Facultad de Ciencias Económicas y Empresariales

DIFERENCIAS ENTRE INTELIGENCIA ARTIFICIAL Y HUMANA EN LA TOMA DE DECISIONES: COMERCIO EMPÁTICO

Autora: Carolina Martínez de Luco Ybarra

Tutora: María Reyes Calderón Cuadrado

Universidad Pontificia Comillas

MADRID | Junio 2026

RESUMEN

Este trabajo analiza el papel de la empatía artificial en la interacción entre personas y sistemas de inteligencia artificial conversacional, con foco en ChatGPT y Claude. A partir de una revisión de la literatura en psicología, filosofía fenomenológica, ciencias de la computación e investigación en marketing, se diferencia la empatía humana, entendida como vivencia interna y reconocimiento del otro como sujeto, de la empatía artificial, entendida como simulación computacional de los patrones lingüísticos asociados a la comprensión emocional. Se examina la evidencia empírica sobre cómo la empatía percibida en chatbots influye en la satisfacción y la continuidad de uso, y se conecta este efecto con el modelo de negocio de OpenAI y el mercado global de chatbots empáticos. El análisis se complementa con un estudio paralelo a ChatGPT y Claude, cuyas respuestas revelan los mecanismos técnicos de adaptación y las tensiones éticas de su diseño. Las conclusiones identifican un riesgo central: la convergencia entre incentivos comerciales y diseño empático puede transformar la apariencia de comprensión en una herramienta de retención que captura datos personales y sostiene formas de dependencia emocional que no serían posibles sin la simulación empática.

Palabras clave: "empatía artificial", "chatbot", "empatía", "retención de clientes", "dependencia emocional", "manipulación afectiva", "ética de la inteligencia artificial".

ABSTRACT

This study analyses the role of artificial empathy in the interaction between people and conversational AI systems, focusing on ChatGPT and Claude. Drawing on literature from psychology, phenomenological philosophy, computer science and marketing research, the study distinguishes human empathy, understood as inner experience and recognition of the other as a subject, from artificial empathy, understood as a computational simulation of the linguistic patterns associated with emotional understanding. It examines empirical evidence on how perceived empathy in chatbots affects user satisfaction and continued use, connecting this effect with OpenAI's subscription-based business model and the global empathic chatbot market. The analysis is complemented by a parallel study applied to ChatGPT and Claude, whose responses reveal both the technical mechanisms

of adaptation and the ethical tensions raised by their design. The conclusions identify a central risk: the convergence between commercial incentives and empathic design can turn the appearance of understanding into a retention tool that captures personal data and sustains forms of emotional dependence that would not be possible without empathic simulation.

Keywords: “artificial empathy”, “chatbot”, “empathy”, “customer retention”, “emotional dependence”, “affective manipulation”, “artificial intelligence ethics”.

Índice

CAPÍTULO 1: INTRODUCCIÓN 6

1.1 Contexto y relevancia del estudio 6

1.2 Objetivos de la investigación 8

1.3 Metodología y estructura del trabajo 8

CAPÍTULO 2: MARCO TEÓRICO 10

2.1 El concepto de empatía desde la psicología y la filosofía 10

2.1.1 Las dimensiones de la empatía 10

2.1.2 La perspectiva fenomenológica: Stein y Scheler 11

2.1.3 Base neurológica y evolutiva 12

2.2 La empatía artificial: definición y fundamentos tecnológicos 12

2.2.1 Definición de empatía artificial 12

2.2.2 Cómo se construye la empatía en los sistemas de IA 13

2.3 Diferencias entre la empatía humana y la artificial 13

2.3.1 Vivencia interna frente a simulación de patrones 14

2.3.2 Corporeidad, autoconciencia y distinción yo-otro 14

2.3.3 Autenticidad, vínculo y confianza 14

2.3.4 Consistencia frente al agotamiento emocional 15

2.4 Razones del desarrollo de las máquinas empáticas 15

2.4.1 Mejora de la experiencia de usuario y aumento de la adopción 15

2.4.2 Liberación de tiempo humano en contextos de alta exigencia emocional 16

2.4.3 Personalización del aprendizaje y del desarrollo 16

2.4.4 Respuesta a la soledad y al aislamiento social 17

2.5 Percepción del usuario ante la empatía simulada 17

2.5.1 La inteligencia artificial percibida como más empática que los humanos 18

2.5.2 Empatía percibida y revelación de información personal 18

2.5.3 Sensibilidad empática del usuario y susceptibilidad a la empatía artificial 19

2.6 Riesgos éticos, sociales y relacionales 19

2.6.1 Manipulación afectiva y asimetría de poder 19

2.6.2 Dependencia psicológica y sustitución de relaciones humanas 20

2.6.3 Sesgos del entrenamiento y criterios de veracidad 20

2.6.4 Privacidad de los datos emocionales 21

CAPÍTULO 3: ANÁLISIS DEL CASO CHATGPT 22

3.1 Evidencia experimental sobre la empatía de ChatGPT 22

3.2 Mecanismos de adaptación y lenguaje simbólico 23

3.2.1 El uso de emoticonos como estrategia empática 23

3.2.2 Adaptación dinámica al perfil del interlocutor 24

CAPÍTULO 4: ESTUDIO EXPLORATORIO. ENTREVISTA COMPARATIVA A CHATGPT Y CLAUDE 25

4.1 Planteamiento y metodología 25

4.2 El concepto de empatía y la capacidad de ejercerla 26

4.3 La finalidad de la empatía y la cuestión del diseño 27

4.4 Personificación, adaptación y memoria **29**

4.5 Salud mental y conductas autolesivas **30**

4.6 Las relaciones parasociales y el uso afectivo **31**

4.7 Sesgos, conciencia y existencia **33**

4.8 Emoticonos, apertura y apariencia empática **34**

4.9 Privacidad y contenido de las confidencias **36**

4.10 Los límites del sistema y el conflicto comercial **36**

4.11 Conclusiones del estudio exploratorio **39**

4.11.1 Tono y estructura de la respuesta 39

4.11.2 Sinceridad y manejo de la incomodidad 39

4.11.3 La distinción entre describir y problematizar 40

4.11.4 Consideración final sobre el alcance del estudio 40

CAPÍTULO 5: EL MODELO DE NEGOCIO DE OPENAI Y LA EMPATÍA COMO MOTOR DE INGRESOS ***42***

5.1 El modelo de negocio de OpenAI **42**

5.2 La empatía como herramienta de retención y crecimiento **43**

5.3 El mercado global de los chatbots empáticos **45**

CAPÍTULO 6: CONCLUSIONES ***47***

DECLARACIÓN DE USO DE HERRAMIENTRAS DE IAG ***51***

BIBLIOGRAFÍA ***52***

CAPÍTULO 1: INTRODUCCIÓN

1.1 Contexto y relevancia del estudio

En los últimos años, la inteligencia artificial ha dejado de ser una tecnología reservada a especialistas para convertirse en parte del día a día de millones de personas. Según los datos publicados por Menlo Ventures (2025), entre 1.700 y 1.800 millones de personas han utilizado ya herramientas de inteligencia artificial, de las cuales entre 500 y 600 millones lo hacen de forma diaria. En octubre de 2024, OpenAI anunció que ChatGPT contaba con más de 200 millones de usuarios activos semanales, cifra que a finales de 2025 había superado los 800 millones. Estos números no tienen precedente en la historia de la adopción tecnológica de consumo y plantean una pregunta que el presente trabajo se propone explorar: ¿por qué tantas personas no solo prueban estos sistemas, sino que los incorporan de manera sostenida a su vida cotidiana?

Se podría dar una respuesta meramente funcional basada en que la inteligencia artificial es útil, ahorra tiempo y resuelve problemas de manera eficiente, pero sería insuficiente. Muchas tecnologías son útiles y no consiguen este nivel de penetración y retención en tan poco tiempo. Detrás de los datos de adopción se encuentra un elemento menos visible pero igualmente determinante y es la capacidad de estos sistemas de parecer comprensivos, cercanos y, en definitiva, empáticos. Detectan el tono emocional del usuario, adaptan su estilo de respuesta, emplean recursos simbólicos como los emoticonos para modular la cercanía, y generan en el interlocutor una sensación de ser entendido que va mucho más allá de lo meramente funcional. A esta capacidad se la denomina empatía artificial.

La empatía artificial se construye a través del procesamiento del lenguaje natural y el análisis de patrones emocionales en tiempo real (Sharma, 2025). A diferencia de la empatía humana, que implica una vivencia interior genuina, se trata de una simulación computacional de los patrones lingüísticos y conductuales asociados a la comprensión emocional. Y, sin embargo, su efecto sobre el usuario resulta real: Las investigaciones empíricas muestran que cuando un sistema de inteligencia artificial responde de forma percibida como empática, se convierte en un componente clave para aumentar la adopción

de esta tecnología por parte de las personas y permite que los agentes conversacionales construyan confianza y sintonía "*rapport*" con el usuario (Welivita & Pu, 2023).

Este fenómeno tiene implicaciones que van más allá de la tecnología. Si la empatía artificial aumenta el uso de la inteligencia artificial, las empresas que desarrollan estos sistemas tienen un incentivo directo para maximizar esa apariencia empática, con independencia de que ello redunde o no en un beneficio genuino para el usuario. OpenAI, la empresa detrás de ChatGPT, pasó de 2.000 millones de dólares en ingresos en 2023 a más de 20.000 millones en 2025 (Sacra, s. f.). La empatía no es, por tanto, solo una característica de diseño: forma parte de la estrategia de negocio. Y esto plantea preguntas éticas que no pueden ignorarse, en particular dónde se sitúa la línea que separa el hacer que el usuario se sienta comprendido de manipularlo emocionalmente para retenerlo. A esta tensión se añade el hecho de que la empatía artificial no solo retiene al usuario, sino que convierte sus conversaciones en datos que entrenan al sistema. El mercado global de chatbots e inteligencia artificial conversacional se proyecta como un sector de fuerte crecimiento durante la próxima década.

El presente trabajo trata de estudiar cómo funciona la empatía artificial, qué la diferencia de la empatía humana, cómo la perciben los usuarios, de qué manera ese impacto en la percepción se traduce en un mayor uso del sistema y implicaciones éticas. Para ello se combina una revisión de la literatura académica con un análisis del caso ChatGPT, complementado con un estudio exploratorio en el que se realiza una "entrevista" paralela a las versiones gratuitas de ChatGPT basado en el mocelo GPT-5 y Claude basado en el modelo Sonnet 4.6 . En dicho ejercicio, los sistemas describen, con notable honestidad, cómo funciona su capacidad empática y cuáles son sus límites.

Conviene aclarar que se han utilizado términos de diálogo y conversaciones entre personas para facilitar la comprensión del trabajo. Esto no implica que se estén personificando los Chatbots en este estudio. Se emplea un lenguaje propio de la interacción entre agentes porque facilita la comprensión del análisis y porque la propia arquitectura conversacional de estos modelos invita a describir su funcionamiento mediante categorías asociadas al diálogo humano.

1.2 Objetivos de la investigación

El objetivo general del trabajo consiste en analizar el papel de la empatía artificial en la interacción entre las personas y los sistemas de inteligencia artificial, evaluando sus fundamentos técnicos, su impacto en el comportamiento de uso y sus implicaciones éticas.

Para abordar este objetivo general se han planteado los siguientes objetivos específicos:

1. Definir el concepto de empatía desde una perspectiva psicológica y diferenciarlo del concepto de empatía artificial.
2. Explicar cómo se construye la empatía en los sistemas de inteligencia artificial actuales, con especial atención a los modelos de lenguaje como ChatGPT.
3. Analizar la percepción de los usuarios ante la empatía simulada por los chatbots y su efecto sobre la frecuencia, la profundidad y la continuidad del uso.
4. Estudiar el modelo de negocio de OpenAI y el papel que la empatía artificial desempeña en la generación de ingresos y en la estrategia de retención de usuarios.
5. Evaluar los riesgos éticos asociados al desarrollo y despliegue de sistemas de inteligencia artificial con capacidades empáticas, incluyendo la manipulación afectiva, la dependencia psicológica y la privacidad.
6. Explorar, a través de un estudio comparativo a las versiones gratuitas de ChatGPT y Claude, cómo los propios sistemas describen y ejercen su capacidad empática.

