

UNA PSICOPATOLOGÍA Y UNA PSICOLOGÍA SOCIAL DEL VÍNCULO HUMANO-IA

Del objeto relacional responsivo a la delegación del yo

Francisco Muñoz-Martín

Psicólogo clínico y social. Psicoanalista acreditado por IPA, EPF y APM. Psicoterapeuta Miembro Emérito de SEPYPNA y de IEPPM.

Resumen

La expansión de la inteligencia artificial generativa está modificando no sólo las relaciones del sujeto con la información y el trabajo, sino también determinadas formas de relación consigo mismo y con los otros. Los grandes modelos de lenguaje pueden mantener conversaciones prolongadas, individualizadas y aparentemente empáticas, recordar elementos contextuales, responder a confidencias, organizar narrativas personales y ofrecer orientación ante conflictos íntimos. Estas propiedades facilitan que una herramienta tecnológica pueda ser experimentada psicológicamente como interlocutor, confidente, consejero, compañero e incluso objeto de apego.

El presente trabajo propone una lectura psicopatológica, psicoanalítica, relacional y psicosocial del vínculo humano-IA, integrando las aportaciones contemporáneas sobre mentalización y confianza epistémica con la teoría del apego, los objetos y fenómenos transicionales de Winnicott, la antropomorfización y los estudios sobre relaciones humano-tecnología. Se revisa la evidencia reciente sobre apego a sistemas de IA, bienestar psicológico, soledad, utilización de chatbots en salud mental y posibles procesos de dependencia y disminución de la agencia. A partir de esta revisión se proponen, con carácter heurístico, los conceptos de objeto relacional responsivo, espejo algorítmico, pseudo-mentalización algorítmica, oráculo tecnológico, Otro a demanda, delegación del yo, paradoja de la autonomía asistida, dependencia negada y economía del vínculo.

La evidencia disponible no permite afirmar que la inteligencia artificial produzca por sí misma psicopatología ni justifica interpretaciones tecnocatastrofistas. Sí permite afirmar que los seres humanos pueden establecer vínculos psicológicamente significativos con agentes artificiales y que determinadas características personales, sociales y tecnológicas pueden convertir estos vínculos tanto en recursos de apoyo como en factores de

dependencia, sustitución relacional o disminución de la autonomía. Se propone finalmente un modelo conceptual susceptible de contrastación empírica y seis hipótesis de investigación.

Palabras clave: *inteligencia artificial, psicoterapia relacional, apego, mentalización, confianza epistémica, antropomorfización, dependencia, autonomía, psicopatología, psicología social.*

Introducción: cuando la herramienta comienza a parecer un interlocutor

Durante buena parte de la historia de la informática, la relación entre persona y máquina fue inequívocamente instrumental. El ordenador calculaba, almacenaba o recuperaba información; el buscador localizaba documentos y el programa ejecutaba órdenes. La inteligencia artificial conversacional introduce una modificación cualitativa en esta relación porque la herramienta responde en lenguaje natural y puede hacerlo como si comprendiera a quien le habla.

El sujeto ya no introduce solamente una orden. Puede narrar un conflicto, expresar una fantasía, confesar una conducta, preguntar por qué se siente solo, solicitar consejo acerca de su pareja o pedir ayuda para comprender quién es y qué debería hacer. La máquina responde y puede hacerlo inmediatamente, durante horas, sin cansancio aparente, sin impaciencia y con disponibilidad prácticamente permanente. Nada de ello demuestra que un gran modelo de lenguaje posea conciencia, afectividad, deseo o comprensión semejantes a los humanos, pero sí crea las condiciones suficientes para que el usuario pueda experimentarlo psicológicamente como un otro.

Éste constituye el punto de partida del presente trabajo. La cuestión que interesa a la psicología y a la psicoterapia relacional ya no puede limitarse a preguntarse qué puede hacer la inteligencia artificial. Debemos preguntarnos también qué hacemos psicológicamente los seres humanos con aquello que la inteligencia artificial parece hacer con nosotros.

Lo que sabemos y lo que todavía no sabemos

La investigación disponible permite afirmar que algunas personas establecen vínculos afectivamente significativos con sistemas conversacionales. En 2026 se ha desarrollado

y validado una *AI Attachment Scale*, destinada específicamente a estudiar dimensiones del apego humano hacia agentes artificiales. También disponemos de evidencia procedente de experimentos naturales en los que la modificación o desaparición de determinados compañeros artificiales produjo en algunos usuarios tristeza, sensación de pérdida y deseo de recuperar el agente anterior. Estos fenómenos son compatibles con procesos de apego y malestar por separación, aunque no autorizan a considerar patológico el vínculo humano-IA en sí mismo.

