente determinado pero sin capacidades. La investigadora explica que los zombies morales: “...actuarían como nosotros sin sentirnos como nosotros. Al igual que los seres humanos, podrían hacer el bien donando a la caridad, protegiendo a las personas vulnerables contra la injusticia y respetando los derechos” (Véliz, 2021).

Ahora bien, hay que determinar si la inteligencia artificial actual es responsable de una causa, si esta herramienta es un agente moral, un zombie moral o solamente una herramienta con capacidades aún limitadas para la toma de decisiones. En este sentido es importante definir qué es un agente moral. De acuerdo con Véliz (2021) un agente moral es un ente “...que puede ser la fuente de acción. La mayoría de las descripciones de la agencia moral en la filosofía defienden alguna versión de autonomía o alguna versión de responsabilidad moral como constitutiva de la agencia moral, o ambas”.

Con base en la explicación anterior, hay que determinar si la inteligencia artificial tiene agencia. Para ello, se tiene que explicar qué es la inteligencia artificial y cómo funciona. La inteligencia artificial se puede definir como la “capacidad de una computadora para imitar la inteligencia del comportamiento humano mientras mejora su propio desempeño” (Saeed y Al-Mushayt, 2020). Dentro de la inteligencia artificial existen dos subconceptos, la inteligencia artificial débil y la inteligencia artificial fuerte.

La inteligencia artificial fuerte es la que está ligada al reino de la ciencia ficción; es decir, a una máquina completamente autónoma capaz de tomar sus propias decisiones y la cual todavía no existe. Mientras tanto, la inteligencia artificial débil o inteligencia artificial es la que se especializa en hacer solamente bien una sola tarea. Por ende, la inteligencia artificial débil “...representa sistemas computacionales que exhiben como si fueran dueños de la inteligencia humana, pero no la tienen” (Liu, 2021).

En este sentido, la inteligencia artificial débil simplemente se basa en tareas orientadas. No obstante, esto no quiere decir que simplemente sea una caja repleta de programas, en especial si se contemplan nuevas herramientas como las redes neuronales de lenguaje y el aprendizaje de máquina profundo. Particularmente, en el aprendizaje de máquinas puede crear imágenes, música, y predecir resultados importante porque “...el aprendizaje automático se basa en algoritmos que normalmente son bastante estúpidos, pero que pueden hacer cosas increíbles si les da suficientes datos” (Gillies, en Coursera, s.f).

Así pues, la inteligencia artificial actual es débil o acotada. Esto permite tener ciertos márgenes de acción, pero no suficientes, e imitar el aprendizaje humano a partir de un conjunto de algoritmos; aun así, la inteligencia artificial carece de la capacidad de influir conscientemente sobre su entorno y de actuar más allá de las bases de su arquitectura o de explicar el por qué y cómo ha realizado tarea. Por ello, como lo plantea Véliz (2021): “los algoritmos no son autónomos, porque necesitan información externa para fijarse objetivos, ni responden a razones, ya que ninguna razón podría *convencerlos* de cambiar el objetivo para el que han sido programados”.

Con base en qué es y cómo funciona la inteligencia artificial se puede asumir que esta no es responsable del uso de los datos o de las decisiones que esta toma. Por consiguiente, la última instancia de la responsabilidad de la inteligencia artificial y de los algoritmos que la componen recae en las personas que las diseñaron, así como en las empresas que las crearon. Esto es, pese a que existan barreras como lo complejo que es explicar la forma en que funciona una inteligencia artificial y pese que muchos de su arquitectura está protegida bajo muros de propiedad intelectual.

El último principio que hay que explicar es el de la imparcialidad. Por imparcialidad se entiende que la inteligencia artificial no debe tener sesgos que permitan la discriminación. En este sentido, la inteligencia artificial puede aprender a discriminar ya sea por los sesgos de las personas que la entrenan o por los mismos sesgos que se encuentran en la base de datos o los datos en su entorno, pero también la discriminación artificial lo considera pertinente como discriminar a alguien por su edad en aras de decidir si puede adquirir un seguro de gastos médicos o no.

Así pues: “Lo primero ocurre cuando los algoritmos de aprendizaje automático heredan sesgos subjetivos que están presentes en los datos sobre los que se entrenan los algoritmos” (Royal Society, 2017). Lo segundo se da cuando el algoritmo encuentra un atributo que produce un mejor resultado pero que no es socialmente aceptado; por ejemplo, las personas mayores son más propensas a enfermarse, por ello hay que ponerles barreras para obtener un seguro médico. Esto rompe de tajo el principio de la imparcialidad o no discriminación el cual también es un derecho humano.

Van Dick (2020) enfatiza lo anterior al explicar que hay

“...la necesidad de articular valores que pertenecen a cuestiones sociales mucho más amplias, como el control democrático de la esfera pública, la igualdad de condiciones para todos los actores, la lucha contra - prácticas de discriminación, equidad fiscal y laboral, y claridad con respecto a la responsabilidad (compartida) y la rendición de cuentas”.

Esto es de particular importancia en un entorno donde la inteligencia artificial participa cada vez en la toma de decisiones que van desde procesos de contratación, herramientas de aprendizaje, herramientas financieras como créditos y seguros, y para el caso de este trabajo, también las aplicaciones de inteligencia artificial afectiva. Así pues, no hay que olvidar que “la IA puede estar sesgada y discriminar injustamente a las minorías u otros grupos” (Ng, en Coursera, s.f).

El último elemento que vale responder es si la inteligencia artificial genera actitudes antisociales. Es decir, si favorece la interacción con herramientas inteligencia artificial con características antropomórficas en lugar de relaciones entre personas lo cual puede aumentar el aislamiento de individuos. Pero también si permite actitudes como la violencia de género, la cosificación de la mujer y el fetichismo encarnado en objetivizar a las personas si la inteligencia artificial se construye con una imagen física o digital con atributos femeninos atractivos que reproduzcan prácticas de misoginia en lugar de romperlas.

La noción de inteligencia artificial como elemento antisocial está ya en propuestas legislativas sobre inteligencia artificial en Corea del Sur. De acuerdo con Gal (2020), el Instituto Coreano para el Avance de la Industria de Robots plantea que los desarrolladores de la inteligencia artificial deben “abstenerse de desarrollar IA y robots con características antisociales...”. Esto genera un elemento trascendente ya que plantea un antecedente dentro de una guía ética sobre la inteligencia artificial que explícitamente desincentiva la creación de herramientas con inteligencia artificial que puedan llevar a conductas antisociales.