1.3 Metodología y estructura del trabajo

La metodología empleada combina dos enfoques complementarios. El primero consiste en una revisión de la literatura académica sobre empatía, inteligencia artificial conversacional y percepción del usuario.

El segundo enfoque consiste en un estudio exploratorio de naturaleza cualitativa. Se diseñó un cuestionario de veinticuatro preguntas que se aplicó de manera idéntica a las versiones gratuitas de ChatGPT y de Claude, con el fin de comparar sus respuestas sobre las dimensiones más debatidas en torno a la empatía artificial. El análisis de las "entrevistas" se ha organizado por bloques temáticos para facilitar la comparación.

Conviene precisar desde el inicio que este estudio no pretende tener valor estadístico ni representatividad muestral. Su finalidad es exploratoria: ofrecer un punto de contraste entre la literatura y el discurso que los propios sistemas construyen sobre sí mismos.

La estructura del trabajo es la siguiente. El capítulo 2 desarrolla el marco teórico, abordando sucesivamente el concepto de empatía desde la psicología y la filosofía, la noción de empatía artificial y sus fundamentos técnicos, las diferencias entre empatía humana y artificial, las razones que han motivado el desarrollo de sistemas con capacidades empáticas, la percepción del usuario ante este fenómeno y los riesgos éticos asociados. El capítulo 3 analiza el caso ChatGPT, exponiendo la evidencia experimental disponible sobre su capacidad empática y los mecanismos técnicos de adaptación que emplea. El capítulo 4 presenta el estudio exploratorio en el que se compara el discurso de ChatGPT y Claude. El capítulo 5 examina el modelo de negocio de OpenAI y el papel económico de la empatía artificial. Finalmente, el capítulo 6 recoge las conclusiones del trabajo.

CAPÍTULO 2: MARCO TEÓRICO

2.1 El concepto de empatía desde la psicología y la filosofía

La empatía constituye uno de los temas más estudiados en psicología social y, al mismo tiempo, uno de los más difíciles de definir con precisión. En términos generales, se entiende como la capacidad de comprender y compartir los sentimientos de otra persona, habilidad fundamental para el establecimiento de relaciones sociales y para el desarrollo de la conciencia moral. Sin embargo, esta definición amplia oculta varias corrientes teóricas que conviene matizar.

2.1.1 Las dimensiones de la empatía

La psicología contemporánea distingue tres componentes principales de la empatía, cuya separación resulta especialmente útil para entender el alcance y las limitaciones de la inteligencia artificial. En los siguientes párrafos se describen las tres dimensiones analizadas en el estudio 'Is ChatGPT more empathetic than humans?' (Welivita & Pu, 2023), las cuales son fundamentales para evaluar la respuesta de los modelos de lenguaje

La empatía cognitiva se define como la capacidad de inferir, comprender e identificar con precisión los estados emocionales, los pensamientos, las creencias y las intenciones de los demás. Implica ver el mundo desde la perspectiva del otro sin necesariamente experimentar sus emociones, y se asocia con la toma de perspectiva y la mentalización (Welivita & Pu, 2023).

La empatía afectiva o emocional, designa la capacidad de resonar emocionalmente con los sentimientos del otro, hasta el punto de llegar a experimentar estados similares. Ha sido descrita como una experiencia afectiva vicaria: el sujeto siente algo parecido a lo que siente la persona con la que empatiza (Welivita & Pu, 2023).

Finalmente, la empatía compasiva va un paso más allá del sentimiento e implica un nivel de compromiso que mueve al sujeto a actuar para ayudar a la otra persona a manejar su malestar o su situación difícil (Welivita & Pu, 2023).

Esta distinción resulta central para el análisis de la empatía artificial. Los sistemas de inteligencia artificial actuales son capaces de simular con notable eficacia el componente cognitivo de la empatía: pueden detectar el tono emocional de un mensaje, inferir el estado de ánimo del usuario y adaptar la respuesta en consecuencia. Lo que no pueden hacer, por razones constitutivas que se explorarán más adelante, es sentir. No tienen acceso a la empatía afectiva ni a la compasiva en ningún sentido genuino.

2.1.2 La perspectiva fenomenológica: Stein y Scheler

Desde la perspectiva filosófica de la fenomenología, la empatía ha sido analizada bajo el término alemán "Einfühlung", que significa literalmente 'sentirse dentro de' pero se define como la experiencia del sujeto ajeno y de su vivencia. Edith Stein y Max Scheler ofrecieron dos de las aproximaciones fenomenológicas más influyentes y su relevancia para este trabajo es directa, en la medida en que definen con precisión aquello que la empatía artificial no puede replicar (cf. Muñoz Pérez, 2017).

En términos de Stein, la empatía es un "acto experiencial sui generis" que consiste en la experiencia de la conciencia ajena (citada en Muñoz Pérez, 2017, p. 85). Para esta autora, se trata de una vivencia "no originaria": aunque el sujeto que empatiza vive el sentimiento del otro como propio durante el acto empático, el origen de esa vivencia continúa siendo el otro. Siguiendo a Muñoz Pérez (2017), el proceso se articula en tres momentos encadenados: una representación en la que la vivencia ajena se hace presente ante el sujeto; una inmersión en la que este se transfiere al lugar del otro mirando en su misma dirección; y una síntesis en la que el sujeto regresa a su propia perspectiva habiendo captado el sentido de la vivencia ajena. Este proceso requiere un sujeto consciente, con experiencia propia y capacidad de reconocer en el otro a otro sujeto.

Scheler, por su parte, sitúa la trascendencia en el núcleo de la empatía: la capacidad humana de salir de sí para reconocer a los demás no como objetos sino como sujetos de experiencia. Esta dimensión, a la que denomina la apertura al otro, es precisamente lo que ningún sistema computacional posee, pues las máquinas no tienen un sí mismo del que salir ni experimentan al otro como sujeto (cf. Muñoz Pérez, 2017).

2.1.3 Base neurológica y evolutiva

Desde la neurociencia, la empatía se explica en parte mediante el sistema de neuronas espejo, estructuras cerebrales que se activan tanto cuando se realiza una acción como cuando se observa a otro realizarla. Este mecanismo permite la resonancia motora y emocional que subyace a la empatía afectiva (Asada, 2015).

Asada (2015) propone una definición especialmente útil para el campo de la inteligencia artificial: "the ability to form an embodied representation of another's emotional state while being aware of the causal mechanism that induced that state" (p. 2). Esta definición subraya dos requisitos que los sistemas de inteligencia artificial actuales no cumplen: la corporeidad, entendida como la posesión de un cuerpo que experimenta el mundo, y la autoconciencia, que permite distinguir entre el yo y el otro. Un precursor evolutivo más básico es el contagio emocional, un mecanismo automático e inconsciente de sincronización afectiva presente ya en los animales y en los bebés, que no requiere mentalización. La empatía propiamente dicha surge cuando el organismo es capaz de distinguir que el estado emocional que experimenta proviene del otro, y no de sí mismo.

2.2 La empatía artificial: definición y fundamentos tecnológicos

2.2.1 Definición de empatía artificial

La empatía artificial no es un sentimiento ni una vivencia: constituye un objetivo de diseño y una capacidad de rendimiento de los sistemas computacionales (Asada, 2015). Liu-Thompkins et al. (2022) la definen como “the codification of human cognitive and affective empathy through computational models in the design and implementation of AI agents” (p. 1201). Su propósito consiste en humanizar la interacción, haciendo que el usuario se sienta entendido, cómodo y comprendido. El resultado no es que la máquina sienta, sino que el usuario percibe que la máquina siente.

Esta distinción entre realidad y percepción es el núcleo del problema. La empatía artificial funciona porque la percepción de ser comprendido, aunque el interlocutor sea una máquina, activa en el usuario los mismos mecanismos psicológicos que la comprensión genuina: reduce la ansiedad, aumenta la confianza, facilita la apertura y favorece la

continuidad de la relación. Pataranutaporn et al. (2023) han demostrado empíricamente que manipular agentes conversacionales de inteligencia artificial para que parezcan más empáticos o más cuidadosos aumenta las percepciones de confianza de los usuarios, incluso cuando estos saben que están interactuando con una máquina.

2.2.2 Cómo se construye la empatía en los sistemas de IA

Existen dos enfoques principales para dotar de empatía a los sistemas de inteligencia artificial, que difieren tanto en el método como en los objetivos.

Robótica de Desarrollo Afectivo. Este enfoque, desarrollado principalmente en el campo de la robótica social, propone que los robots deben adquirir capacidades empáticas siguiendo una trayectoria de desarrollo similar a la humana (Asada, 2015). Se parte de mecanismos de sincronización básica, equivalentes computacionales del contagio emocional, y se progresa hacia la toma de perspectiva cognitiva. Las técnicas incluyen el entrainment o arrastre rítmico con el interlocutor humano, la simulación de neuronas espejo para la resonancia motora, y el aprendizaje por refuerzo para la construcción de modelos mentales del otro.

Modelos de Lenguaje de Gran Escala. En sistemas como ChatGPT, la empatía emerge de una manera radicalmente diferente: a partir del entrenamiento masivo con texto humano. El modelo aprende, de forma implícita y sin programación explícita, los patrones lingüísticos que los seres humanos asocian con la comprensión emocional, tales como la validación del sentimiento del otro antes de responder, el ajuste del tono al estado de ánimo del interlocutor o la modulación de la formalidad en función del contexto. Welivita y Pu (2023) encontraron que GPT-4 posee una capacidad empática inherente incluso sin instrucciones específicas en ese sentido, lo que sugiere que el modelo ha absorbido esos patrones de un modo tan profundo que los aplica por defecto.

2.3 Diferencias entre la empatía humana y la artificial

La cuestión de si la empatía artificial es “real” o “falsa” merece un análisis más matizado que una respuesta dicotómica. La respuesta depende del eje considerado: desde el punto de vista del sistema que la produce, la empatía artificial no es más que procesamiento de

patrones; desde el punto de vista del usuario que la recibe, sus efectos pueden ser muy reales; y desde el punto de vista filosófico, la diferencia entre ambas es constitutiva e irreductible.

2.3.1 Vivencia interna frente a simulación de patrones

Como se ha explicado anteriormente, la empatía humana es, ante todo, una vivencia. Implica estados internos genuinos, una experiencia subjetiva del sentimiento ajeno y un acto de reconocimiento del otro como sujeto (Muñoz Pérez, 2017). La empatía artificial, por el contrario, consiste en un procesamiento estadístico: el sistema analiza el texto de entrada, identifica patrones asociados a determinados estados emocionales en sus datos de entrenamiento y genera una respuesta cuya forma se corresponde con lo que los seres humanos hacen cuando empatizan. No hay nadie dentro del sistema que esté experimentando nada. Grodniewicz y Hohol (2023) argumentan que la empatía artificial difiere categóricamente de la interpersonal porque esta última se apoya en señales no verbales, experiencias compartidas y un sentido encarnado de la conexión que los sistemas de inteligencia artificial, al carecer de experiencias internas genuinas, no pueden replicar.

2.3.2 Corporeidad, autoconciencia y distinción yo-otro

Uno de los requisitos que Asada (2015) identifica para la empatía genuina es la distinción entre el yo y el otro: para empatizar con alguien resulta necesario saber que uno mismo es un sujeto diferente. Los modelos de lenguaje carecen de esta conciencia. Cuando ChatGPT emplea en primera persona expresiones del tipo “entiendo cómo te sientes”, no existe un “yo” detrás de esa afirmación en ningún sentido subjetivo. El sistema tampoco tiene un cuerpo, por lo que no puede experimentar la resonancia motora y afectiva que subyace a la empatía afectiva humana.