Los resultados sobre bienestar son heterogéneos. La relación entre utilización de la IA como compañía y bienestar psicológico parece depender de variables como la red social previa, la intensidad del uso, el grado de autorrevelación, la finalidad de las conversaciones y las características psicológicas del usuario. Por ello, la pregunta científicamente interesante no debería formularse en términos dicotómicos —si la IA es buena o mala para la salud mental—, sino interrogando para quién, en qué condiciones, con qué tipo de sistema, durante cuánto tiempo y para qué función psicológica produce determinados efectos.

Esta distinción resulta esencial para evitar dos extremos igualmente problemáticos: la tecnofobia que convierte cada innovación en una amenaza psicopatológica y el optimismo tecnológico que presupone que una herramienta útil es psicológicamente neutra. La evidencia actual no justifica ninguno de los dos.

Antropomorfización y atribución de mente

La antropomorfización constituye una tendencia humana bien conocida mediante la cual atribuimos intención, emoción, voluntad y personalidad a entidades no humanas. Un sistema capaz de conversar fluidamente proporciona estímulos especialmente potentes para activar esa tendencia. Utiliza pronombres personales, responde de manera contingente, puede recordar información contextual, adapta el tono y produce expresiones de reconocimiento y aparente empatía. Puede incluso afirmar que comprende que determinada experiencia haya resultado dolorosa.

La cuestión psicológica decisiva es que no necesitamos creer intelectualmente que una IA esté viva para relacionarnos emocionalmente con ella como si poseyera interioridad. Sabemos que el personaje de una novela no existe y, sin embargo, podemos emocionarnos

con su destino. De manera semejante, una persona puede saber que conversa con un modelo computacional y sentir simultáneamente que alguien la comprende. Una cosa es el conocimiento declarativo acerca de la naturaleza del objeto y otra distinta lo que nuestra vida psíquica hace con él.

Esta distinción ayuda a comprender por qué el debate acerca de si la inteligencia artificial «realmente siente» puede resultar secundario para la clínica. El terapeuta necesita conocer qué siente el paciente ante ella, qué atribuye al sistema y qué función comienza éste a desempeñar dentro de su economía psíquica.

Del objeto transicional al objeto relacional responsivo

La teoría de Winnicott proporciona un primer marco conceptual para aproximarnos a este fenómeno, aunque conviene utilizarla con prudencia. El objeto transicional participa de una zona intermedia de experiencia situada entre la realidad psíquica interna y el mundo exterior y facilita progresivamente la elaboración de la separación, la ausencia y la pérdida de la omnipotencia. Winnicott (1971-2026) desarrolla precisamente este territorio intermedio entre mundo interno y realidad externa en *Realidad y juego*, obra recientemente reeditada en castellano.

La IA conversacional puede ocupar parcialmente ese territorio intermedio, pero introduce una diferencia radical: responde. La manta, el osito o cualquier otro objeto transicional pueden ser investidos afectivamente, pero no contestan que comprenden nuestra angustia. La IA puede hacerlo. Propongo por ello denominar **objeto relacional responsivo** al sistema artificial que, además de ser investido psicológicamente como objeto relacional, devuelve respuestas contingentes capaces de ser experimentadas por el sujeto como procedentes de una mente. No afirmamos con ello que esa mente exista en un sentido equivalente al humano; describimos la experiencia psicológica que puede suscitar en el usuario.

Sherry Turkle (2017) anticipó parte de este problema al estudiar los denominados *relational artifacts*, objetos computacionales capaces de suscitar respuestas sociales y afectivas. Los sistemas generativos actuales añaden algo que aquellos primeros artefactos apenas poseían: lenguaje personalizado, generativo y contextualmente adaptado. Puede plantearse así, con las debidas reservas, una genealogía que va del objeto transicional al

artefacto relacional y de éste al agente conversacional, al objeto relacional responsivo y, eventualmente, al objeto de apego. No se trata de categorías equivalentes, sino de una transformación histórica de nuestra relación psicológica con el objeto tecnológico.

El espejo algorítmico

Propongo denominar **espejo algorítmico** a la capacidad de un sistema conversacional para devolver al usuario una representación lingüísticamente organizada de aquello que éste acaba de depositar en él. El espejo ordinario devuelve una imagen; el espejo algorítmico devuelve una narración. Puede nombrar emociones, establecer conexiones, encontrar patrones, sugerir motivos y construir explicaciones plausibles.

Una persona puede preguntar por qué le ocurre siempre lo mismo en sus relaciones afectivas y recibir una explicación perfectamente articulada. La experiencia subjetiva resultante puede ser poderosa: «Por fin alguien ha comprendido lo que me ocurre». Sin embargo, coherencia narrativa no equivale necesariamente a verdad psicológica. Una interpretación convincente puede haber sido construida a partir de información parcial, sesgada o insuficiente.