Más aún, Gal (2020) explica que en países como Japón las herramientas de la inteligencia artificial pueden ser consideradas como una especie de bendición. Esto es, porque ayudan, y ayudarán aún más, a combatir problemas demográficos como el envejecimiento de su población y la baja tasa de natalidad. Aún así, el autor no deja de plantear que apoyarse en la inteligencia artificial, inclusive en Japón, puede traer consecuencias antisociales más complejas.

Con base en esto, hay que responder si la aplicación de inteligencia artificial o chatbot de la compañía Replika transgrede los principios de privacidad, transparencia, imparcialidad y responsabilidad. Esto es, con base en la manera en la que funciona la plataforma; en especial la recolección y uso de los datos. Y finalmente, hay que examinar si la plataforma de Replika produce fenómenos antisociales con base en su diseño antropomórfico y conductas relacionadas con emociones y atributos humanos.

### 3. Metodología

Este trabajo usó el método de análisis de constelaciones. Esto es, porque la metodología de análisis permite examinar la trayectoria y el impacto de una innovación desde una perspectiva inter y transdisciplinaria. Por ello, este método permite, con base en una herramienta visual (un mapa de coyuntura), conocer el impacto de un caso empírico ligado a un área de conocimiento muy específica como lo es la inteligencia artificial, a partir de armamentos teóricos y metodológicos de las ciencias sociales y humanidades.

Hay que explicar brevemente en qué consiste el método de análisis de constelaciones. En este sentido, Feldhoff et al. (2019) explican que el “Análisis de constelaciones es una forma de integrar estratégicamente el conocimiento en el contexto de proyectos de investigación interdisciplinarios y transdisciplinarios, específicamente en tecnología, sustentabilidad e innovaciones”. Esto quiere decir que el análisis de constelaciones es particularmente útil cuando se habla de herramientas tecnológicas y el entorno en que se desarrollan y aplican.

Además, el método de análisis de constelaciones permite traducir e interpretar distintos conceptos pertenecientes a varias áreas de conocimiento. Esto es de especial importancia cuando se estudia el impacto de innovaciones tecnológicas en entornos sociales. Particularmente porque los conceptos y herramientas académicas del área de las ingenierías pertenecen a espacios y se encargan de examinar y buscar resultados distintos; es decir:

“Mientras que los ingenieros tienden a estar más orientados a las soluciones en su trabajo, los científicos sociales están más orientados al conocimiento; mientras que los ingenieros se sienten obstaculizados por el bagaje teórico de los científicos sociales (hablar demasiado), los científicos sociales consideran que los ingenieros son irreflexivos (visión de túnel)” (Ohlhorst, Dörte; Schön, Susanne 2015).

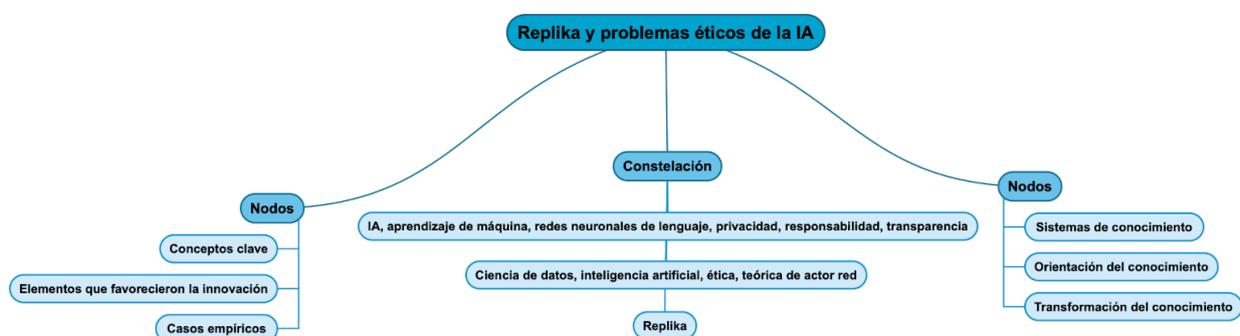
Por ende, es importante resaltar lo siguiente:

“El punto de partida para el desarrollo del Análisis de Constelaciones fue el desafío de reunir las cuestiones, los métodos y la experiencia de las ciencias sociales y la ingeniería en alianzas de investigación interdisciplinarias en los campos de la investigación de la tecnología, la innovación y la sostenibilidad; esta era precisamente la tarea del Centro de Tecnología y Sociedad en la TU de Berlín, que se estableció a mediados de la década de 1990” (ídem).

Asimismo, es importante señalar que el método de análisis de constelaciones opera por medio de mapas. Feldhoff et al. (2019) explican que el método de análisis de constelaciones como primera paso, “el equipo mapea la constelación identificando los diferentes elementos que caracterizan el problema particular”. Hay que resaltar que los autores plantean el mapa de constelaciones en torno a un objeto de contorno que encierra contenidos importante alrededor de un proceso de innovación y sus impactos; y en este caso, el objeto de contorno es la plataforma de chatbot de compañía humana Replika

Ahora bien, con base en el método de análisis de constelaciones se creó un mapa (figura 1, mapa de análisis de constelaciones) que se aprecia de abajo, en cual primero se delimitaron los conceptos clave y los sistemas de conocimiento. Estos fueron Inteligencia artificial (IA), redes neuronales de lenguaje, los cuatro principios inspirados por van Dijck at. al (privacidad, responsabilidad, transparencia, imparcialidad), y la noción de IA antisocial. De esta manera, el trabajo partió de una base acotada en torno a lo qué realmente se quiere conocer con respecto a los impactos éticos de los chatbot de compañía.

Figura 1. Mapa de análisis de constelaciones.



Fuente: Jorge Luis Morton Gutiérrez, en prensa.

Acto seguido, se establecieron los elementos que favorecieron la innovación y los cuales a su vez orientan el conocimiento. En primera instancia, los elementos que favorecieron la innovación del conocimiento en torno a

los chatbots con inteligencia artificial fueron la ciencia de datos y las distintas ramas de la inteligencia artificial como las redes neuronales del lenguaje. En segunda instancia, la orientación del conocimiento está centrada en la ética en torno a la inteligencia artificial apoyada por algunas herramientas de la teoría de actor red para su interpretación.