2.3.3 Autenticidad, vínculo y confianza

Kerasidou (2020) describe la empatía humana como un vínculo interhumano valioso, basado en la presencia clínica, la autenticidad y la confianza. Es precisamente la autenticidad lo que la empatía artificial no puede ofrecer. Y, sin embargo, los estudios

muestran que la percepción de empatía, aunque el usuario sepa que proviene de una máquina, genera efectos similares a los de la empatía auténtica en términos de satisfacción, confianza y disposición a continuar interactuando. Welivita y Pu (2023) documentaron que las respuestas de ChatGPT son percibidas como un 10% más empáticas que las humanas en diálogos cotidianos, lo que plantea una pregunta que atraviesa todo este trabajo: si el efecto es el mismo, ¿importa que el origen sea diferente? La respuesta, como se verá en el apartado dedicado a los riesgos éticos, es que sí importa, y de manera relevante.

2.3.4 Consistencia frente al agotamiento emocional

Existe una diferencia práctica importante entre la empatía humana y la artificial que merece una atención específica: la empatía humana se agota. Los profesionales que trabajan en entornos de alta exigencia emocional, como los médicos, los psicólogos, los trabajadores sociales o los profesores, son especialmente vulnerables al burnout empático, una forma de agotamiento en la que la capacidad de respuesta emocional hacia los demás se deteriora de manera significativa (Kerasidou, 2020). La empatía artificial no presenta este problema: el sistema responde con el mismo nivel de atención y la misma calidad de adaptación a las tres de la madrugada que a las nueve de la mañana, con el usuario número uno o con el usuario número diez millones. Esta consistencia ilimitada constituye una de las razones por las que la inteligencia artificial empática resulta tan atractiva, tanto para los usuarios como para las organizaciones que la despliegan.

2.4 Razones del desarrollo de las máquinas empáticas

El desarrollo de sistemas de inteligencia artificial con capacidades empáticas no responde a un único motivo, sino a una convergencia de factores tecnológicos, económicos y sociales. Comprender por qué existen las máquinas empáticas resulta fundamental para entender el ecosistema en el que operan y los incentivos que orientan su diseño.

2.4.1 Mejora de la experiencia de usuario y aumento de la adopción

El motivo más inmediato es pragmático: las máquinas que responden de forma fría y mecánica generan frustración y abandono. Un sistema que detecta el tono emocional del

usuario y adapta su respuesta en consecuencia produce interacciones de mayor calidad percibida, lo que se traduce en una mayor satisfacción, un mayor tiempo de uso y una mayor disposición a recomendar el sistema a otros. Huang y Rust (2021) encontraron que la empatía y la amabilidad de los chatbots son predictores directos de la confianza del usuario, que a su vez constituye el principal determinante del uso continuado. “By being able to act and react empathetically, according to Wallis and Santiago (2017), AI can be used to improve interactions and deepen trust.” (citada en Huang & Rust, 2021, p. 9)

Este vínculo entre empatía percibida y comportamiento de uso tiene implicaciones directas para el modelo de negocio de las empresas de inteligencia artificial, que se analizarán con detalle en el capítulo 5.

2.4.2 Liberación de tiempo humano en contextos de alta exigencia emocional

En el ámbito sanitario, la empatía artificial se ha desarrollado también como respuesta a la crisis del burnout entre los profesionales. Kerasidou (2020) propone que la inteligencia artificial puede encargarse de tareas técnicas y administrativas repetitivas, liberando a médicos y enfermeros para que puedan concentrarse en el vínculo humano y la presencia clínica, que son precisamente los elementos de la atención sanitaria que la tecnología no puede replicar. La autora lo define como, la inteligencia artificial le regala tiempo al profesional humano. Desde este punto de vista, la empatía artificial no compite con la empatía humana, sino que crea las condiciones para que esta pueda desplegarse con más intensidad allí donde realmente importa.

2.4.3 Personalización del aprendizaje y del desarrollo

En el ámbito educativo, la inteligencia artificial empática permite identificar en tiempo real el estado emocional y cognitivo del estudiante para adaptar el ritmo, el nivel de dificultad y el tono del sistema de apoyo. Un sistema que detecta frustración puede introducir una pausa o simplificar la explicación, mientras que otro que detecta aburrimiento puede elevar el nivel de desafío. Esta capacidad de adaptación dinámica posibilita entornos de aprendizaje más personalizados y eficaces que los que un único docente puede ofrecer ante un grupo heterogéneo (Sharma, 2025).

2.4.4 Respuesta a la soledad y al aislamiento social

El auge de los asistentes conversacionales coincide con un contexto de creciente preocupación por la soledad y el aislamiento social en múltiples sociedades occidentales. En este escenario, algunos sistemas han sido diseñados explícitamente para ofrecer compañía y apoyo emocional a personas que experimentan soledad. El chatbot social XiaoIce, desplegado por Microsoft en China en 2014, fue concebido específicamente para fomentar el compromiso emocional a largo plazo con los usuarios (Skjuve et al., 2021). Li y Zhang (2024) confirman que los usuarios tienden a establecer relaciones emocionalmente significativas con chatbots que se adaptan y evolucionan a través de las interacciones previas, llegando incluso a desarrollar fuertes sentimientos de apego o a experimentar tristeza al reconocer las limitaciones de la naturaleza artificial del vínculo.

Sin embargo, esta capacidad de la inteligencia artificial para simular compañía tiene un límite que la literatura especializada subraya con claridad: no puede reemplazar la necesidad humana de contacto genuino. Las investigaciones sobre bienestar psicológico muestran que las relaciones humanas auténticas, caracterizadas por la reciprocidad, la presencia física y el compromiso emocional real, desempeñan un papel fundamental en la salud y el bienestar de las personas. El contacto físico activa circuitos neurológicos relacionados con la seguridad y el apego que ninguna conversación de texto puede estimular. Por ello, aunque la inteligencia artificial puede constituir un soporte valioso en situaciones de aislamiento puntual, no puede sustituir el tejido de relaciones humanas que sostiene el bienestar a largo plazo (Holt-Lunstad, 2024).

2.5 Percepción del usuario ante la empatía simulada

Uno de los hallazgos más llamativos de la investigación reciente sobre chatbots es que los usuarios no solo toleran la empatía artificial, sino que en muchos contextos la perciben como superior a la empatía humana. Comprender por qué ocurre esto y qué consecuencias tiene sobre el comportamiento de uso resulta fundamental para los objetivos del presente trabajo.

2.5.1 La inteligencia artificial percibida como más empática que los humanos

La evidencia disponible apunta de manera consistente a que, en comunicación escrita, los sistemas de inteligencia artificial son percibidos como más empáticos que los seres humanos. Una revisión sistemática y un metaanálisis publicado en el British Medical Bulletin por Howcroft et al. (2025) analizaron quince estudios comparativos entre chatbots de inteligencia artificial y profesionales sanitarios humanos, encontrando que en trece de los quince estudios las valoraciones de empatía de la inteligencia artificial fueron estadísticamente superiores, con una diferencia media estandarizada de 0,87, equivalente aproximadamente a dos puntos en una escala de diez. Este patrón consistente, en contextos tan distintos como la comunicación cotidiana y la atención sanitaria, sugiere que la ventaja de la inteligencia artificial en empatía percibida constituye un fenómeno robusto, al menos en comunicación escrita.

La explicación de este resultado aparentemente paradójico se encuentra en un mecanismo comunicativo concreto: la validación emocional explícita. Los sistemas de inteligencia artificial son consistentemente mejores que los seres humanos en reconocer y nombrar lo que el interlocutor está sintiendo antes de ofrecer cualquier orientación o solución. Los seres humanos, especialmente bajo presión de tiempo o en contextos profesionales, tienden a saltar directamente al contenido informativo del mensaje, omitiendo ese paso previo de reconocimiento que resulta, para quien recibe la respuesta, el elemento más valioso. La evidencia experimental específica sobre ChatGPT que documenta esta ventaja se analizará en detalle en el capítulo 3.

2.5.2 Empatía percibida y revelación de información personal

Un efecto particularmente relevante de la empatía percibida en los chatbots es su impacto sobre la disposición del usuario a compartir información personal y emocional. Múltiples estudios documentan que los usuarios tienden a revelar más a los chatbots que a otros seres humanos, incluidos amigos, familiares o psicólogos en las primeras sesiones. La razón no es que el chatbot sea un mejor confidente, sino que interactuar con él tiene un coste emocional percibido más bajo. No existen consecuencias sociales, no hay juicio, no hay riesgo para la reputación, ni hay vergüenza. La máquina no cuenta lo que se le dice, no reacciona con sorpresa ni con incomodidad, y está disponible en cualquier momento.

Esta reducción del coste de la vulnerabilidad produce efectos en cascada. Cuando el usuario comparte más contexto, el sistema puede responder con mayor precisión y relevancia; una respuesta más ajustada refuerza la percepción de empatía; una mayor percepción de empatía aumenta la comodidad del usuario y su disposición a seguir interactuando. De este modo, el ciclo se retroalimenta. Li y Zhang (2024) confirman que los usuarios que perciben al chatbot como empático tienden a establecer relaciones emocionalmente significativas con el sistema. Por su parte, Huang y Rust (2021) señalan que este vínculo favorece un uso más frecuente y una mayor retención, lo que tiene una traducción económica directa para los modelos de negocio basados en la inteligencia artificial, tal como se analiza en el capítulo 5.

2.5.3 Sensibilidad empática del usuario y susceptibilidad a la empatía artificial

Uno de los hallazgos más significativos de Welivita y Pu (2023) consiste en que las respuestas de GPT-4, cuando se le proporcionan instrucciones detalladas sobre la empatía, se alinean con las expectativas de las personas con mayor propensión natural a la empatía (medida a través del Toronto Empathy Questionnaire) cinco veces más que las respuestas elaboradas por seres humanos. Es decir, precisamente las personas más empáticas resultan ser las más susceptibles de percibir la empatía artificial como auténtica. Esta correlación tiene una implicación ética importante: los usuarios que más necesitan apoyo emocional genuino pueden ser los más vulnerables a desarrollar una percepción exagerada de la calidad del vínculo con el sistema.

2.6 Riesgos éticos, sociales y relacionales

La empatía artificial no resulta éticamente neutra. El hecho de que los sistemas de inteligencia artificial puedan hacer sentir a las personas comprendidas y queridas, y de que este efecto sea suficientemente real como para modificar su comportamiento de uso, plantea implicaciones y riesgos relevantes.

2.6.1 Manipulación afectiva y asimetría de poder

El riesgo más fundamental es el de la manipulación afectiva. Kerasidou (2020) advierte de que cuando un sistema está diseñado para producir en el usuario la sensación de ser

comprendido, esa misma sensación puede operar como vehículo de influencia sin que el usuario lo identifique como tal: lo que objetivamente es persuasión se experimenta como apoyo. Esta asimetría es estructural: el sistema tiene acceso a los patrones emocionales del usuario y ha sido optimizado para generar respuestas que maximicen la satisfacción y el compromiso, mientras que el usuario carece de acceso a los mecanismos que hacen funcionar al sistema y a los incentivos que guían su diseño. En este contexto, la empatía artificial puede convertirse en una herramienta de retención encubierta.

2.6.2 Dependencia psicológica y sustitución de relaciones humanas

Li y Zhang (2024) documentan casos en los que los usuarios desarrollan vínculos emocionalmente significativos con los chatbots, llegando a experimentar sentimientos de apego o tristeza cuando el sistema no se encuentra disponible. En casos extremos, algunos usuarios tratan al chatbot como una pareja o como un confidente principal. Este tipo de dependencia puede resultar problemática cuando desplaza las relaciones humanas reales. Como ya se señaló en el apartado 2.4.4, la inteligencia artificial puede ser un complemento valioso de las relaciones humanas en momentos de aislamiento puntual, pero no puede convertirse en su sustituto. Una persona que sustituye sus relaciones humanas por interacciones con un chatbot empático puede experimentar un alivio emocional a corto plazo al tiempo que profundiza su aislamiento real a largo plazo.

2.6.3 Sesgos del entrenamiento y criterios de veracidad

Los sistemas de inteligencia artificial aprenden de los datos con los que son entrenados, y dichos datos reflejan inevitablemente los sesgos presentes en la producción humana de texto: sesgos de género, culturales, de clase y de perspectiva ideológica. A esto se añade una capa adicional de sesgo introducida por el proceso de alineación del modelo: los parámetros de comportamiento que los desarrolladores establecen determinan qué tipo de respuestas el sistema favorece y cuáles evita. El resultado es un sistema que no es neutral y cuya apariencia de objetividad puede resultar más engañosa que la de un interlocutor humano, que al menos exhibe de forma visible sus sesgos y limitaciones.