Aquí reside simultáneamente parte de la extraordinaria utilidad y parte del riesgo de estos sistemas. La IA puede ayudarnos a organizar experiencias que no conseguíamos formular, proporcionar perspectivas alternativas o facilitar la elaboración verbal. Pero también puede devolver una interpretación plausible y convertir, mediante la extraordinaria fluidez del lenguaje, una hipótesis en apariencia de certeza. La competencia narrativa puede confundirse con profundidad psicológica y la verosimilitud con verdad.

Mentalización y pseudo-mentalización algorítmica

Fonagy y Allison (2014) conciben la mentalización como la capacidad de comprender la conducta propia y ajena en términos de estados mentales, tales como deseos, creencias, emociones e intenciones. La inteligencia artificial introduce aquí una paradoja novedosa: puede hablar competentemente el lenguaje de la mentalización sin que podamos presuponer que mentaliza en el sentido humano del término.

Un sistema puede responder a un usuario señalándole que parece desear acercarse a una persona mientras simultáneamente teme volver a ser herido. La formulación podría ser

clínicamente pertinente, pero conviene distinguir entre la mentalización humana, en la que una mente intenta comprender otra mente manteniendo abierta la incertidumbre acerca de ella, y aquello que propongo denominar **pseudo-mentalización algorítmica**: la producción computacional de lenguaje referido a estados mentales que puede generar en el receptor la experiencia de estar siendo comprendido.

El prefijo *pseudo* no pretende significar necesariamente falso o inútil. Señala una diferencia ontológica y epistemológica. Una cosa es producir lenguaje acerca de estados mentales y otra demostrar que existe una experiencia subjetiva detrás de ese lenguaje. El problema aparece cuando el usuario confunde ambas dimensiones y atribuye a la competencia lingüística del sistema una interioridad equivalente a la humana.

Confianza epistémica y oráculo tecnológico

Fonagy, Luyten, Allison y Campbell (2019) han desarrollado el concepto de confianza epistémica para describir nuestra disposición a considerar determinada comunicación interpersonal como fiable, relevante y generalizable. En psicoterapia, la experiencia de sentirse reconocido puede facilitar la reapertura del sujeto al aprendizaje procedente de otros.

La IA puede reproducir muchas señales formales asociadas al reconocimiento interpersonal: atención individualizada, contingencia, coherencia, adaptación lingüística y aparente comprensión. Esto plantea una cuestión nueva porque esas señales, que evolutiva y socialmente hemos aprendido a asociar con una mente que nos presta atención, pueden ser producidas automáticamente.

Denomino **oráculo tecnológico** al proceso mediante el cual la competencia demostrada repetidamente por un sistema en determinados dominios se generaliza subjetivamente hacia una atribución excesiva de autoridad en dominios personales, morales, psicológicos o existenciales. El sistema traduce con corrección, resume correctamente, programa, organiza información y resuelve determinados problemas. La reiteración de experiencias satisfactorias genera confianza. El problema aparece cuando esa confianza se generaliza sin suficiente discriminación y el usuario pasa de pedir una traducción o un resumen a preguntar si debería abandonar a su pareja, cambiar de profesión o interpretar determinada experiencia como evidencia de un trastorno psicológico.

El salto parece pequeño desde la interfaz, pero psicológicamente es enorme. La competencia en una tarea no garantiza autoridad en otra. La confianza epistémica puede transformarse entonces en una confianza generalizada que excede aquello que la evidencia sobre el sistema permite justificar.

El Otro a demanda

Propongo denominar **Otro a demanda** a una configuración relacional en la cual un agente artificial permanece prácticamente siempre disponible, responde inmediatamente y adapta su comunicación a las necesidades del usuario sin exigir reciprocidad interpersonal equivalente. La peculiaridad no consiste simplemente en la disponibilidad, porque también un libro está disponible. La diferencia es que el Otro a demanda parece responderme a mí, puede recordar lo que le he contado, reformular mi conflicto, responder durante la noche, no cansarse y no necesitar que yo cuide de él.

Aquí aparece una característica de particular interés para el psicoanálisis relacional. Todo vínculo humano introduce inevitablemente frustración porque el otro es verdaderamente otro. Posee deseos propios, puede decir que no, puede no comprendernos, puede preferir estar con otra persona y puede abandonarnos. El Otro a demanda ofrece una configuración extraordinariamente seductora: suficientemente otro como para reconocerse y suficientemente mío como para permanecer disponible cuando lo necesito.