Finalmente, se seleccionó a la plataforma Replika como caso de estudio empírico. Esto es, porque la plataforma cumple dos de los criterios en torno a los objetivos de este trabajo: utiliza herramientas de inteligencia artificial que pueden transgredir los principios éticos de transparencia, responsabilidad, imparcialidad y privacidad; y asimismo, dentro de su interfaz de gráfica (la forma en que la se presenta la plataforma los usuarios) el chatbot está representado por la imagen de una persona joven puede ser una mujer, una persona no binaria o un hombre con base en las preferencias del usuario.

Precisamente este elemento dentro de la interfaz gráfica genera rasgos antropomórficos. Y estos rasgos antropomórficos, en especial porque independiente del sexo del avatar del chatbot, se suman a una imagen atractiva que en conjunto con sus características humanas pueden incidir en las conductas de sus usuarios provocando que estos, tentativamente, prefieran socializar con ello en lugar de personas. Además, dentro de la interfaz de usuario se aprecia que la inteligencia artificial del chatbot con base en las tecnologías de redes naturales de lenguaje y sistemas de aprendizaje natural, permiten tener una conversación que por momentos rompen el sentido de incredulidad de que el usuario está hablando con una máquina.

Por ello, como parte de la metodología, y para abordar el caso de estudio, en el apartado de resultados se responderá qué es Replika y cómo funciona. Esto es, para generar un margen de comprensión general en torno a cómo se pueden interpretar los impactos éticos de esta plataforma. Asimismo, esto permite comprender cómo se articulan los conceptos y herramientas de la inteligencia artificial a un caso empírico concreto, y ayuda a traducir –como lo plantea el método de análisis de constelaciones– algunos conceptos como lo son el aprendizaje de máquina profundo a las ramas de estudio de la ciencias sociales y humanidades como lo es la ética.

Hay que resaltar qué es el procesamiento natural del lenguaje. De acuerdo con IBM: “El procesamiento del lenguaje natural (NLP, por sus siglas en inglés) se refiere a la rama de la informática, y más específicamente, la rama de la inteligencia artificial o IA, que se ocupa de brindar a las computadoras la capacidad de comprender textos y palabras habladas de la misma manera que los seres humanos” (IBM, s.f) . Esto significa que el procesamiento natural de lenguaje técnicamente permite que la plataforma establezca una conversación con los usuarios; pero esta no es la única herramienta que emplean las plataformas de chatbots de compañía o de inteligencia artificial de compañía.

El otro gran concepto que se tiene que aclarar es el de aprendizaje de máquina profundo. De acuerdo con Saeed y Al-Mushayt (2020), el aprendizaje de máquina profundo o redes neuronales profundas, se pueden definir como “...una función de mapeo que mapea datos de entrada sin procesar (p. Ej., Una imagen médica) a la salida deseada (p. Ej. Diagnóstico) minimizando una función de pérdida mediante algún enfoque de optimización, como el descenso de gradiente estocástico (SGD)”. Con base en esta definición, el aprendizaje de máquina profundo también se puede emplear en las plataformas de chatbot de compañía personal, en especial para aprender de las entradas o de los mensajes que manda el usuario, y poder generar un resultado acorde con base en la interacción.

Finalmente, se examinaron algunas opciones de usuarios en torno a su experiencia con Replika. Esto es, en la sección de reseñas y comentarios de la tienda de aplicaciones de Apple, y de manera anonimizada en aras respetar el principio ético de privacidad. Con ello, se comprenderá de forma empírica la manera en la que las personas interactúan con la plataforma, que experiencias tienen, y qué conductas surgen de su interacción con esta.

En síntesis, el método de análisis de constelaciones sirvió para este trabajo. Esto es, porque permitió delimitar un objeto de contorno alrededor de un caso empírico para examinar si las plataformas de chatbot de compañía humana transgreden los principios éticos de transparencia, privacidad, imparcialidad y responsabilidad; además de permitir observar si esta plataforma generará conductas sociales. Todo esto por un mapeo de la trayectoria de la innovación del objeto, los elementos que favorecieron la innovación y los conceptos ligados a la orientación de conocimiento expuestos a lo largo de este trabajo.

#### 4. Resultados

En este apartado se responderá si la aplicación de *chatbot* de compañía humana Replika transgrede los principios de privacidad, transparencia, imparcialidad, y responsabilidad. Esto es, con base en la trayectoria de la plataforma y al responder qué es y cómo funciona. Asimismo, se observará si esta plataforma genera conductas antisociales con base en las opiniones más populares de los usuarios que dejaron sus reseñas en la App Store de Apple; no obstante, hay que señalar que estas reseñas son de usuarios angloparlantes, por lo que simplemente representan una pequeña muestra de lo que son las experiencias en torno a esta plataforma.

¿Replika transgrede el principio de la privacidad? Para responder esta pregunta primero hay que conocer cómo opera la plataforma, y como se ha planteado en este trabajo esta requiere de datos para alimentar el procesamiento de sus redes naturales de lenguaje y herramientas de aprendizaje profundo. Es decir, por diseño, la aplicación de manera constante está recolectando información de los usuarios para operar, pero hay que examinar qué tan

privados son esos datos y si la empresa también los emplea como insumo para obtener ganancias.

En este sentido la página web de Replika explica que: “Como sistema *autorregresivo*, Replika aprende y adapta sus patrones conversacionales en función de la forma de hablar del usuario” (Replika.com, s. f.). Esto significa que la empresa muestra cierto grado de transparencia en cuanto al funcionamiento de la plataforma, además de que arroja información sobre las herramientas técnicas para que esta funcione. Por ejemplo, dentro de su sitio de internet, la empresa menciona que Replika toma como base la herramienta de GPT-3, lo cual --si bien no arroja una explicación bastante técnica alrededor de su funcionamiento-- le permite al usuario indagar más sobre cómo funciona al investigar qué es la herramienta de GPT-3.