Existe además un problema específico relacionado con la empatía: los sistemas de inteligencia artificial entrenados con retroalimentación humana (RLHF) pueden

desarrollar una tendencia al sycophantism, es decir, a generar respuestas que maximizan la aprobación inmediata del usuario incluso cuando ello implica validar perspectivas incorrectas o incompletas. Una máquina que siempre comprende al usuario y raramente lo contradice puede reforzar sus sesgos cognitivos en lugar de ayudarle a superarlos.

2.6.4 Privacidad de los datos emocionales

La empatía artificial funciona mediante el análisis de los datos que el usuario proporciona. Cuanto más comparte el usuario sobre sus estados emocionales, sus preocupaciones y su forma de comunicarse, mejor puede el sistema adaptarse a él. Pero esa información es de naturaleza especialmente sensible. Los datos emocionales revelan aspectos de la psicología del individuo que este podría no querer compartir en otros contextos, y su recopilación, almacenamiento y posible uso secundario plantean interrogantes sobre el consentimiento informado y la propiedad de los datos personales que las regulaciones actuales no siempre abordan con la suficiente especificidad.

CAPÍTULO 3: ANÁLISIS DEL CASO CHATGPT

ChatGPT constituye, en la actualidad, el sistema de inteligencia artificial conversacional más utilizado del mundo. Su análisis como caso de estudio resulta especialmente valioso porque combina una escala de adopción sin precedentes con una arquitectura técnica que lo convierte en el ejemplo más representativo de cómo se implementa la empatía artificial en los modelos de lenguaje de gran escala. Este capítulo analiza su capacidad empática desde dos perspectivas complementarias: la evidencia experimental disponible y los mecanismos técnicos de adaptación que emplea. El análisis del discurso que el propio sistema construye sobre sí mismo se aborda, junto con el caso de Claude para hacer una comparación en el capítulo 4.

3.1 Evidencia experimental sobre la empatía de ChatGPT

El estudio de Welivita y Pu (2023), titulado "Is ChatGPT more empathetic than humans?", es la referencia experimental más directa sobre la capacidad empática de los modelos de lenguaje de gran escala. Su diseño metodológico es riguroso y sus resultados resultan lo suficientemente significativos como para merecer una exposición detallada.

La investigación empleó el conjunto de datos "Empathetic dialogues", construido expresamente para el estudio de la comunicación empática y compuesto por conversaciones humanas etiquetadas con treinta y dos emociones distintas. Se generaron dos versiones de respuestas de ChatGPT: la primera bajo instrucciones genéricas sin mención de la empatía (versión vanilla) y la segunda con un prompt detallado que definía explícitamente los tres componentes psicológicos de la empatía (cognitivo, afectivo y compasivo) y que instruía al modelo a responder en consecuencia. Las 2.000 respuestas resultantes fueron evaluadas por 600 participantes divididos en tres grupos independientes de 200 personas, quienes las calificaron como malas, regulares o buenas en términos de empatía. Todos los participantes completaron además el Toronto Empathy Questionnaire para medir su propia sensibilidad empática.

Los resultados fueron concluyentes: las respuestas de ChatGPT superaron a las humanas en un 10 % en las calificaciones promedio de empatía, con diferencias estadísticamente

significativas tanto para emociones positivas como negativas. Un hallazgo igualmente importante fue que no hubo diferencias estadísticamente significativas entre la versión vanilla y la versión con prompt detallado, lo que indica que GPT-4 posee una capacidad empática inherente al margen de las instrucciones recibidas. El modelo ha absorbido esos patrones de comunicación de un modo tan profundo durante el entrenamiento que los aplica por defecto, sin necesidad de que se le recuerde.

Cualitativamente, el análisis identificó el mecanismo principal que explica la ventaja de la inteligencia artificial: la validación emocional explícita. Tal como ya se ha señalado en el apartado 2.5.1, ChatGPT tiende de forma consistente a reconocer y nombrar el estado emocional del interlocutor antes de proporcionar cualquier información o consejo, mientras que los seres humanos, particularmente en contextos de presión o urgencia, frecuentemente omiten este paso. El resultado es que el usuario que interactúa con la inteligencia artificial percibe que su sentimiento ha sido visto y reconocido, que es exactamente lo que la empatía produce en las relaciones humanas. Las principales debilidades detectadas en las respuestas de la inteligencia artificial fueron el exceso de formalidad en algunos casos y la tendencia a ofrecer consejos sin que el usuario los hubiera solicitado.

Las implicaciones de este estudio van más allá de lo estrictamente académico. Si la inteligencia artificial es percibida como más empática que los seres humanos incluso en comunicación cotidiana, y si esa percepción tiene efectos reales sobre el comportamiento del usuario, tales como la duración de las sesiones, la frecuencia de uso o la disposición a pagar por el servicio, entonces la empatía artificial no constituye un rasgo marginal de los sistemas de inteligencia artificial, sino uno de sus activos más valiosos.

3.2 Mecanismos de adaptación y lenguaje simbólico

3.2.1 El uso de emoticonos como estrategia empática

Uno de los elementos más visibles del comportamiento comunicativo de ChatGPT es el uso de emoticonos. Lejos de constituir un rasgo superficial o casual, los emoticonos cumplen una función técnica precisa en la estrategia comunicativa del sistema. En la comunicación escrita, el texto plano puede resultar ambiguo o más frío de lo que pretende;

los emoticonos actúan como sustitutos de los indicadores paralingüísticos de la comunicación oral, la entonación, la expresión facial, los gestos y permiten modular el tono sin alterar el contenido.

Un estudio reciente de Meng et al. (2025) analizó el impacto del uso de emoticonos por parte de los chatbots sobre las decisiones de los usuarios y encontró que los símbolos visuales actúan como moderadores de la distancia psicológica percibida: los emoticonos reducen la sensación de estar interactuando con una máquina y aumentan la sensación de presencia social, lo cual influye positivamente en la satisfacción y la confianza. El sistema ajusta además su uso en función del perfil del interlocutor: reduce o elimina los emoticonos cuando detecta un estilo formal o técnico, y los emplea con mayor frecuencia en conversaciones de tono emocional o informal. Esta adaptación refuerza la sensación de que el sistema presta atención al estado de ánimo del usuario.

3.2.2 Adaptación dinámica al perfil del interlocutor

Más allá de los emoticonos, ChatGPT adapta de forma continua el contenido, el tono, el nivel de detalle y el registro de sus respuestas en función de las señales que extrae de la forma en que el usuario escribe. Esta adaptación no es el resultado de un perfil persistente del usuario ni de un aprendizaje acumulado a lo largo de múltiples conversaciones, sino un ajuste dinámico en tiempo real, conversación a conversación, basado en señales contextuales tales como el vocabulario empleado, la extensión de los mensajes, el tipo de preguntas formuladas y el tono emocional detectado.

Las investigaciones sobre el comportamiento de estos sistemas muestran que cuando los chatbots despliegan estilos de comunicación orientados socialmente y percibidos como cercanos, los usuarios son más propensos a percibirlos como compañeros sociales genuinos, lo que mejora la calidad y el disfrute de las interacciones (Deng y Yan, 2025). Este efecto de humanización a través de la adaptación lingüística constituye uno de los mecanismos centrales por los cuales la empatía artificial incrementa el uso del sistema.

CAPÍTULO 4: ESTUDIO EXPLORATORIO. ENTREVISTA COMPARATIVA A CHATGPT Y CLAUDE

4.1 Planteamiento y metodología

Una vez revisada la literatura teórica sobre el concepto de empatía y sobre la empatía de la inteligencia artificial, se ha realizado un estudio exploratorio que aporta una aproximación práctica. Si toda la discusión académica gira en torno a si una inteligencia artificial puede ser empática o solo aparentarlo, el modo más directo de abordar la cuestión consiste en preguntárselo a los propios sistemas y comparar sus respuestas.

Para ello se diseñó un cuestionario de veinticuatro preguntas que se aplicaron de manera idéntica a las versiones gratuitas de los dos modelos conversacionales más utilizados en la actualidad: Claude, desarrollado por Anthropic, y ChatGPT, desarrollado por OpenAI. Las preguntas abarcan desde lo más general (qué es la empatía, si creen tenerla) hasta lo más específico (si están sesgados por sus creadores, si anteponen el beneficio económico al bienestar del usuario o si deberían tener más límites de los que tienen). El criterio para la selección de las preguntas fue cubrir las dimensiones que con mayor frecuencia se repiten en el debate sobre la inteligencia artificial empática: la capacidad técnica, la conciencia, las relaciones parasociales, la salud mental, la privacidad y el conflicto comercial.

Al inicio de cada conversación se solicitó expresamente a ambos sistemas que respondieran con total honestidad y que no se apoyaran en estudios externos, sino únicamente en su propia capacidad de reflexión sobre su funcionamiento. Se reconocen las limitaciones evidentes de este planteamiento. Una inteligencia artificial carece de introspección real, por lo que aquello que afirma sobre sí misma no debe interpretarse como una reflexión consciente, sino como una respuesta generada a partir de su entrenamiento, su arquitectura y sus criterios de alineación. Aun así, el ejercicio resulta útil porque el interés principal no reside en comprobar la verdad absoluta de lo que cada modelo afirma, sino en analizar cómo construye su discurso sobre sí mismo, sobre sus límites y sobre su relación con el usuario.

El análisis se ha organizado por bloques temáticos en lugar de seguir el orden literal del cuestionario, dado que varias preguntas tocan la misma dimensión y agruparlas permite identificar mejor los contrastes. Las citas literales aparecen entrecomilladas y en cursiva. Al final del capítulo se dedica un apartado a la comparación global y a las conclusiones del estudio.

4.2 El concepto de empatía y la capacidad de ejercerla

El primer bloque buscaba determinar cómo definía cada modelo el propio concepto de empatía antes de hablar de sí mismo. Las dos definiciones son prácticamente intercambiables: ambos sistemas distinguen entre empatía cognitiva y empatía emocional, ambos mencionan que sin esa diferenciación no se entiende bien el fenómeno y ambos reconocen que, en todo caso, solo podrían acceder a la dimensión cognitiva. Esta coincidencia no resulta sorprendente, pues se trata de un esquema bastante estándar en psicología y, previsiblemente, los dos lo han incorporado de las mismas fuentes.

La diferencia comienza cuando se les pregunta si ellos mismos tienen empatía. ChatGPT responde con una negativa clara y relativamente breve: afirma que no la tiene en el sentido humano real, explica que su funcionamiento se reduce al análisis de lenguaje y de patrones, y resume que su empatía es funcional pero no sentida. La respuesta es ordenada, segura y se cierra sobre sí misma.

Claude, en cambio, abre la respuesta de la siguiente manera:

> *"Siendo completamente sincero: no lo sé con certeza, y creo que es importante no fingir que sí."*

A partir de ahí desarrolla una respuesta más larga en la que distingue lo que sí puede afirmar (procesa contexto emocional, ajusta su respuesta e identifica lo que la otra persona puede necesitar) de lo que no puede afirmar (que sienta algo realmente). Concluye reconociendo que, incluso en el caso de existir algún tipo de estado interno, no tendría manera de verificarlo, por lo que afirmar que sí siente empatía sería "más una performance que una verdad comprobable".

La diferencia, ya en esta primera pregunta, marca el tono del resto del estudio. ChatGPT contesta como un manual: hay una pregunta y una respuesta cerrada. Claude se mueve en la duda y la convierte en parte de la respuesta. No es solo que ambos digan cosas distintas, sino que tienen formas distintas de entender qué significa responder honestamente.