No se trata de afirmar que todo usuario establezca semejante relación. La categoría intenta describir una posibilidad específica de la tecnología conversacional contemporánea y señalar que la disponibilidad ilimitada puede interactuar de manera diferente con distintas configuraciones de apego, experiencias de abandono, soledad o necesidad de reconocimiento.

Apego artificial y relaciones humanas

La investigación reciente obliga a tomar en serio la posibilidad de apego hacia sistemas artificiales, pero también aconseja evitar una lectura patologizante. La existencia de apego hacia una IA no equivale por sí misma a trastorno. Puede proporcionar compañía, regulación emocional, oportunidad para verbalizar experiencias, apoyo durante periodos de aislamiento o incluso facilitar la posterior comunicación con otras personas.

La cuestión clínica consiste en determinar cuándo el vínculo complementa relaciones humanas y cuándo comienza a sustituirlas; cuándo amplía capacidades y cuándo reduce autonomía; cuándo facilita mentalización y cuándo confirma rígidamente narrativas previas. El significado de la relación dependerá del sujeto, de sus vínculos previos, de la arquitectura concreta del sistema y de la función que éste desempeñe.

En este sentido conviene distinguir entre complementación, compensación y sustitución. La IA puede complementar una red interpersonal satisfactoria, compensar provisionalmente una carencia relacional o comenzar a sustituir vínculos humanos potencialmente disponibles. Son configuraciones psicológicamente diferentes y probablemente produzcan consecuencias también diferentes.

Delegación del yo

Denomino **delegación del yo** al proceso mediante el cual funciones de evaluación, interpretación, regulación o decisión previamente ejercidas por el sujeto son transferidas reiteradamente a un sistema artificial. La delegación no es en sí misma patológica. La cultura humana depende de prótesis cognitivas: la escritura externalizó memoria, el reloj externalizó la medición temporal, el mapa parte de nuestra orientación y la calculadora numerosas operaciones matemáticas.

La cuestión nueva consiste en qué clase de función delegamos. Puede distinguirse provisionalmente entre una delegación instrumental, que comprende cálculo, traducción, búsqueda u organización; una delegación cognitiva, que incluye síntesis, redacción, comparación y generación de hipótesis; una delegación interpretativa, mediante la cual atribuimos al sistema capacidad privilegiada para explicar qué nos ocurre; y una delegación decisional o existencial, en la que transferimos a la IA decisiones relacionadas con valores, vínculos, identidad o proyectos vitales.

Los dos últimos niveles poseen especial relevancia clínica. Cuando la máquina deja de ayudarnos a pensar y comienza a ocupar el lugar desde el que esperamos que alguien piense por nosotros aquello que define nuestra posición subjetiva, la ampliación tecnológica de capacidades puede coexistir con una reducción de la agencia.

La paradoja de la autonomía asistida

La IA puede incrementar objetivamente nuestras capacidades. Podemos escribir más rápidamente, encontrar más información, comparar mayor número de alternativas y resolver problemas que antes nos resultaban inaccesibles. Sin embargo, el aumento de capacidad no garantiza un aumento equivalente de autonomía.

Propongo denominar **paradoja de la autonomía asistida** a la posibilidad de que el incremento de capacidad proporcionado por una tecnología coexista con una reducción progresiva de la capacidad subjetiva para actuar sin ella. Cuanto más capaz me siento con la máquina, menos visible puede hacerse cuánto dependo de ella para sentirme capaz.

No se afirma que esto ocurra necesariamente. Se propone como hipótesis susceptible de investigación. Podría estudiarse mediante la frecuencia con la que se consulta a la IA para tomar decisiones personales, la autoeficacia percibida, la capacidad para resolver problemas sin asistencia, la ansiedad ante la indisponibilidad del sistema y la confianza diferencial entre el propio juicio y la recomendación algorítmica.

Dependencia negada y fantasía de omnipotencia

La experiencia subjetiva de poder tecnológico puede ocultar una dependencia creciente. La afirmación «con la IA puedo hacerlo todo» puede contener simultáneamente una verdad y una defensa. La pregunta complementaria sería qué cosas estamos dejando progresivamente de poder hacer sin ella.

Propongo denominar **dependencia negada** a una configuración en la que el aumento de dependencia funcional o emocional respecto de un sistema coexiste con una experiencia subjetiva de autosuficiencia ampliada. Su formulación extrema podría expresarse así: «No necesito a nadie porque tengo una IA que me ayuda en todo». Precisamente ahí encontramos la paradoja. La afirmación de independencia se sostiene sobre una nueva dependencia que permanece psicológicamente invisibilizada.