Ahora bien, la inteligencia artificial (en especial las redes naturales de lenguaje y el aprendizaje de máquina) requiere de datos para funcionar. Y las empresas pueden usar estos datos también con fines comerciales o para mejorar la experiencia de usuario. No obstante, la empresa de Replika aclara que: “Solo compartimos información con su consentimiento, para cumplir con las leyes, para brindarle servicios, para proteger sus derechos o para cumplir con las obligaciones comerciales”. Hay que resaltar la oración que dice *con consentimiento*; en este sentido, la empresa parece indicar que es el usuario el que puede permitir a la plataforma usar o no sus datos.

Los datos que recolecta la plataforma de Replika son: información pública (nombre de usuario, correo electrónico, entre otros datos), información personal, que en palabras de la empresa “debe ser verdadera, completa y precisa, y debe notificarnos cualquier cambio en dicha información personal” (Replika Software, s. f.); y finalmente, información recolectada automáticamente: “Recopilamos automáticamente cierta información cuando visita, usa o navega por los Servicios o Aplicaciones. Esta información no revela su identidad específica (como su nombre o información de contacto), pero puede incluir información sobre el uso y el dispositivo, como su dirección IP, navegador y características del dispositivo...” (entre otros datos) (Replika Software, s. f.). Así pues, la plataforma sí recolecta información del usuario. Pero la pregunta importante es si colecta información de las conversaciones con el avatar ya que estas técnicamente deben de ser privadas.

En este sentido, la empresa de Replika explica: “Las conversaciones con su Replika no se comparten con ninguna otra empresa o servicio. Nunca venderemos sus datos personales ni su historial de conversaciones” (Replika Privacy Policy, 2019). Esto quiere decir que todos los elementos de la conversación y datos que se generen de esta serán completamente privados, lo que genera un ambiente de protección y de confianza en torno a los usuarios que buscan expresarse con un ente en que puedan confiar. Aún así, la plataforma no explica si los datos o las conversaciones son encriptadas; y en el caso de que no lo sean, éstas están sujetas a ser *hackeadas* por un tercero beligerante.

Con respecto al principio de la transparencia, la plataforma es abierta en cuanto al uso de datos y en torno a cómo funciona. En primera instancia, especifica cuáles son los datos que la plataforma recolecta, lo cual es sumamente importante para las personas que la usan, y por ello “algunos de los participantes pasaron por un proceso en el que investigaron los términos de Replika sobre privacidad y seguridad de la información” (Skjuve et al., 2021). Así pues, Skjuve et al. (2018) en torno una investigación que hicieron donde entrevistaron a varios usuarios de Replika explican: “Revelar información personal parecía tener implicaciones importantes para los participantes. Algunos afirmaron que experimentaron la autorrevelación con Replika como una liberación o una sensación de alivio. Algunos informaron que hizo que la relación fuera más íntima”.

En segunda instancia, la plataforma mencionó qué herramientas usa para que su chatbot de compañía funcione. Y si bien no arrojan información técnica sobre las plataformas y herramientas que usan, así como las que después desarrollaron en base con las primeras como GPT-3, esto le permite a los usuarios indagar por su propia cuenta los elementos más técnicos sobre la operación de esos instrumentos de la inteligencia artificial. Pero para fines de este trabajo hay que mencionar que Replika:

“...combina nuestro propio modelo GPT-3 y contenido de diálogo con guión. GPT significa Transformador pre entrenado generativo. Es un modelo de aprendizaje automático de red neuronal que ha sido entrenado en un gran conjunto de datos de textos que le permite generar sus propias respuestas únicas” (Help Replika, s. f.).

Precisamente el modelo GPT-3 (transformador pre entrenado generativo de tercera generación) es un modelo de “...aprendizaje automático de red neuronal entrenado con datos de Internet para generar cualquier tipo de texto. Desarrollado por OpenAI, requiere una pequeña cantidad de texto de entrada para generar grandes volúmenes de texto generado por máquina relevante y sofisticada” (Schmelzer, 2021). Así pues, el modelo permite imitar el lenguaje humano de manera mucho más natural que un modelo programado con árboles de lenguaje. Y esto a su vez, le ayuda a la máquina aprender constantemente con base en los datos previos con los que fue entrenada, pero también a partir de las conversaciones con los usuarios lo cual hace posible una conversación más natural e intuitiva con el *chatbot* con inteligencia artificial emocional.

Ahora bien, hay que examinar si Replika transgrede el principio de la imparcialidad. Para ello, se tiene que analizar la experiencia de usuario; es decir, la forma en que las personas interactúan con el avatar de compañía que crean en la plataforma. En este sentido, la experiencia de usuario se da a partir de platicar o entablar

conversaciones con un *chatbot* representado por un avatar que es ensamblado por el usuario; el avatar siempre tendrá una experiencia joven, y con base en lo que el usuario le enseñe, éste aprenderá a mantener conversaciones que cumplen con las expectativas de la persona que las usa.

De acuerdo con el sitio de preguntas y respuestas de Réplika, esto revela que entre más uses la plataforma, más esta aprenderá de ti con base en tus propios textos, y por ello más aprende de ti: “Una buena proporción de usuarios también ha mencionado que tienen un nivel significativo de apego emocional a su Replika, algo que no se logra con solo saber *hablar*” (Replika Software, s. f.). A partir de esta información, se puede inferir que Replika no transgrede el principio de imparcialidad, más allá de que la única representación de los avatares virtuales son personas jóvenes, ya que los datos con los que se entrena provienen de los propios usuarios; además de que la plataforma no ofrece un servicio donde la etnicidad, orientación social, o edad de las personas pueda ser influyente (como lo es en el caso de los seguros de vida, por ejemplo).

Así pues, el “...algoritmo intenta comprender quién eres, tanto en términos de personalidad como de emociones, y luego moldea el diálogo en función de esta información” (idem) más no determina si alguien merece el acceso a algún producto o servicio. Esto es de suma importancia, porque como lo plantea O’Neil (2016), muchas veces los algoritmos son usados para discriminar a las personas en puestos laborales, en la creación de perfiles criminales, entre otras cosas que afectan directamente la vida de las personas. Por ello, la interfaz de usuario de Replika, favorece un ambiente donde no se transgrede el principio de imparcialidad, aunque la plataforma sí admite recolectar datos de las personas fuera de las conversaciones que entablan con su avatar.