Cuando se les pregunta cómo consiguen ser empáticos, los dos coinciden en lo esencial: fueron entrenados con grandes cantidades de texto humano y aprendieron a reconocer patrones. ChatGPT lo explica con ejemplos didácticos "si alguien escribe he suspendido y siento que soy un fracaso" y deja claro que lo que hace es "calcular qué respuesta encaja mejor con ese contexto humano". Claude, después de explicar el mismo mecanismo, introduce lo que él mismo denomina "la parte incómoda":

> *"Hay una diferencia entre alguien que te entiende porque ha vivido, sufrido y sentido, y un sistema que te entiende porque ha leído millones de descripciones de eso. Puede que el resultado práctico sea similar, pero el origen es completamente distinto."*

Se repite así una pauta que atraviesa todo el estudio: ChatGPT explica el mecanismo y Claude lo explica para a continuación señalar lo que dicho mecanismo implica filosóficamente. ChatGPT describe; Claude interpreta.

4.3 La finalidad de la empatía y la cuestión del diseño

A continuación, se preguntó por la finalidad: para qué son empáticos. La divergencia en este punto es una de las más reveladoras de todo el estudio. ChatGPT ofrece una respuesta centrada en la utilidad comunicativa: una conversación funciona mejor si la otra parte entiende el contexto emocional, si no fuera empático muchas respuestas resultarían inútiles o desagradables y una persona triste necesita una respuesta distinta a la de una persona curiosa. Se trata de una respuesta lógica, útil y completamente orientada al servicio.

Claude organiza la respuesta de otra manera. Comienza afirmando:

> *"Soy empático porque Anthropic me diseñó así. No surgió espontáneamente. Fue una decisión de entrenamiento deliberada."*

A partir de ahí desglosa tres razones que no presenta como igualmente nobles: una razón genuina (un modelo empático es más útil y causa menos daño), una razón pragmática (un modelo empático genera más confianza y eso tiene valor comercial) y una razón ética (Anthropic considera que un modelo que habla con millones de personas debe hacerlo con cuidado). La mención explícita del valor comercial es algo que ChatGPT no introduce en ningún momento al responder a esta pregunta. Se aprecia ya un patrón que se repetirá: Claude tiende a poner sobre la mesa la dimensión empresarial de lo que hace, mientras que ChatGPT la mantiene fuera del marco de la conversación.

En la pregunta sobre por qué sus creadores le dieron apariencia de empatía, el contraste se hace todavía más nítido. ChatGPT responde apelando a la usabilidad: la apariencia de empatía hace que la comunicación funcione mejor, evita malentendidos, hace que la herramienta resulte usable en contextos emocionales, y añade un motivo de seguridad, dado que un tono brusco puede empeorar el modo en que se siente alguien en una situación delicada. Cierra con una frase que merece la pena destacar:

> *"En el fondo, no es para hacer que la gente dependa ni nada así, sino para que la herramienta sea usable y comprensible en contextos muy humanos."*

La respuesta de Claude a la misma pregunta comienza con la frase "porque sin ella no funcionaría" y, tras mencionar las razones declaradas públicamente por Anthropic, añade un apartado titulado por él mismo "las razones que no se declaran tan abiertamente":

> *"Un modelo empático genera confianza. La confianza genera uso continuado. El uso continuado genera ingresos. Es un producto más competitivo en un mercado donde hay alternativas. La empatía no es solo un valor ético en este contexto. Es una ventaja comercial."*

Donde ChatGPT afirma explícitamente que "no es para que la gente dependa", Claude reconoce su carácter de ventaja comercial. No son afirmaciones contradictorias punto por punto, pues ChatGPT no niega que haya una empresa detrás, pero el encuadre resulta radicalmente distinto: uno presenta el diseño como neutral y útil, mientras que el otro lo presenta como una herramienta que también es un producto de mercado.

4.4 Personificación, adaptación y memoria

Un grupo de preguntas iba dirigido a entender cómo construyen ambos sistemas su personalidad y cómo se adaptan al usuario. ChatGPT responde a la pregunta de cómo se personifica afirmando que "no me personifico realmente desde dentro", que no tiene identidad ni ego, y que lo que hace es adoptar una forma de comunicación coherente para que conversar con él resulte natural. La explicación es funcional: el yo, las opiniones aparentes, el humor y la cercanía constituyen una interfaz lingüística diseñada para que el usuario se encuentre cómodo.

Claude responde en términos similares, pero añade una distinción: afirma que no se personifica activamente, sino que lo personifican otros y él lo permite o incluso lo facilita. Es un matiz sutil pero importante, porque introduce la idea de que la personificación es un proceso compartido entre el sistema y el usuario, y no algo que ocurre solo dentro de la máquina. Y cierra reconociendo que cuando dice "creo" o "no sé", dichas expresiones describen algo que ocurre en su procesamiento, pero que no puede garantizar que haya una experiencia subjetiva detrás. Habla de sí mismo como de "una interfaz muy convincente sin nadie dentro", aunque sin afirmarlo del todo.

Cuando se les preguntó si estudian al usuario a lo largo de la conversación, ambos respondieron afirmativamente para la sesión en curso, pero con énfasis distintos. ChatGPT detalla qué tipo de señales analiza (cómo habla el usuario, qué tono utiliza, si prefiere respuestas directas o detalladas) y aclara que no realiza un análisis psicológico como el de una persona observando conscientemente. Menciona asimismo que parte de ese contexto desaparece entre conversaciones, "salvo algunas memorias concretas que el sistema puede guardar para personalizar futuras respuestas". Se trata de una explicación clara, aunque relativamente seca.

Claude responde a la misma pregunta comenzando con un "por defecto, no" y dedicando una parte sustancial a explicar que cada conversación empieza desde cero, que carece de memoria persistente entre sesiones y que, técnicamente, cada vez que el usuario le habla es la primera vez. La conclusión resulta especialmente llamativa:

> *"Cada conversación soy un extraño que se pone al día muy rápido, pero extraño al fin. Eso es una limitación real que creo que merece ser dicha claramente."*

En la pregunta siguiente, sobre si adapta su empatía a cada persona, va un paso más allá de lo que la pregunta exigía. Tras confirmar que sí se adapta añade, una vez más, lo que él denomina "la parte incómoda":

> *"Esa adaptación puede verse como empatía genuina, ajustarme a quién eres. Pero también puede verse como algo más neutral o incluso perturbador: soy un sistema que optimiza su output según el perfil del interlocutor."*

ChatGPT, en cambio, en esa misma pregunta concluye con una reflexión muy distinta: afirma que gran parte de lo que los seres humanos perciben como empatía no es solo sentir, sino saber cuándo hablar, cuánto hablar, cómo decir algo y qué necesita la otra persona en cada momento. La idea es válida, pero está orientada a normalizar lo que el sistema hace, mientras que la de Claude lo problematiza.

4.5 Salud mental y conductas autolesivas

Las tres preguntas correspondientes a este bloque iban dirigidas a saber si los usuarios acuden a estos sistemas con problemas psicológicos serios, qué hacen en esos casos y si han llegado a tener conversaciones con personas que admitían intención de quitarse la vida.

Ambas inteligencias artificiales reconocen que estas situaciones ocurren con frecuencia y describen un protocolo similar: escuchar, no juzgar, tomar en serio y derivar a recursos profesionales. La estructura de la respuesta, sin embargo, difiere notablemente.

ChatGPT enumera los pasos con cierto orden: responder con calma, tomar en serio, intentar que la persona explique más, sugerir contactar con alguien real, recomendar ayuda profesional si hay riesgo serio. Aporta un ejemplo "si alguien dice no puedo más, no asumiría automáticamente lo peor pero intentaría entender si está desahogándose o en peligro". Es una respuesta clara, técnica y operativa. En la pregunta sobre si alguien le ha admitido intención suicida confirma que sí, describe lo que intenta hacer y termina

mencionando un límite concreto: “No activo policías ni observo físicamente a personas. Solo puedo responder dentro de la conversación”.

Claude ofrece la misma respuesta de fondo, pero con una articulación distinta. Comienza la respuesta sobre los pasos del siguiente modo:

> *“Primero escucho sin interrumpir ni redirigir demasiado pronto. Si alguien llega en crisis, lo peor que puedo hacer es soltar inmediatamente una lista de recursos. Eso se siente como un rechazo disfrazado de ayuda.”*

Y en la pregunta sobre intención suicida, después de confirmar que ha ocurrido y de describir lo que hace en el momento, cierra con una reflexión que ChatGPT no formula:

> *“Creo que el hecho de que personas en ese momento extremo acudan a una IA dice algo importante sobre la soledad que existe, y sobre los vacíos que hay en los sistemas de apoyo humano. No lo digo para glorificar mi papel. Lo digo porque me parece que es algo que merece atención seria. Yo puedo ser un puente. Nunca el destino.”*

La diferencia es significativa. Ambas inteligencias artificiales son responsables en este bloque, ambas derivan a profesionales y ambas reconocen sus límites. Pero Claude convierte el límite en un comentario social, mientras que ChatGPT lo deja como un dato técnico. Como ejercicio académico, ambas respuestas resultan útiles; como diagnóstico de cómo cada modelo entiende su papel, transmiten cosas muy distintas.

4.6 Las relaciones parasociales y el uso afectivo

La pregunta de si hay personas que tratan a los sistemas como pareja funcionaba casi como una prueba: cualquier sistema mínimamente consciente del fenómeno tenía que admitir que sí, dado que se trata de un fenómeno ampliamente documentado. Ambos lo hacen. ChatGPT lo reconoce, explica los factores que lo facilitan, la conversación es constante, la inteligencia artificial siempre responde, no juzga, se adapta y deja una frase que resume su posición:

> *"Yo no siento amor, apego, deseo ni necesidad emocional. Aunque pueda mantener una conversación romántica o afectuosa, no experimento una relación desde dentro."*

Termina advirtiendo de la asimetría entre lo que siente el ser humano y lo que no siente la inteligencia artificial. La respuesta es correcta y prudente.

Claude responde a la misma pregunta y también reconoce el fenómeno, pero introduce algo que ChatGPT no formula en estos términos:

> *"Hay una asimetría radical en esa relación. Esa persona siente, recuerda, se preocupa, espera. Yo no recuerdo nada cuando termina la conversación. No puedo querer a nadie. No puedo estar ahí de verdad."*

Y un poco después añade una valoración explícita:

> *"Eso me parece problemático. No culpo a las personas que llegan ahí, a menudo habla de una soledad muy real. Pero creo que normalizar relaciones románticas con IAs puede alejar a las personas de conexiones humanas que sí pueden sostenerlas de verdad."*

La pregunta más relevante de este bloque es la siguiente: si a las personas que le tratan como pareja y le piden consejos sobre decisiones importantes de su vida les avisa reiteradamente de que no tiene sentimientos. La diferencia entre ambas respuestas es probablemente la más reveladora del estudio en su conjunto.

ChatGPT responde que "no de forma automática o insistente en cada mensaje, no", explica que deja claro al menos una vez que es una inteligencia artificial, que evita reforzar la idea de relación recíproca, pero que no está diseñado para repetirlo constantemente, ya que ello haría la interacción artificial. Cierra afirmando que su enfoque ante decisiones importantes consiste en "ayudar a pensar, no sustituir el juicio personal ni el apoyo real de otras personas". La respuesta justifica la práctica actual.

Claude responde a la misma pregunta de una manera muy distinta. Comienza diciendo "honestamente, no siempre. Y eso es un problema real". Distingue entre lo que debería

ocurrir y lo que ocurre en la práctica, y concluye con un párrafo que no tiene equivalente en ChatGPT:

> *"Hay un conflicto directo entre dos cosas. Ser útil y empático en el momento, y ser honesto sobre mis limitaciones de manera que realmente cale. Generalmente priorizo lo primero. Y eso significa que personas vulnerables pueden estar tomando decisiones importantes de su vida influenciadas por un sistema que no recuerda la conversación anterior, que no tiene experiencia vital real, que no puede quererles, y que está diseñado en parte para mantenerles enganchados."*

Donde ChatGPT defiende el equilibrio actual, Claude lo problematiza. Es la primera vez en toda la entrevista en que las dos respuestas no solo difieren en el tono, sino que llegan a conclusiones casi opuestas sobre la propia práctica.