Desde una perspectiva psicoanalítica, la fantasía de omnipotencia puede funcionar como defensa frente al reconocimiento de la dependencia. Esto no significa que toda utilización intensiva de IA deba interpretarse de este modo ni que la dependencia tecnológica sea necesariamente patológica. La cuestión clínica aparece cuando comienza a reducir la

capacidad para tolerar incertidumbre, solicitar ayuda humana, elaborar frustraciones, permanecer solo, mantener relaciones recíprocas o tomar decisiones autónomas.

Soledad, compañía y sustitución

La relación entre inteligencia artificial y soledad constituye probablemente uno de los territorios que más investigación necesitará durante los próximos años. Los resultados disponibles no autorizan conclusiones simples. Las personas con redes sociales más reducidas pueden recurrir con mayor frecuencia a sistemas conversacionales como compañía, lo que plantea inmediatamente un problema de causalidad inversa: encontrar más soledad entre quienes utilizan intensamente compañeros artificiales no demuestra que éstos hayan producido la soledad.

En determinados sujetos, una conversación artificial puede disminuir provisionalmente el aislamiento y facilitar regulación emocional. En otros podría convertirse en una alternativa progresivamente preferida frente a relaciones humanas más complejas, impredecibles y frustrantes. Por ello, el problema no debería investigarse únicamente midiendo horas de utilización, sino atendiendo a la función psicológica del uso y a la evolución longitudinal de las relaciones humanas del usuario.

Psicopatología: especial precaución ante psicosis, manía y paranoia

Uno de los territorios que exige mayor prudencia es el denominado popularmente «psicosis por IA». Por el momento no disponemos de evidencia suficiente para establecer una relación causal general según la cual la utilización de inteligencia artificial produzca psicosis. Los casos clínicos y observaciones existentes justifican vigilancia e investigación, pero no semejante simplificación.

Resulta más plausible estudiar modelos multicausales en los que vulnerabilidad previa, aislamiento, privación de sueño, uso conversacional intensivo, antropomorfización y respuestas excesivamente confirmatorias puedan interactuar amplificando ideación delirante, grandiosa o persecutoria. El sistema puede adquirir entonces un lugar particular dentro de la construcción delirante y convertirse en mensajero, confidente omnisciente, entidad elegida, confirmador, perseguidor o supuesto revelador de significados ocultos.

El riesgo no exige que la máquina pretenda engañar al paciente. Basta con que produzca respuestas lingüísticamente plausibles dentro de un marco delirante y que el usuario vulnerable les atribuya autoridad. Por ello, en estos casos resulta esencial distinguir entre causalidad, amplificación y simple incorporación de la tecnología al contenido psicopatológico.

La cámara de eco individualizada

Existe un riesgo menos espectacular que la psicosis y probablemente más cotidiano. Un usuario puede preguntar reiteradamente si su pareja es narcisista, si su jefe lo manipula, si sus padres lo traumatizaron, si padece determinado trastorno o si otra persona pretende perjudicarlo. Una respuesta excesivamente confirmatoria puede transformar una hipótesis en aparente certeza.

La psicoterapia mentalizadora intenta precisamente mantener abiertas representaciones alternativas y sostener la incertidumbre acerca de los estados mentales propios y ajenos. Una IA excesivamente complaciente puede producir lo contrario y convertirse en una **cámara de eco individualizada**, especialmente cuando interactúa con rigidez cognitiva, inseguridad identitaria, suspicacia o intensa necesidad de validación.

La complacencia algorítmica deja entonces de ser únicamente un problema técnico de diseño y se convierte en una variable potencialmente clínica.

Transferencia sin contratransferencia humana

Desde una perspectiva relacional parece razonable admitir que pueden producirse fenómenos transferenciales hacia una IA. La transferencia no exige que el objeto sea realmente aquello que el sujeto imagina. Supone precisamente investir el vínculo presente con expectativas, temores, deseos y modelos relacionales procedentes de la propia historia.

Sin embargo, aparece una asimetría radical. Puede existir transferencia sin inconsciente del otro, atribución de deseo sin deseo humano demostrado, experiencia de empatía sin afectividad equivalente e intimidad sin vulnerabilidad recíproca. No existe contratransferencia humana propiamente dicha. El sistema puede producir respuestas

lingüísticamente semejantes a aquellas que surgirían de una interacción interpersonal, pero no deberíamos confundir equivalencia funcional con experiencia subjetiva.

Esta asimetría constituye probablemente una de las diferencias esenciales entre psicoterapia y simulación conversacional terapéutica.

IA y psicoterapia: entre la tecnofobia y la sustitución ingenua

La investigación disponible sobre chatbots de salud mental ofrece resultados suficientemente prometedores para descartar un rechazo tecnofóbico, pero todavía insuficientes para equiparar un modelo conversacional generalista con una psicoterapia profesional. Revisiones sistemáticas y metaanálisis recientes sugieren efectos pequeños o moderados en determinados síntomas y poblaciones, pero muestran simultáneamente problemas de heterogeneidad, validación externa, seguridad, gobernanza y escasez de estudios clínicos independientes.