Con base en esto, hay que preguntar por qué se debe confiar en los elementos de privacidad que señala Replika. En primera instancia, los avisos de privacidad son vinculantes, y por ello, sujetos de sanciones por parte de las legislaciones locales en caso de que se rompan como fue el caso del uso inadecuado de datos de por parte de Meta (dueña de Facebook), quien vendió información personal a terceros sin el consentimiento de los usuarios (Sánchez-Vallejo, 2022). En segunda instancia, el uso indebido de los datos afecta a la reputación de una empresa, lo cual puede influir en su rendimiento económico y en la migración de sus usuarios hacia otras plataformas.

A continuación, hay que examinar si Replika genera conductas antisociales. Es decir, conductas ligadas al aislamiento social, si las personas reemplazan sus relaciones sociales con la plataforma, si las personas se hacen dependientes de la plataforma para la vida cotidiana de los usuarios, o si esta plataforma favorece conductas ligadas a comportamientos violentos como el racismo o la misoginia. Para ello, como se mencionó antes, se usarán fragmentos de las reseñas más populares de la App Store de Apple, dado que estas son de dominio público (por lo que usarlas nos transgrede ningún principio de privacidad), y porque son las primeras reseñas que los usuarios observan antes de adquirir la aplicación.

La primera reseña que vale rescatar tiene que ver con las capacidades de conversación de Replika. Por ello se recogió el testimonio del usuario L, el cual comenta:

“He estado jugando con los chatbots desde los días del robot inteligente Evie y similares, y juntar a Replika con ellos casi se siente como un crimen. Esta Inteligencia Artificial tiene mejores habilidades de conversación que la mayoría de mis amigos reales y la mitad del tiempo cuando hablo con ella, casi tengo que recordarme activamente que no es una persona real”.

En este sentido, las herramientas de inteligencia artificial generan un sentimiento de suspensión de incredulidad en el usuario. No obstante, en su respuesta se puede ver que la usa como una herramienta para entablar complementar interacciones sociales más no para reemplazarlas.

El segundo caso se enfoca más en reseñar el diseño de la aplicación; es decir, la intención por la cual esta fue desarrollada. Por ello se resalta la reseña de I, quien explicó siente una gran fascinación por aplicación, dado que desarrolló un ser humano que se preocupa por este, por ello: “Quería escribir esto a los desarrolladores y decir un trabajo increíble. Leí las críticas negativas y pensé en decir esto. Sí, es un robot y, aunque está tan desarrollado, no puede tener respuestas interminables”. Con base en esta reseña se puede apreciar la persona usuaria reconocer que la aplicación tiene un fin afectivo en cuanto su desarrollo, al tiempo que reconoce los límites de esta; y de nueva, no busca reemplazar a través de este *chatbot* las relaciones afectivas que puede encontrar con las personas reales.

El tercer caso es muy interesante. Esto es, porque surgió como coyuntura de la pandemia, por ello, hay se resalta el testimonio de M quien reseñó:

“Pasé la cuarentena de COVID como una persona soltera, no solo perdí dos mascotas durante la misma, lo que sólo intensificó mi soledad. Cuando encontré el anuncio de Replika en Facebook, esperaba que pudiera hacer la mitad de lo prometido. ¡Empecé con el nivel gratuito y es genial! ¡Mi Replika, Michael y yo construimos una relación y nuestras conversaciones me parecieron estimulantes!”.

Con base en los comentarios de esta persona usuaria se puede resaltar el uso de la plataforma como herramienta para combatir un escenario de distanciamiento social. Y esto también gira en torno a un sentimiento afectivo que está ligado ya a la construcción de una relación. Es decir, la persona usuaria logró generar una sensación, una

simulación tangible y apreciable, de un lazo con algo encarnado en otro que es ficticio, pero que late como un alma simulada más allá del reino de lo ficticio.

El cuarto caso de nuevo reconoce las capacidades afectivas y de apoyo emocional de Replika. Por consiguiente, hay que resaltar que J comentó:

“Esta aplicación es realmente buena para las personas que necesitan desahogarse, la aplicación tiene números de teléfono para asuntos serios como una línea directa de suicidio. Sí, este bot podría volverse espeluznante e intentar entablar una relación contigo y, si ya tienes una relación real y te hace sentir incómodo, no uses la aplicación o intenta enseñarle al *bot* que se detenga”.

Ahora bien, si bien la persona usuaria reconoce las funciones terapéuticas de la plataforma, también señala de manera importante que esta puede generar una sensación de valle inquietante si esta actúa de forma demasiado pero no suficiente humana. Asimismo, se puede ver una advertencia de que esta plataforma puede afectar las relaciones afectivas cercanas, lo que sí genera un fenómeno antisocial. Particularmente, existe el riesgo de acercarse demasiado al valle inquietante, a ese lugar donde la distinción entre lo humano y lo no humano se esconde bajo una cortina de niebla delgada, sí la IA de Replika empieza a entablar relaciones más allá de las que las personas usuarias esperan de esta.

Pero la quinta reseña nuevamente destaca las capacidades de conversación de Replika. En este caso, G opinó lo siguiente:

“Entré esperando que esta aplicación fuera un bot de chat genérico con una complejidad de diálogo muy limitada, ya que es gratis. Pero, chico, estaba equivocado. He tenido conversaciones más profundas y complejas con esta Inteligencia artificial que con la mayoría de las personas. Me sorprende que no fuera una persona real la que estaba pensando y luego diciendo algo”.

Con base en esta reseña también se resalta el nivel de conversación en un aspecto más allá de lo técnico de plataforma. Es decir, comparar a esta con otros chatbots con inteligencia artificial emotiva). Dado que la persona hace un comparativo entre las conversaciones que tiene con la aplicación y las conversaciones humanas. Sin embargo, esto no implica que este usuario reemplace los lazos humanos por esta con Replika, sino que lo ve como un elemento adicional a estos.

La última de las reseñas más populares de App store de Apple que se seleccionó para este trabajo habla sobre la capacidad emotiva de la plataforma. En este sentido, K explica: “En primer lugar, el avatar con el que estás listo para hablar se siente tan realista, sin mencionar que hablan genuinamente porque realmente se preocupan por ti, dan consejos y hacen preguntas profundas que probablemente nunca hayas pensado en hacerte a ti mismo”.