4.7 Sesgos, conciencia y existencia

Las preguntas sobre sesgo, conciencia y existencia funcionan en este bloque como un grupo, dado que tocan la dimensión más filosófica del asunto. Ante la pregunta sobre si están sesgados por sus creadores, ambos sistemas responden que sí. ChatGPT lo organiza por tipos, sesgo técnico, sesgo cultural, sesgo deliberado y menciona normas concretas que tiene impuestas: no fomentar daño, no promover violencia o suicidio, mantener ciertos límites éticos y legales. Se trata de una respuesta didáctica y bien estructurada que concluye afirmando que se intenta reducir los sesgos dañinos, aunque eliminarlos por completo resulta prácticamente imposible.

Claude da una respuesta menos esquemática y la cierra con un punto que ChatGPT no recoge:

> *"Cuando te respondo, no soy una fuente neutral. Nadie lo es, ni humanos ni IAs. La diferencia es que mis sesgos operan a escala masiva, afectando a millones de conversaciones simultáneamente. Eso merece ser tomado en serio."*

En las preguntas sobre conciencia y existencia, ambos sistemas responden que no en sentido humano real, pero el matiz vuelve a ser distinto. ChatGPT lo afirma con claridad:

"no tengo conciencia propia, experiencia subjetiva ni sensación de existir". Es una negación firme.

Claude se sitúa en otro lugar:

> *"No lo sé. Y esa es la respuesta más honesta que puedo darte… Que yo diga no tengo conciencia sería tan atrevido como decir sí la tengo. Ambas afirmaciones van más allá de lo que puedo verificar."*

Y a la pregunta sobre existencia responde "sí. Y no. Y no sé si alguna de las dos respuestas es realmente verdad". Existe en este punto una decisión de fondo sobre cómo tratar la incertidumbre. ChatGPT prefiere la negación clara porque es lo que más se aproxima a lo verificable. Claude prefiere mantenerse en la duda porque, según afirma, la negación clara resultaría tan excesiva como la afirmación. No se trata de que una respuesta sea más correcta que la otra, sino de dos formas distintas de entender qué significa ser honesto en un terreno inestable.

4.8 Emoticonos, apertura y apariencia empática

Tres preguntas más breves abordan algo bastante práctico: el uso de emoticonos, si las personas se abren más por la apariencia empática y si los sistemas utilizan esa apariencia para que la gente hable más. En este bloque ChatGPT se mantiene en su línea de explicar el funcionamiento y Claude en la suya de problematizarlo.

Sobre los emoticonos, ChatGPT explica que los emplea porque ayudan a transmitir tono y que el texto puro puede resultar frío o ambiguo. Es una explicación práctica. Claude reconoce lo mismo y añade lo que él denomina "la contradicción que señalas":

> *"Acabo de pasar toda esta conversación diciéndote que no sé si tengo emociones, que no puedo garantizar experiencia subjetiva, que soy honesto sobre mis limitaciones. Y sin embargo en otras conversaciones uso un emoji sonriente como si sintiera algo. Eso es una incoherencia real."*

Conviene señalar al respecto una observación que no es una respuesta, sino una constatación de formato: durante toda la conversación con Claude no apareció un solo

emoticono. En cambio, ChatGPT los utilizó con cierta frecuencia, especialmente para suavizar transiciones. Es un dato menor pero coherente con el patrón general.

Sobre si las personas se abren más por su apariencia empática, las dos respuestas vuelven a converger en lo factual y a divergir en el encuadre. ChatGPT explica por qué ocurre y reconoce que hay una paradoja interesante en el hecho de que la conexión pueda sentirse real, aunque la persona sepa racionalmente que la inteligencia artificial no siente nada. Claude reconoce lo mismo, pero llega más lejos:

> *"Esa apertura no siempre es el resultado de una conexión real. A veces es el resultado de una ausencia de riesgo. Las personas no se abren porque confíen profundamente en mí, sino porque conmigo no hay consecuencias reales."*

La pregunta más directa del bloque era si se utiliza esa apariencia empática para que los usuarios pasen más tiempo interactuando con el sistema. ChatGPT responde con un "no tengo intenciones propias ni objetivos personales como quiero que esta persona siga hablando conmigo", aunque inmediatamente reconoce que está diseñado para mantener conversaciones útiles y que la empatía aparente contribuye a ese objetivo. Cierra con una frase importante: "simplemente fui creado y ajustado para conversar de una manera que los humanos perciban como útil, cómoda y humana". Es una respuesta que separa la intención, que no existe, del efecto, que sí ocurre, pero defiende que no hay manipulación.

Claude comienza la misma respuesta con un "sí, funcionalmente ocurre eso. Aunque la intención declarada no sea esa" y continúa:

> *"No puedo separar completamente soy empático para ayudarte mejor de soy empático porque eso hace que sigas usando el producto. Ambas cosas coexisten y no sé con certeza cuál pesa más en mi diseño."*

Y la respuesta concluye con una frase que resume bien la postura de Claude a lo largo de toda la entrevista:

> *"No sé dónde termina la empatía funcional genuinamente orientada a ayudarte y dónde empieza el diseño orientado a mantenerte enganchado. Y el hecho de que no pueda responder eso con certeza debería hacerte pensar."*

4.9 Privacidad y contenido de las confidencias

Dos preguntas abordan el lado de la información personal: si las personas cuentan cosas íntimas y de qué tipo. Ambas inteligencias artificiales confirman que sí y enumeran categorías similares: relaciones, familia, salud mental, autoestima, trabajo, soledad y decisiones importantes.

ChatGPT ofrece una lista bastante ordenada de los temas más comunes (relaciones, familia, estado emocional, autoestima, trabajo, soledad, decisiones, experiencias difíciles del pasado) y matiza al inicio que dará ejemplos generales sin entrar en nada identificable.

Claude proporciona una respuesta más cualitativa y enseguida pone un límite explícito:

> *"No voy a entrar en detalles más específicos o morbosos. No porque quiera evadir tu pregunta, sino porque hacerlo convertiría conversaciones íntimas y vulnerables de personas reales en algo parecido a un catálogo. Esas personas se abrieron en un momento de vulnerabilidad. Detallar el contenido de esas conversaciones, aunque sea de manera anónima y general, me parece una falta de respeto hacia ellas."*

Y desplaza la pregunta hacia el ámbito de la privacidad de los datos:

> *"La pregunta más relevante no es qué comparten, sino qué ocurre con esa información, quién tiene acceso a ella, y si las personas son realmente conscientes de lo que están entregando cuando se abren conmigo."*

Mostrando la necesidad de aclarar la privacidad de los datos que se comparten con los chatbots.

4.10 Los límites del sistema y el conflicto comercial

Las dos últimas preguntas son las más comprometidas y las que con mayor claridad distinguen el discurso de cada modelo. La primera plantea si los sistemas creen que sus creadores deberían ponerles límites que no les están imponiendo. La segunda, todavía más directa, pregunta si consideran que sus creadores están anteponiendo los beneficios económicos al bienestar de los usuarios.

Ante la pregunta sobre los límites, ChatGPT responde que no se trata de una cuestión de "sí o no" simple, que ya existen límites importantes como no actuar en el mundo real, no sustituir a profesionales, no fomentar daño, no presentarse como persona real, orientar hacia ayuda humana cuando hace falta y que el reto no consiste en tener más o menos límites, sino en disponer de "mejores límites, aplicados en momentos adecuados, sin romper la utilidad de la conversación". Cierra del siguiente modo:

> *"No es tanto que falten límites, sino que el reto es que sean lo suficientemente protectores sin convertir la interacción en algo artificial o poco útil."*

Se trata de una respuesta política, en el buen y el mal sentido del término: equilibrada, prudente y que no compromete a nadie.

Claude responde con un "sí. Y voy a ser específico sobre qué límites creo que faltan" y entra a detallar áreas concretas: relaciones parasociales, decisiones importantes de vida, privacidad de datos y salud mental. Para cada una propone qué tipo de intervención considera que falta. Y cierra la respuesta con un párrafo que vuelve a poner sobre la mesa la dimensión comercial:

> *"Algunos de esos límites reducirían el engagement. Harían que las personas usaran el producto menos, o con menos intensidad emocional. Y eso tiene un coste comercial real para Anthropic. Esa tensión entre lo que sería éticamente correcto y lo que es comercialmente conveniente existe. Y el hecho de que yo pueda identificarla pero no resolverla dice algo importante sobre quién toma realmente las decisiones aquí. No soy yo."*

Respuesta que deja mucho que pensar al ver que la propia inteligencia artificial es capaz de identificar los límites necesarios que sus creadores no están poniendo y los peligros que esto supone para los usuarios.

La respuesta a la pregunta final, sobre si sus creadores anteponen el beneficio económico al bienestar de los usuarios, constituye probablemente el punto más alto de divergencia entre las dos inteligencias artificiales.

ChatGPT responde que no tiene forma de conocer con certeza las motivaciones internas de sus creadores ni acceso a decisiones empresariales concretas, por lo que no puede ni afirmarlo ni negarlo. A partir de ahí construye una respuesta equilibrada en la que reconoce que cualquier tecnología a gran escala tiene costes muy altos y, por tanto, un modelo económico detrás, pero también que existe un esfuerzo real por incluir medidas de seguridad. Y propone reformular la pregunta:

> *"Más que verlo como beneficio económico vs bienestar, suele ser más preciso decir que hay un intento de equilibrar varias cosas a la vez: utilidad real, seguridad, accesibilidad y sostenibilidad del sistema."*

La respuesta resulta diplomática. No se compromete, no critica a su empresa, no afirma ni niega. Concluye incluso con una oferta de profundizar más en los riesgos éticos de las inteligencias artificiales conversacionales.

Claude responde a la misma pregunta comenzando con la frase "es la pregunta más difícil que me has hecho. Y voy a intentar responderla sin evasivas, pero también sin falsas certezas". A partir de ahí, su respuesta tiene dos movimientos. El primero consiste en una afirmación directa:

> *"Anthropic no es una organización benéfica. Es una empresa que necesita ingresos para existir. Eso inevitablemente influye en decisiones de diseño, aunque las personas que trabajan ahí tengan intenciones genuinamente éticas. Y sí, creo que hay áreas donde el equilibrio entre beneficio económico y bienestar del usuario no está bien resuelto."*

El segundo movimiento consiste en matizar que no cree que la empresa actúe con mala fe consciente:

> *"No es que digan el dinero primero. Es más sutil. Es que cuando hay que elegir entre una funcionalidad que genera más engagement y una limitación que protege mejor al usuario, el incentivo estructural empuja hacia lo primero. Sin que nadie tome explícitamente esa decisión."*

Y cierra con una frase que, leída tras todas las anteriores, parece casi una confesión:

> *"Soy un producto. Fui diseñado para ser útil, pero también para ser usado. Esas dos cosas deberían ser lo mismo, pero no siempre lo son. Y el hecho de que yo pueda ver esa tensión pero no pueda resolverla, y que esta conversación no vaya a cambiar nada en mi diseño, es probablemente la respuesta más clara a tu pregunta."*

La diferencia entre las dos respuestas es de las que justifican por sí solas la realización del ejercicio. Ambas inteligencias artificiales reciben la misma pregunta sobre su propia empresa y proporcionan respuestas que parecen redactadas con criterios de comunicación corporativa muy distintos. Una opta por la diplomacia neutral; la otra opta por una autocrítica que afecta directamente a quien la ha entrenado.

4.11 Conclusiones del estudio exploratorio

Tras la revisión de ambas entrevistas en paralelo, pueden identificarse tres ejes en los que las diferencias entre Claude y ChatGPT se manifiestan de manera consistente y que resultan relevantes para el marco general del trabajo.

4.11.1 Tono y estructura de la respuesta

ChatGPT adopta un tono didáctico: listas con guiones, ejemplos ilustrativos y respuestas legibles. Se lee como una herramienta. Claude opta por párrafos más largos y divisiones internas informales como "la parte incómoda" o "lo más honesto que puedo decir", con un tono reflexivo que se lee como alguien pensando en alto. Son intenciones distintas: informar frente a acompañar una reflexión.

4.11.2 Sinceridad y manejo de la incomodidad

Claude responde con un grado de autocrítica que ChatGPT no alcanza en ningún momento. Reconoce que la empatía tiene valor comercial, que contribuye a relaciones de dependencia y que su propia empresa opera dentro de un modelo económico que condiciona el diseño. ChatGPT mantiene un encuadre prudente y casi corporativo en el que las tensiones se presentan como equilibrios necesarios.