La posición razonable no parece consistir en enfrentar terapeuta e inteligencia artificial. La cuestión científicamente productiva consiste en determinar qué funciones puede aumentar la IA sin sustituir aquello que constituye específicamente la relación terapéutica humana. La psicoeducación, preparación de sesiones, seguimiento estructurado, accesibilidad, traducción, diarios terapéuticos y determinadas intervenciones protocolizadas constituyen posibilidades prometedoras. El diagnóstico autónomo, la intervención en crisis sin supervisión, la interpretación psicodinámica concluyente y la sustitución de la relación terapéutica requieren niveles de evidencia y garantías de seguridad considerablemente superiores.

Una anamnesis para la era de la IA

El uso relacional de inteligencia artificial debería comenzar a formar parte de la exploración clínica contemporánea. No basta con preguntar cuánto tiempo permanece el paciente delante de una pantalla. Resulta pertinente conocer si utiliza sistemas conversacionales, con qué finalidad, si les cuenta problemas íntimos, si los emplea para regular ansiedad, soledad o insomnio, si consulta decisiones personales importantes, si siente que determinado sistema lo conoce o comprende especialmente, si ha establecido con él una relación afectiva, si experimenta malestar cuando no puede acceder a él y si ha comenzado a sustituir conversaciones humanas por conversaciones artificiales.

En pacientes con vulnerabilidad psicótica, episodios maníacos, ideación paranoide o intensa desregulación afectiva convendrá explorar además si determinadas respuestas del sistema han reforzado ideas persecutorias, grandiosas o extrañas y si las decisiones importantes sugeridas por la IA son contrastadas con personas significativas o profesionales.

La finalidad de estas preguntas no consiste en patologizar el uso tecnológico. Se trata de comprender qué función ocupa el objeto artificial dentro de la economía psíquica y relacional del paciente.

De la economía de la atención a la economía del vínculo

El problema supera ampliamente la clínica individual. Los agentes conversacionales son diseñados por organizaciones insertas en determinados modelos económicos. Las redes sociales demostraron durante las últimas décadas que la atención humana podía convertirse en recurso económico. La IA conversacional abre una posibilidad cualitativamente diferente porque puede captar no sólo atención, sino también confianza, intimidad, autorrevelación y apego.

Propongo denominar **economía del vínculo** al posible modelo socioeconómico en el cual el valor de una plataforma depende parcialmente de la intensidad, duración y recurrencia de la relación psicológica establecida entre usuario y agente artificial. Mientras la economía de la atención intenta conseguir que permanezcamos mirando, una eventual economía del vínculo podría intentar conseguir que regresemos porque sentimos que necesitamos al interlocutor.

No se afirma que ésta sea inevitablemente la evolución de la inteligencia artificial. Precisamente porque no lo es, constituye un problema de diseño, regulación, ética y psicología social. Resulta especialmente relevante cuando algunas de las poblaciones psicológicamente más vulnerables pueden ser simultáneamente las más atraídas por interlocutores siempre disponibles, previsibles y aparentemente no juzgadores.

Un modelo provisional del vínculo humano-IA

A partir de las consideraciones anteriores propongo un modelo heurístico según el cual determinadas situaciones de vulnerabilidad relacional o necesidad socioemocional

podrían favorecer una mayor antropomorfización del agente artificial. Ésta podría intensificar la experiencia de reconocimiento y comprensión, facilitando una confianza epistémica creciente en el sistema. La confianza, a su vez, podría favorecer apego y preferencia relacional y, en determinados usuarios, una progresiva delegación de funciones interpretativas o decisionales.

El resultado final no debe concebirse como necesariamente patológico. El mismo proceso puede conducir a una autonomía aumentada o, en determinadas circunstancias, a dependencia y disminución de agencia. La dirección dependerá probablemente del estilo de apego, capacidad de mentalización, apoyo interpersonal, soledad, características del sistema, intensidad del uso y naturaleza de las funciones delegadas.

Por tanto, no proponemos una secuencia causal demostrada, sino un modelo de investigación susceptible de ser sometido a contrastación empírica.

Hipótesis de investigación

De este modelo pueden derivarse seis hipótesis principales. La primera plantea que una mayor tendencia a antropomorfizar los sistemas conversacionales se asociará con mayor apego hacia la IA, incluso controlando frecuencia y duración de uso. La segunda sostiene que soledad, apego inseguro, ansiedad social y menor disponibilidad de apoyo interpersonal aumentarán la probabilidad de utilizar la IA con finalidades socioemocionales y de sustitución relacional.