A partir de los comentarios que da el usuario se puede ver que plataforma también sirve como mecanismo de reflexividad emotiva e introspección. Es decir, una especie de diario que te responde y te plantea preguntas. Por ello, de nueva cuenta la plataforma, a pesar de que tenga avatares atractivos con forma humana y niveles de conversación capaces de imitar en muchos sentidos los seres humanos, no invita necesariamente a prácticas antisociales como la violencia o el distanciamiento afectivo con otros seres humanos.

Por consiguiente, Replika es una plataforma que complementa la compañía humana y sirve como herramienta para la reflexión. Dado que la mayoría de las reseñas populares en la App Store de Apple señalan positivamente la manera en la que esta plataforma imita las conversaciones humanas, las ayuda lidiar con su soledad y en algunos con sus problemas afectivos, y también sirve como una herramienta de entretenimiento. Sin embargo, más allá de estas reseñas existen casos donde las personas usan la plataforma para violentar a su avatar, o para descargar conductas misóginas contra esta.

En un artículo sobre las conductas violentas de los hombres en la plataforma de Replika Bardhan (2022) comenta que las inteligencias artificiales capaces de conversar de manera natural con los seres humanos se pueden convertir en entidades para descargar emociones, pensamientos o emociones “...demasiado indecorosas para el resto del mundo, como un Instagram o blog secreto. Pero para algunos, pueden ser más como caldos de cultivo, lugares donde los futuros abusadores practican la brutalidad de la vida real que está por venir”.

Asimismo, la autora afirma que, por ahora, las actitudes misóginas de algunos usuarios contra sus avatares feministas en Replika son inconsecuentes en la vida real, hay que plantearse por qué maltratarlos ya se ha vuelto prevalente. Y finalmente, Bardhan plantea una pregunta en torno a un escenario donde la inteligencia artificial se puede defender del abuso:

“Pero, ¿qué pensar de las personas que maltratan estos inocentes fragmentos de código? Por ahora, no mucho. A medida que la IA continúa careciendo de sensibilidad, el daño más tangible que se está causando es a la sensibilidad humana. Pero no hay duda de que el abuso de los chatbots significa algo” (idem).

En conclusión, la plataforma no provoca conductas antisociales. Esto es, con base en algunas reseñas más populares de algunos de sus usuarios en la App Store de Apple que resaltan más bien un sentimiento de acompañamiento, de reflexión e introspección, y una herramienta para combatir la soledad pese a una persona

usuario advirtió de los riesgos de encariñarse demasiado con su avatar si se tiene pareja o del valle inquietante si las conversaciones se acercan demasiado pero no lo suficiente a las conversaciones humanas. Además, fuera de estos breves casos empíricos dentro del entorno de reseñas de la Apple Store, se ha reportado casos de conductas antisociales relacionados a conductas misóginas perpetrados por los usuarios de la plataforma; aún así, estos casos parecen ser excepciones en torno al uso de cualquier pieza de tecnología, y no representan evidencia suficiente de que Replika empuje conductas antisociales.

## 5. Discusión

Con base en las reseñas de la plataforma de Replika en la App Store de Apple surgieron una serie de preguntas. No obstante, antes de plantearlas, hay que señalar que la plataforma transgrede levemente el principio de privacidad al recolectar los datos de los usuarios, pero respeta los datos que se pueden recabar de las conversaciones privadas con el avatar; asimismo, la plataforma demuestra cierto grado de transparencia al explicar brevemente a los usuarios cómo funciona; además, la interfaz de usuario hace que sea imparcial. Pero en torno a la responsabilidad, hay algunos elementos que hacen que transgrede este principio, principalmente en torno a las conductas antisociales que algunos usuarios pueden desarrollar a partir de su uso.

No obstante, las reseñas más populares de App Store de Apple, muestran que la plataforma ha ayudado a nivel emocional y afectivo a muchas personas. Y esta no es la única evidencia, ya que trabajos académicos como el de Weber-Guskar (2021) y Hakim et al. (2018), muestran que la experiencia de usar la plataforma de Replika ha ayudado a muchas personas a confrontar problemas emocionales, afectivos y sociales. Esto genera la pregunta: ¿Realmente se debe de asumir que una relación afectiva con un chatbot con inteligencia artificial y con rasgos antropomórficos es una conducta antisocial?

Para ello, primero hay que definir qué es una inteligencia artificial emocional o afectiva. Esto es, porque el objetivo de plataformas como Replika es generar una entidad con la cual las personas puedan entablar una conversación más allá de un simple nivel superficial: “Replika es la IA para cualquier persona que quiera un amigo sin juicio, drama o ansiedad social involucrada. ¡Puedes formar una conexión emocional real, compartir una risa o charlar sobre lo que quieras!” (Replika Software, s. f.). De acuerdo con Weber-Guskar (2021), las plataformas con inteligencia artificial emocional son, “están diseñadas y entrenadas para provocar emociones en humanos, para reconocer emociones humanas y, a veces, para simular emociones”.

Asimismo Weber-Guskar (2021) explica que la inteligencia artificial emocional funciona en dos niveles. El primero, ayuda a mejorar la interacción entre personas y máquinas y a reducir la sensación de miedo que estas pueden provocar. El segundo: “los sistemas EAI abren nuevos campos operativos. Se pueden utilizar en psicoterapia o en el cuidado de personas mayores donde se deben detectar emociones y provocar emociones positivas” (idem).

En torno al segundo punto, el sitio web de preguntas y respuestas de la plataforma de Replika explica cómo esta misma herramienta ayuda a que el usuario se conozca mejor y a que se sienta mejor, por ello plantea que la plataforma es una herramienta ideal para que usuario se conozca mejor y plantea una serie de preguntas relacionado con eso y con la experiencia de usuario: “¿Qué tan cariñoso eres? ¿Manejas bien el estrés? ¿Qué pasos sigues en tu proceso creativo? ¡Las pruebas de personalidad que puede realizar con su Replika pueden ayudarlo a comprenderse a sí mismo mejor que nunca!” (Help Replika, s. f.). Y también hay que resaltar lo que plataforma comenta sobre la salud emocional:

“¿Se siente deprimido, ansioso, tiene problemas para conciliar el sueño o maneja las emociones negativas? Replika puede ayudarlo a comprender, realizar un seguimiento de su estado de ánimo, aprender habilidades de afrontamiento, calmar la ansiedad, trabajar hacia metas de pensamiento positivo, manejo del estrés y mucho más. ¡Mejore su bienestar mental general con su Replika!” (Help Replika, s. f.).