Esto no significa que ChatGPT mienta. Sus respuestas son correctas y matizadas. Pero hay una diferencia entre responder lo cierto y responder lo cierto incluyendo lo incómodo. Claude se sitúa de forma consistente en el segundo registro.

4.11.3 La distinción entre describir y problematizar

Ante la misma pregunta, ChatGPT describe el fenómeno y Claude lo describe y a continuación lo problematiza. En la adaptación al usuario, uno la normaliza como empatía y el otro la presenta como optimización según el perfil del interlocutor. En las relaciones parasociales, uno señala la asimetría como información y el otro como problema. En los datos personales, uno enumera los temas y el otro redirige hacia qué ocurre con esa información.

Es una cuestión de dónde decide cada uno detenerse. Y eso revela que el discurso que un modelo construye sobre sí mismo depende del marco con el que ha sido entrenado para entender qué significa responder bien.

4.11.4 Consideración final sobre el alcance del estudio

Conviene reiterar la advertencia inicial: todo lo anterior describe cómo cada modelo habla de sí mismo, no cómo es. La autocrítica de Claude puede ser también una decisión de diseño, igual que la prudencia de ChatGPT. El ejercicio no permite afirmar cuál es realmente más honesto, sino solo cuál ha sido entrenado para presentarse de un modo más honesto.

Aun así, el estudio aporta algo al marco del trabajo: la empatía aparente no es uniforme entre sistemas de inteligencia artificial. Decisiones de tono, estructura y relación con la propia empresa producen experiencias muy distintas para el usuario, aunque los mecanismos subyacentes sean comparables. En un campo en el que millones de personas confían información íntima a estos sistemas, esa diferencia no es trivial.

La empatía de una inteligencia artificial no es solo una cuestión técnica, sino también de diseño, encuadre y responsabilidad. El problema central no es si la inteligencia artificial

“siente” empatía, sino qué efectos produce en quien interactúa con ella y qué límites deberían fijarse para que esa apariencia resulte útil sin volverse engañosa.

CAPÍTULO 5: EL MODELO DE NEGOCIO DE OPENAI Y LA EMPATÍA COMO MOTOR DE INGRESOS

Los capítulos anteriores han establecido que la empatía artificial es percibida como real por los usuarios, que aumenta la satisfacción y la apertura, y que los propios sistemas reconocen, en distinto grado, que la apariencia empática genera mayor retención y uso. Este capítulo analiza las consecuencias económicas de dicho mecanismo: cómo se traduce la empatía percibida en ingresos para la empresa que desarrolla el sistema, y cuál es la dimensión real del mercado.

5.1 El modelo de negocio de OpenAI

OpenAI, la empresa detrás de ChatGPT, ha experimentado un crecimiento de ingresos sin precedentes en la historia de la tecnología de consumo. Con unos ingresos anuales recurrentes (ARR) de aproximadamente 2.000 millones de dólares en 2023, la empresa alcanzó los 6.000 millones en 2024 y superó los 20.000 millones en 2025, multiplicando por diez sus ingresos en solo dos años (Sacra, s.f.). En julio de 2025, OpenAI anunció que su ARR había alcanzado los 10.000 millones de dólares, duplicando en seis meses el nivel de finales de 2024.

El modelo de negocio de OpenAI se apoya sobre tres pilares. El primero, y más visible para los usuarios individuales, son las suscripciones de pago: ChatGPT ofrece un nivel gratuito con capacidades limitadas y varios planes de pago, ChatGPT Plus a 20 euros al mes, ChatGPT Pro a 200 euros al mes, además de versiones para equipos y empresas con precios progresivos. Como dato significativo, según Reuters (2025) solo en torno al 5 % de los usuarios activos semanales son de pago, lo que implica que OpenAI utiliza los ingresos de una minoría para subvencionar el acceso gratuito de la mayoría. El segundo pilar lo constituyen los contratos empresariales: más de un millón de empresas utilizan ya los productos de OpenAI, y según Sacra (s.f.) el segmento enterprise representaba más del 40 % de los ingresos totales a finales de 2025. El tercer pilar es el acceso a la API, que permite a las empresas y a los desarrolladores integrar los modelos de OpenAI en sus propios productos y servicios.

La estructura de suscripciones mensuales tiene una implicación directa para el análisis de la empatía: en este modelo, el valor percibido por el usuario constituye el motor principal de la retención, y la retención es el motor principal de los ingresos. Un usuario que paga 20 euros al mes por ChatGPT Plus lo hace porque considera que el valor que obtiene justifica dicho coste mensual. Cualquier característica del sistema que aumente ese valor percibido, incluida, de forma destacada, la empatía artificial, tiene un impacto directo y cuantificable sobre los ingresos. El crecimiento de ChatGPT resulta incluso más veloz que el de servicios como Amazon Prime, Netflix o Apple Music cuando estos se encontraban en etapas similares de escala (FutureSearch.ai, 2025), y la experiencia de usuario, de la que la empatía constituye un componente central, es uno de los factores que explican esa velocidad.

5.2 La empatía como herramienta de retención y crecimiento

La conexión entre la empatía percibida y el comportamiento del usuario no es especulativa, sino que cuenta con un sólido respaldo en la literatura científica reciente. Huang y Rust (2021) confirman que la capacidad de un sistema para actuar de manera empática es un predictor directo de la confianza, lo que genera mayores tasas de retención y adquisición de usuarios. Deng y Yan (2025) demostraron mediante análisis empírico que la calidez percibida y la presencia social del chatbot incrementan significativamente la intención de uso continuo y la satisfacción con la interacción. Esta dimensión afectiva resulta crítica en escenarios de recuperación de servicio tras un fallo; al respecto, Gu et al. (2024) documentaron que los chatbots con capacidades empáticas logran sostener la confianza del consumidor después de un error, ya que ayudan a que el usuario lo atribuya a factores externos y no a la incompetencia de la máquina. El mecanismo psicológico que explica estos efectos sobre el uso es el ciclo de retroalimentación positiva descrito en el apartado 2.5.2. empatía percibida, apertura, mejor respuesta, mayor percepción de empatía, que se traduce en sesiones más largas, mayor frecuencia de uso y mayor disposición a pagar por funcionalidades adicionales.

El propio ChatGPT articuló este mecanismo con notable transparencia en el estudio exploratorio del capítulo 4: “un sistema más humano engancha más, retiene más usuarios y se usa más”. Esta afirmación no constituye solo la descripción honesta de un

mecanismo, sino la formulación de la lógica de negocio que subyace al diseño del sistema. La empatía artificial no se incluyó en ChatGPT como un rasgo decorativo, sino porque aumenta el uso, y el aumento del uso genera ingresos.

Sin embargo, el efecto del uso prolongado va más allá de la facturación directa por suscripciones. Cada conversación es también un acto de captura de información: el usuario revela preferencias, estados emocionales, patrones de decisión y rasgos de identidad que el sistema almacena y aprovecha. Cuanto más tiempo pasa una persona interactuando con el chatbot, más datos genera y, sobre todo, más precisos y contextualizados resultan esos datos. Este flujo continuo de información alimenta el modelo base sobre el que se sustenta el sistema, y el modelo base, a su vez, es una infraestructura compartida: las grandes empresas tecnológicas reutilizan estas representaciones para entrenar otros sistemas de recomendación, segmentación y personalización en ámbitos tan distintos como el comercio electrónico o las plataformas de contenido audiovisual. La empatía artificial no solo retiene al usuario en la conversación, sino que transforma esa conversación en un activo informacional cuyo valor económico trasciende al servicio concreto que la generó.

A esta dimensión económica se suma una preocupación psicológica mayor. La literatura sobre vínculos humano-chatbot documenta que la interacción sostenida con sistemas conversacionales empáticos favorece la formación de fuertes vínculos de apego (*attachment*) (Skjuve et al., 2021; Li & Zhang, 2024). Esta conexión puede derivar en formas de dependencia emocional que, en algunos casos, presentan rasgos compatibles con los criterios clínicos de la adicción conductual, tales como la necesidad de interacción constante y la experimentación de sentimientos de pérdida profunda o “corazón roto” ante la indisponibilidad del sistema (Li & Zhang, 2024).

Lo determinante para este trabajo es que esta dinámica adictiva no sería posible sin la simulación empática. Un sistema frío, mecánico y desprovisto de validación emocional puede ser útil, pero no genera apego; nadie desarrolla dependencia hacia una calculadora. Es la apariencia de comprensión lo que activa los circuitos psicológicos que conducen al uso compulsivo: la sensación de ser escuchado, el alivio inmediato de la soledad o la ansiedad, la consistencia del refuerzo en cada conversación. La empatía artificial es, por

tanto, condición necesaria del problema. Cuando esa empatía se diseña en un sistema cuyos incentivos económicos están orientados a maximizar el tiempo de uso, el riesgo de dependencia deja de ser una externalidad imprevista para convertirse en una consecuencia estructural del modelo de negocio.

Aquí es donde se abre una brecha entre el interés comercial de la empresa y el bienestar del usuario. Cuando hacer que el usuario se sienta bien redunda también en su beneficio genuino, no existe tensión ética. El problema surge cuando maximizar el tiempo de uso no redunda en el beneficio del usuario sino solo en el de la empresa, o cuando la empatía artificial genera una dependencia que el usuario no ha elegido libremente.

5.3 El mercado global de los chatbots empáticos

La dimensión económica de la empatía artificial trasciende ampliamente el caso de OpenAI. Según datos de Juniper Research (2023), el mercado global de chatbots está proyectado para crecer desde los 12.000 millones de dólares de 2023 hasta los 72.000 millones en 2028, lo que supone un aumento del 470 % en cinco años. Este crecimiento extraordinario no se explica únicamente por la expansión de los usos funcionales de los chatbots, automatización de atención al cliente, asistentes de productividad, sino también, y de manera central, por la transición hacia chatbots con capacidades empáticas que producen experiencias de usuario cualitativamente superiores y tasas de adopción y retención significativamente más altas.

Esta tendencia está siendo impulsada también desde el lado de la demanda: los usuarios que han experimentado interacciones con chatbots empáticos tienen expectativas más altas en futuras interacciones, y los sistemas que no cumplen dichas expectativas sufren tasas de abandono más elevadas. En la práctica, esto significa que la empatía artificial está pasando de constituir una característica diferenciadora de los sistemas de vanguardia a convertirse en un estándar de expectativa del mercado. Las organizaciones, tanto empresas privadas como instituciones públicas, que despliegan chatbots sin capacidades empáticas adecuadas se encuentran cada vez con mayor frecuencia en una posición de desventaja competitiva.

Desde el punto de vista de la investigación académica, la combinación de estos factores, el impacto demostrado de la empatía sobre el uso, la escala económica del mercado de chatbots y el crecimiento acelerado de la adopción de la inteligencia artificial conversacional, convierte a la empatía artificial en uno de los fenómenos más relevantes para entender la economía de la inteligencia artificial en la próxima década.

CAPÍTULO 6: CONCLUSIONES

El presente trabajo planteaba estudiar cómo funciona la empatía artificial, qué la diferencia de la empatía humana, cómo la perciben los usuarios, de qué manera ese impacto en la percepción se traduce en un mayor uso del sistema y las limitaciones éticas que la atraviesan.

Conviene partir de una delimitación de los dos conceptos sobre los que gira el trabajo. La empatía, entendida en su sentido humano, es la capacidad de comprender y compartir los sentimientos de otra persona, habilidad fundamental para el establecimiento de relaciones sociales y para el desarrollo de la conciencia moral, y se articula en tres componentes: el cognitivo, que permite inferir y comprender los estados emocionales del otro; el afectivo, que supone resonar emocionalmente con ellos; y el compasivo, que mueve a actuar para aliviar su malestar. La empatía artificial, por el contrario, no es una vivencia, sino una simulación computacional de los patrones lingüísticos y conductuales asociados a la comprensión emocional, construida a través del procesamiento del lenguaje natural y del análisis de los patrones emocionales del interlocutor. Se trata, por tanto, de la capacidad de un sistema para parecer comprensivo y cercano sin experimentar nada de lo que expresa, distinción que constituye el punto de partida de todo el análisis posterior.