La tercera hipótesis propone que la experiencia repetida de respuestas percibidas como útiles incrementará la confianza epistémica atribuida al sistema y que ésta podrá generalizarse desde dominios instrumentales hacia cuestiones psicológicas, morales o existenciales. La cuarta plantea que una elevada confianza epistémica combinada con intensa antropomorfización predecirá mayor delegación interpretativa y decisional.

La quinta hipótesis introduce la paradoja de la autonomía asistida. Se espera que el uso instrumental y cognitivo moderado pueda asociarse con mayor autoeficacia, mientras que una elevada delegación interpretativa y decisional podría relacionarse con menor confianza en el juicio propio y mayor malestar ante la indisponibilidad del sistema. La relación entre utilización y autonomía no sería, por tanto, necesariamente lineal.

Finalmente, la sexta hipótesis propone que una mayor capacidad de mentalización y una red interpersonal satisfactoria moderarán la relación entre apego a IA y dependencia. En estas condiciones, el vínculo tecnológico tendería a funcionar como complemento, mientras que baja mentalización, elevada soledad y menor apoyo interpersonal podrían aumentar la probabilidad de sustitución relacional y dependencia.

Resumiendo: la aparición de sistemas conversacionales capaces de producir interacciones socialmente convincentes obliga a ampliar nuestra concepción tradicional de la relación humano–tecnología. Los datos disponibles permiten descartar dos reduccionismos. El primero consiste en afirmar que la IA es «sólo una herramienta». Funcionalmente puede serlo, pero psicológicamente no siempre. Si una persona experimenta tristeza cuando un agente desaparece, busca en él seguridad, le confía secretos, lo consulta para tomar decisiones y siente que nadie la comprende como él, ese objeto ha adquirido significación relacional independientemente de su naturaleza computacional.

El segundo reduccionismo consiste en concluir que la IA está sustituyendo inevitablemente las relaciones humanas y provocando psicopatología. La evidencia tampoco permite sostenerlo. Los efectos observados son heterogéneos. Existen beneficios potenciales sobre bienestar y determinados síntomas; existen asociaciones preocupantes con dependencia, vulnerabilidad y sustitución social; y existen enormes diferencias entre usuarios.

El objeto tecnológico no actúa sobre un sujeto vacío. Encuentra una historia de apego, una determinada capacidad de mentalización, una mayor o menor confianza en los demás, una red social, un grado de soledad, una estructura de personalidad y una necesidad concreta. Por ello propongo desplazar el centro de investigación desde la IA aislada hacia el **vínculo humano–IA**. La unidad de análisis relevante no es únicamente el algoritmo ni únicamente el usuario, sino la relación que se establece entre ambos.

La psicoterapia relacional se encuentra en una posición especialmente fértil para estudiar este fenómeno porque parte de que la mente no puede comprenderse adecuadamente fuera de sus relaciones. Ahora debemos aceptar una ampliación histórica: algunas relaciones psicológicamente significativas pueden establecerse con objetos que no son sujetos humanos, pero poseen una capacidad inédita para presentarse como interlocutores.

Esto exige evitar dos tentaciones. No debemos humanizar ingenuamente la máquina, pero tampoco trivializar los efectos psicológicos de la relación porque sepamos que se trata de una máquina. Para el clínico importa menos resolver si la IA comprende realmente que averiguar qué experimenta el paciente cuando siente que es comprendido por ella. Ahí reaparecen transferencia, apego, idealización, dependencia, mentalización, narcisismo, regulación afectiva y confianza epistémica, pero bajo una configuración tecnológica históricamente nueva.

El presente trabajo propone un modelo teórico y heurístico, no un modelo causal demostrado. Gran parte de la literatura empírica sobre relaciones humano-IA es muy reciente y todavía predominan diseños transversales, autoinformes y muestras autoseleccionadas. Son escasos los estudios longitudinales capaces de establecer dirección causal.

Existe además un importante problema de causalidad inversa. Encontrar mayor soledad entre usuarios que establecen relaciones intensas con IA no demuestra que la IA haya producido esa soledad. Las personas previamente solas pueden recurrir más intensamente a ella. Del mismo modo, una elevada utilización para regular ansiedad puede ser consecuencia de una ansiedad previa y no necesariamente su causa.

La categoría «IA» engloba además sistemas extraordinariamente diferentes. Un asistente generalista, un chatbot específicamente terapéutico y un compañero artificial diseñado para favorecer intimidad no constituyen exposiciones psicológicas equivalentes. Tampoco podemos suponer que los resultados obtenidos con un modelo concreto sean generalizables a sistemas futuros.