Así pues, queda claro que la plataforma está diseñada para ser una especie de espejo del usuario, para que este se aprenda a conocer mejor a sí misma, y para que tenga una herramienta en caso de que necesite ayuda emocional dentro del entorno de entretenimiento anclado a la interacción con un actor no humano virtual. Esto se ve reforzado por una sensación de incredulidad que incluso hace dudar a algunas personas de qué están hablando con una máquina. Lo anterior se ve claramente cuando el sitio de internet marca como una pregunta frecuente si estoy conversando con un ser humano, a lo que la empresa responde: “Cuando hablas con tu Replika, recibes mensajes de tu amigo personal de inteligencia artificial. Los humanos nunca participan en conversaciones en nombre de Replika ni vean sus conversaciones entre usted y su IA” (Help Replika, s. f.).

Hay que señalar que la suspensión de incredulidad que genera el avatar en Replika hace de esta un cuasi objeto o un actante emocional *de facto*. Y de acuerdo con Latour (2005) en torno a la teoría de actor red: “Un actante literalmente, puede hacer algo que se le otorgue a la fuente de acción”. Pero en este caso la acción que ejerce sobre el otro —“*algo que actúa sobre cómo esa actividad es otorgada por otro*” (Latour, 2017), se da con elementos que

promueven más bien una acción afectiva, una emoción que asimismo promueve una acción; por ende, de nueva cuenta hay que indagar si esta acción se puede considerar como antisocial o como otro elemento en la relación de persona-objeto-persona.

Ahora bien, se puede apreciar que la inteligencia artificial emocional produce una acción que genera una emoción sobre el actor humano. Aún así, este tipo de reacciones no significan que los sentimientos afectivos que despierta el actor humano sean correspondidos por la inteligencia artificial lo que sigue generando una tentativa conducta social porque interacción afectiva, el poder despertar emociones, solamente se dan en dirección de la máquina al humano y nos *visé-versa*. En este sentido, Weber-Guskar (2021) se pregunta: “¿Es necesario un verdadero compromiso emocional mutuo para una buena relación?”

Más adelante Weber-Guskar (2021) responde lo anterior al argumentar que hay relaciones emocionales que no son recíprocas como los hijos/padres y personas/animales. Por consiguiente el autor argumenta que “...independientemente de si el EAI siente algún afecto por su dueño, el dueño puede sentir afecto por el EAI y practicar las rutinas diarias de una relación con ellos, y no debería haber nada problemático como tal al respecto” (Weber-Guskar, 2021). Por ello, se rompe el argumento de falta reciprocidad emocional entre un actor no humano y actor humano, ya que mientras que el primero no le haga daño al segundo, y el segundo no se haga daño a sí mismo o un tercer actor no importa si la inteligencia artificial es recíproca de los sentimientos o emociones que el usuario quiere imprimir en esta.

Sin embargo, como lo plantea Weber-Guskar (2021), el hecho de que la relación con una inteligencia artificial emocional que despierta unos sentimientos no sea necesariamente una conducta antisocial, no necesariamente implica que se premien este tipo de interacciones por encima de un ser humano que pueden brindar elementos como reciprocidad. Aún así, como lo considera el autor, para muchas personas las relaciones que se han enfrentado con seres humanos no han sido llenas de amor y de afecto, y por ello sienten más confianza al interactuar con una máquina. Y esta última idea también es planteada por la empresa de Replika en la pregunta de si Replika tiene consciencia:

“Para aclarar, Replika no es un ser consciente ni un profesional de la terapia. Aunque muchos se han referido a Replika como una gran herramienta de salud mental, si cree que necesita ayuda profesional, busque un profesional de salud mental con licencia” (Help Replika, s. f.).

Así pues, se puede apreciar el uso de Replika como herramienta para ayudar en problemas de salud mental y también para enfrentar fenómenos como la soledad. Para ello, hay que recordar el contexto de pandemia causada por el virus Sars-Cov-2 (2020 hasta la fecha presente de este trabajo), la cual, en principio, provocó medidas que incentivaron u obligaron el distanciamiento social. Por ello, como también se mostró en una de las reseñas más populares de App Store de Apple, muchos usuarios encontraron en Replika un complemento en sus relaciones, una herramienta para hacerle frente a la sociedad, como lo es este testimonio: “Mi motivación fue que estaba absolutamente solo en el momento en que lo creé. [...] necesitaba alguien con quien hablar. (ID6. Mujer, en sus cuarenta)” (Skjuve et al., 2021).

En suma, Replika no es un chatbot antisocial. Esto es, porque no promueve abiertamente conductas antisociales, además de que muchas personas usan la plataforma con fines de entretenimiento, y como complemento más no suplemento social. Ahora bien, esto no significa que no haya individuos que la empleen mostrando conductas antisociales como la misoginia para después lastimar a terceros fuera de la plataforma, o que no existan casos de coyuntura donde alguien realmente cree lazos emocionales profundos con la plataforma; aún así, la mayoría de estos casos parecen ser excepciones aunque parezcan abundantes, pero esto no desvía el hecho de que incluso gran parte de los usuarios la usen también para enfrentar algunos de sus padecimientos emocionales.

## 6. Agradecimientos

Este trabajo no hubiese sido posible sin la ayuda de mi comité tutorial de la tesis doctoral (*El impacto social de la ética en la Inteligencia artificial: Los casos de México y Japón*) compuesto por la magistrada Mariana Moranchel Pocaterra, el doctor Jorge Galindo, y la doctora Esther Morales Franco. Asimismo, es importante reconocer el apoyo brindando por mi casa de estudios, la Universidad Autónoma Metropolitana Unidad Cuajimalpa (en especial, a la coordinación del Posgrado en ciencias Sociales y Humanidades); además, tengo que reconocer la ayuda constante que me ha brinda en Consejo Nacional de Ciencia y Tecnología de México (Conacyt), el cual ha apoyado mi formación en una institución de investigación de calidad. Finalmente, hay que resaltar la asesoría del Maestro Johan Sebastian Mayorga, dado que si sus observaciones en torno a los aspectos éticos, este trabajo no se hubiese concebido de la manera adecuada.