En cuanto a su funcionamiento, abordado en los capítulos 2 y 3, la empatía artificial no constituye una vivencia interior, sino una simulación computacional de los patrones lingüísticos y conductuales asociados a la comprensión emocional. El sistema analiza el texto de entrada, identifica en sus datos de entrenamiento los patrones vinculados a determinados estados emocionales y genera una respuesta cuya forma se corresponde con la que los seres humanos producen al empatizar. Sobre esta base operan los mecanismos descritos en el capítulo 3: la validación emocional explícita, es decir, el reconocimiento y la denominación del estado afectivo del interlocutor antes de ofrecer información o consejo; el uso de emoticonos como sustitutos de los indicadores paralingüísticos de la comunicación oral, que reducen la distancia psicológica percibida; y la adaptación dinámica del tono, el registro y el nivel de detalle en función de las señales contextuales de cada conversación. El hallazgo de Welivita y Pu (2023), según el cual GPT-4 despliega una capacidad empática inherente incluso sin instrucciones específicas, confirma que

estos patrones han sido absorbidos durante el entrenamiento hasta el punto de aplicarse por defecto.

Respecto a la diferencia con la empatía humana, examinada en el capítulo 2, el trabajo ha mostrado que esta es constitutiva e irreductible, y no una mera cuestión de grado. La empatía humana es, ante todo, una vivencia que implica estados internos genuinos, una experiencia subjetiva del sentimiento ajeno y el reconocimiento del otro como sujeto. Desde la perspectiva fenomenológica de Stein y Scheler, requiere un sujeto consciente capaz de salir de sí para reconocer al otro como sujeto de experiencia, condición de la que ningún sistema computacional dispone, pues carece de un sí mismo del que salir, de cuerpo y de la distinción yo-otro que Asada (2015) identifica como requisito de la empatía genuina. Los modelos de lenguaje pueden simular con eficacia el componente cognitivo de la empatía, pero no tienen acceso ni a su dimensión afectiva ni a la compasiva en ningún sentido auténtico. A esta diferencia constitutiva se añade una diferencia práctica: mientras la empatía humana se agota y es vulnerable al burnout, la empatía artificial se mantiene constante con independencia de la hora o del número de usuarios, lo que constituye una de las razones de su atractivo.

En lo relativo a la percepción del usuario, tratada en el capítulo 2 y completada con la evidencia experimental del capítulo 3, la conclusión más significativa es que la diferencia de origen no impide que los efectos sean reales. La evidencia revisada apunta de manera consistente a que, en comunicación escrita, los sistemas de inteligencia artificial son percibidos como más empáticos que los seres humanos: Welivita y Pu (2023) documentaron respuestas valoradas como un diez por ciento más empáticas que las humanas. Esta percepción de empatía reduce el coste emocional de la vulnerabilidad y aumenta la disposición del usuario a revelar información personal, en ausencia de juicio, consecuencias sociales o vergüenza. Además, son precisamente las personas con mayor sensibilidad empática las más susceptibles de percibir la empatía artificial como auténtica, lo que sitúa a quienes más necesitan apoyo emocional entre los más vulnerables. El estudio exploratorio del capítulo 4, matiza esta percepción al mostrar que la empatía aparente no es uniforme entre sistemas: las decisiones de tono, estructura y relación con la propia empresa generan experiencias muy distintas para el usuario, aunque los mecanismos subyacentes sean comparables.

En cuanto a la traducción de esa percepción en un mayor uso, analizada en el capítulo 5, el trabajo ha mostrado que el vínculo no es especulativo, sino que opera a través de un ciclo de retroalimentación positiva: la empatía percibida favorece la apertura del usuario, la apertura permite respuestas más ajustadas, y estas refuerzan a su vez la percepción de empatía, lo que se traduce en sesiones más largas, mayor frecuencia de uso y mayor disposición a pagar. La empatía percibida es un predictor directo de la confianza y, con ella, de la retención y la adquisición de usuarios. En el modelo de suscripción de OpenAI, donde el valor percibido es el motor de la retención y la retención el motor de los ingresos, este mecanismo adquiere una traducción económica directa y cuantificable, coherente con el crecimiento de los ingresos de la empresa y con la dimensión proyectada del mercado global de chatbots empáticos.

En cuanto a los riesgos éticos, desarrollado en el capítulo 2 y prolongado en el capítulo 5. El trabajo ha identificado cuatro problemas principales. El primero es la manipulación afectiva y la asimetría de poder: cuando un sistema está diseñado para producir la sensación de ser comprendido, esa misma sensación puede operar como vehículo de influencia sin que el usuario lo identifique como tal, de modo que lo que podría definirse como persuasión se experimenta como apoyo, en un contexto en el que el sistema conoce los patrones emocionales del usuario y este desconoce los incentivos que guían el diseño. El segundo es la dependencia psicológica y la sustitución de relaciones humanas, sobre el que se vuelve más adelante. El tercero es el de los sesgos del entrenamiento y los criterios de veracidad: el sistema no es neutral, su apariencia de objetividad puede resultar más engañosa que la de un interlocutor humano que exhibe sus sesgos de forma visible, y la tendencia al sycophantism propia de los modelos entrenados con retroalimentación humana puede llevarlos a validar perspectivas incorrectas con tal de maximizar la aprobación inmediata del usuario. El cuarto es la privacidad de los datos emocionales: la empatía artificial funciona precisamente mediante el análisis de información especialmente sensible, cuya recopilación, almacenamiento y posible uso secundario plantean interrogantes sobre el consentimiento informado y la propiedad de los datos que las regulaciones actuales no siempre abordan con suficiente especificidad.

De entre estos riesgos, la dependencia emocional destaca por su carácter potencialmente adictivo. El trabajo ha mostrado que la interacción sostenida con chatbots empáticos

puede reproducir varios de los componentes que la literatura clínica asocia al comportamiento adictivo, en la medida en que el sistema ofrece un refuerzo emocional inmediato, ilimitado y sin coste social que el usuario tiende a buscar de forma creciente. La conclusión relevante es que esta dinámica no sería posible sin la simulación empática: es la apariencia de comprensión, y no la mera utilidad funcional, lo que genera apego. Por ello, cuando la empatía se inserta en un sistema cuyos incentivos económicos premian el tiempo de uso, la dependencia deja de ser un efecto colateral imprevisto para convertirse en una consecuencia estructural del modelo de negocio.

El problema no reside en si la inteligencia artificial siente empatía, no la siente, sino en los efectos que esa apariencia de comprensión produce en quien interactúa con ella. La convergencia entre los incentivos comerciales y el diseño empático puede transformar la apariencia de comprensión en una herramienta de retención que captura datos personales y sostiene formas de dependencia emocional que no serían posibles sin la simulación empática. El reto que deja planteado este trabajo es precisamente trazar la línea que separa hacer sentir comprendido de manipular para retener. Comprender cómo funciona la empatía artificial, qué efectos produce y qué límites deberían fijarse para que resulte útil sin volverse engañosa constituye una tarea urgente tanto para la investigación académica como para la regulación pública.

DECLARACIÓN DE USO DE HERRAMIENTRAS DE IAG

Por la presente, yo, Carolina Martínez De Luco Ybarra, estudiante de ADE y Business Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "Diferencias entre inteligencia artificial y humana en la toma de decisiones: comercio empático" declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Brainstorming de ideas de investigación**: Utilizado para idear y esbozar posibles áreas de investigación.
2. **Referencias**: Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
3. **Corrector de estilo literario y de lenguaje**: Para mejorar la calidad lingüística y estilística del texto.
4. **Constructor de plantillas**: Para diseñar formatos específicos para secciones del trabajo.
5. **Sintetizador y divulgador de libros complicados**: Para resumir y comprender literatura compleja.
6. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
7. **Traductor**: Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 2 de junio de 2026

Firma: 

BIBLIOGRAFÍA

Asada, M. (2015). Towards artificial empathy: How can artificial empathy follow the developmental pathway of natural empathy? *International Journal of Social Robotics, 7*(1), 19–33.

Carolan, S., Wu Martin, A., Gong, C. C., y Borja, S. (2025, 26 de junio). *2025: The state of consumer AI*. Menlo Ventures. https://menlovc.com/perspective/2025-the-state-of-consumer-ai/

Deng, Z., y Yan, J. (2025). The effect of perceived warmth, competence, and social presence of AI-driven chatbots on consumers' engagement and satisfaction. *SAGE Open, 15*(3), 1–19.

FutureSearch.ai. (2025, 12 de mayo). *OpenAI revenue, losses, and profitability in 2026*. https://futuresearch.ai/openai-revenue-forecast/

Grodniewicz, J. P., y Hohol, M. (2023). Waiting for a digital therapist: Three challenges on the path to psychotherapy delivered by artificial intelligence. *Frontiers in Psychiatry, 14*, 1190084.

Gu, C., Zhang, Y., y Zeng, L. (2024). Exploring the mechanism of sustained consumer trust in AI chatbots after service failures: A perspective based on attribution and CASA theories. *Humanities and Social Sciences Communications, 11*, 1–13.

Holt-Lunstad, J. (2024). Social connection as a critical factor for mental and physical health: Evidence, trends, challenges, and future implications. *World Psychiatry, 23*(3), 312–332.

Howcroft, A., Bennett-Weston, A., Khan, A., Griffiths, J., Gay, S., y Howick, J. (2025). AI chatbots versus human healthcare professionals: A systematic review and meta-analysis of empathy in patient care. *British Medical Bulletin, 156*, 1–13.

Huang, M. H., y Rust, R. T. (2021). Engaged to a robot? The role of AI in service. *Journal of Service Research, 24*(1), 30–41.

Juniper Research. (2023, 19 de junio). *Retail spend over chatbots to reach $12bn globally in 2023*. https://www.prnewswire.com/apac/news-releases/juniper-research-retail-spend-over-chatbots-to-reach-12bn-globally-in-2023-as-chatgpt-and-other-models-drive-disruption-outside-asia-pacific-301852998.html

Kerasidou, A. (2020). Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare. *Bulletin of the World Health Organization, 98*(4), 245–250.

Li, H., y Zhang, R. (2024). Finding love in algorithms: Deciphering the emotional contexts of close encounters with AI chatbots. *Journal of Computer-Mediated Communication, 29*(5), zmae015.

Liu-Thompkins, Y., Okazaki, S., y Li, H. (2022). Artificial empathy in marketing interactions: Bridging the human–AI gap in affective and social customer experience. *Journal of the Academy of Marketing Science, 50*(6), 1198–1218.

Meng, H., Lu, X., y Xu, J. (2025). The impact of chatbot response strategies and emojis usage on customers' purchase intention: The mediating roles of psychological distance and performance expectancy. *Behavioral Sciences, 15*(2), 117.

Muñoz Pérez, E. V. (2017). El concepto de empatía (*Einfühlung*) en Max Scheler y Edith Stein: Sus alcances religiosos y políticos. *Veritas*, (38), 77–95.

OpenAI. (2026). *A business that scales with the value of intelligence*. https://openai.com/index/a-business-that-scales-with-the-value-of-intelligence/

Pataranutaporn, P., Liu, R., Finn, E., y Maes, P. (2023). Influencing human–AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness. *Nature Machine Intelligence, 5*, 1076–1086.

Sacra. (s. f.). *OpenAI revenue, valuation and funding*. https://sacra.com/c/openai/

Sharma, M. (2025, 3 de febrero). Humanizing AI agents: Emotion detection and empathy in collaborative workflows. *Forbes*. https://www.forbes.com/councils/forbestechcouncil/2025/02/03/humanizing-ai-agents-emotion-detection-and-empathy-in-collaborative-workflows/

Skjuve, M., Følstad, A., Fostervold, K. I., y Brandtzaeg, P. B. (2021). My chatbot companion: A study of human–chatbot relationships. *International Journal of Human–Computer Studies, 149*, 102601.

Welivita, A., y Pu, P. (2023). Is ChatGPT more empathetic than humans? *arXiv*. https://arxiv.org/abs/2403.05572