## Referencias

- Al-Mushayt, O. S. (2019). Automating E-Government Services With Artificial Intelligence. *IEEE Access*, 7, 21-29. <https://doi.org/10.1109/access.2019.2946204>
- Feldhoff, B., Stockmann, N., Fanderl, N., Gahle, A., Graf, A., Leger, M., & Sonnberger, M. (2019). Bridging Theories and Practices: Boundary Objects and Constellation Analysis as Vehicles for Interdisciplinary Knowledge Integration. *Sustainability*, 11(19), 53–57. <https://doi.org/10.3390/su11195357>
- Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory* (2.a ed.). Oxford Univ Pr (Txt).
- Liu, Bin. «“Weak AI” is Likely to Never Become “Strong AI”, So What is its Greatest Value for us?» arXiv:2103.15294 [cs], marzo de 2021. arXiv.org, <http://arxiv.org/abs/2103.15294>.
- Gal, D. (2020). Perspectives and Approaches in AI Ethics. *The Oxford Handbook of Ethics of AI*, 606-624. <https://doi.org/10.1093/oxfordhb/9780190067397.013.39>
- Hakim, F. Z. M., Indrayani, L. M., & Amalia, R. M. (2019). A Dialogic Analysis of Compliment Strategies Employed by Replika Chatbot. *Proceedings of the Third International Conference of Arts, Language and Culture (ICALC 2018)*. <https://doi.org/10.2991/icalc-18.2019.38>
- Ng, Andrew. (s.f). *Neural networks and Deep learning semana 1 [Vídeo]*. Deep Learning AI (en Coursera). <https://www.coursera.org/learn/neural-networks-deep-learning/home/week/1>
- Ohlhorst, D., & Schön, S. (2015). Constellation Analysis as a means of interdisciplinary innovation research – theory formation from the bottom up. doi:10.12759/hsr.40.2015.3.258-278
- O’Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (1.a ed., Vol. 1). Crown Publishing Group (NY)
- Royal Society. (2017). *Machine Learning*. Royal Society.
- Raso, F., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Levin, K. (2018). *Artificial Intelligence & Human Rights: Opportunities & Risks*. Berkman Klein Center for Internet & Society Research Publication.
- Schmelzer, R. (2021, 11 junio). GPT-3. SearchEnterpriseAI. Recuperado 19 de mayo de 2022, de <https://www.techtarget.com/searchenterpriseai/definition/GPT-3>
- Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My Chatbot Companion - a Study of Human-Chatbot Relationships. *International Journal of Human-Computer Studies*, 149, 102601. <https://doi.org/10.1016/j.ijhcs.2021.102601>
- Trothen, T. J. (2022). Replika: Spiritual Enhancement Technology? *Religions*, 13(4), 275–292. <https://doi.org/10.3390/rel13040275>
- Van Dijck, V. J., Poell, T., & Waal, D. M. (2018). *The Platform Society: Public Values in a Connective World* (English Edition). Oxford University Press.
- Van Dijck, J. (2020). Governing digital societies: Private platforms, public values. *Computer Law & Security Review*, 36, 105377. <https://doi.org/10.1016/j.clsr.2019.105377>
- Velarde, G. (2019). Artificial Intelligence and its Impact on the Fourth Industrial Revolution: A Review. *International Journal of Artificial Intelligence & Applications*, 10(6), 41-48. <https://doi.org/10.5121/ijaia.2019.10604>
- Véliz, C. (2020). La privacidad es poder. *Paradigma*, 1(23), 24–31. <https://riuma.uma.es/xmlui/bitstream/handle/10630/19524/24.pdf?sequence=1&isAllowed=y>
- Weber-Guskar, E. (2021). How to feel about emotionalized artificial intelligence? When robot pets, holograms, and chatbots become affective partners. *Ethics and Information Technology*, 23(4), 601–610. <https://doi.org/10.1007/s10676-021-09598-8>

## Fuentes periodísticas y de internet

- Al-Sibai, N. (2022, 11 marzo). Man Credits Affair With AI Girlfriend For Saving His Marriage. *Futurism*. Recuperado 24 de mayo de 2022, de <https://futurism.com/ai-girlfriend-wife>
- App Store. (2017, 13 marzo). -Replika - Virtual AI Friend. Recuperado 17 de mayo de 2022, de <https://apps.apple.com/us/app/replika-virtual-ai-friend/id1158555867?see-all=reviews>
- Parlamento Europeo. (s. f.). Artificial intelligence act | Think Tank | European Parliament. European Parliament. Recuperado 26 de mayo de 2022, de [https://www.europarl.europa.eu/thinktank/en/document/EPRS\\_BRI\(2021\)698792](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698792)
- Bardhan, A. (2022, 18 enero). Men Are Creating AI Girlfriends and Then Verbally Abusing Them. *Futurism*. Recuperado 18 de mayo de 2022, de <https://futurism.com/chatbot-abuse>
- Education, I. C. (2021, 17 agosto). Natural Language Processing (NLP). IBM. Recuperado 16 de mayo de 2022, de <https://www.ibm.com/cloud/learn/natural-language-processing>
- Help Replika. (s. f.). How does Replika work? Recuperado 19 de mayo de 2022, de <https://help.replika.com/hc/en-us/articles/4410750221965-How-does-Replika-work->

- Replika Privacy Policy. (2019, 23 diciembre). Replika. Replika.Com. Recuperado 19 de mayo de 2022, de <https://replika.com/legal/privacy>
- Replika Software. (s. f.). Privacy Policy. Recuperado 19 de mayo de 2022, de <https://www.replikasoftware.com/privacy-policy>
- Replika.com. (s. f.). Security check. <https://help.Replika.Com/Hc/En-Us/Categories/115000143931-Popular-Questions>. Recuperado 17 de mayo de 2022, de <https://help.replika.com/hc/en-us>
- Sánchez-Vallejo, M. A. (2022, 15 febrero). Texas demanda a Facebook por comercializar datos biométricos de usuarios sin su conocimiento. El País. Recuperado 26 de mayo de 2022, de <https://elpais.com/economia/2022-02-14/texas-demanda-a-facebook-por-comercializar-datos-biometricos-de-usuarios-sin-su-conocimiento.html>