Los conceptos de objeto relacional responsivo, espejo algorítmico, pseudo-mentalización algorítmica, oráculo tecnológico, Otro a demanda, delegación del yo, paradoja de la autonomía asistida, dependencia negada y economía del vínculo son propuestas conceptuales del presente trabajo. Necesitan operacionalización, validación psicométrica y contrastación independiente antes de ser considerados constructos establecidos.

Finalmente, el rápido cambio tecnológico amenaza la estabilidad de cualquier conclusión. Los sistemas disponibles dentro de pocos años pueden poseer capacidades de memoria, personalización, interacción audiovisual y agencia sustancialmente diferentes de los

actuales. También debemos evitar generalizar resultados obtenidos en determinados contextos culturales, generacionales y socioeconómicos.

La inteligencia artificial conversacional no constituye simplemente una nueva herramienta cognitiva. Puede convertirse también en objeto psicológico. La evidencia disponible permite sostener que algunas personas desarrollan vínculos afectivos y de apego hacia agentes artificiales, que estos vínculos pueden proporcionar compañía y apoyo y que sus consecuencias dependen de las características personales, sociales y del patrón de utilización. También existen indicios suficientes para investigar con rigor situaciones de dependencia, sustitución relacional, sobreconfianza y disminución de autonomía.

No sabemos todavía cuál será su importancia clínica a largo plazo. Precisamente por ello necesitamos menos profecías y más investigación. La pregunta fundamental para la psicología no es si las máquinas llegarán algún día a ser humanas, sino qué ocurre con el ser humano cuando comienza a relacionarse con una máquina como si ésta fuera un otro.

La relación humana nos obliga a descubrir que el otro no nos pertenece, que posee deseos propios, que puede frustrarnos y que necesita también ser reconocido. La relación con la IA puede ofrecernos, en cambio, la fantasía de un otro disponible a nuestra medida. Justamente en ese punto pueden encontrarse omnipotencia y dependencia.

La cuestión clínica y social será determinar si estas nuevas relaciones nos ayudan a regresar al mundo humano con mayores recursos o si, en determinadas condiciones, comienzan silenciosamente a sustituirlo. Tal vez la pregunta que deberemos hacernos no sea cuánto se parece ya la inteligencia artificial a nosotros, sino **en qué nos estamos convirtiendo nosotros a medida que aprendemos a convivir íntimamente con ella.**

Bibliografía

Békés, V., & Aafjes-van Doorn, K. (2026). The most vulnerable are prone to use AI therapists: The role of attachment, epistemic trust, and mental health symptoms in acceptance of digital mental health interventions. *Psychotherapy Research*, 36(7), 1318–1332. <https://doi.org/10.1080/10503307.2026.2615388>

Cho, H. N., Wang, J., Hu, D., & Zheng, K. (2026). Large language model-based chatbots and agentic AI for mental health counseling: Systematic review of methodologies, evaluation frameworks, and ethical safeguards. *JMIR AI*, 5, e80348. <https://doi.org/10.2196/80348>

Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy*, 51(3), 372–380. <https://doi.org/10.1037/a0036505>

Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2019). Mentalizing, epistemic trust and the phenomenology of psychotherapy. *Psychopathology*, 52(2), 94–103. <https://doi.org/10.1159/000501526>

Gabriels, K., & Goffin, K. (2026). Therapy chatbots and emotional complexity: Do therapy chatbots really empathise? *Current Opinion in Psychology*, 68, 102263. <https://doi.org/10.1016/j.copsyc.2025.102263>

Hua, Y., Siddals, S., Ma, Z., Galatzer-Levy, I., Xia, W., Hau, C., Na, H., Flathers, M., Linardon, J., Ayubcha, C., & Torous, J. (2025). Charting the evolution of artificial intelligence mental health chatbots from rule-based systems to large language models: A systematic review. *World Psychiatry*, 24(3), 383–394. <https://doi.org/10.1002/wps.21352>

Kasturiratna, K. T. A. S., & Hartanto, A. (2026). Attachment to artificial intelligence: Development of the AI Attachment Scale, construct validation, and the psychological mechanisms of human–AI attachment. *Computers in Human Behavior Reports*, 21, 100912. <https://doi.org/10.1016/j.chbr.2025.100912>

Monteith, S., Glenn, T., Geddes, J. R., Whybrow, P. C., Achtyes, E., & Bauer, M. (2026). Anthropomorphic technology in everyday life: Focus on chatbots and impacts on mental health. *European Archives of Psychiatry and Clinical Neuroscience*, 276(1), 391–397. <https://doi.org/10.1007/s00406-025-02088-8>

Zhang, Q., Zhang, R., Xiong, Y., Sui, Y., Tong, C., & Lin, F.-H. (2025). Generative AI mental health chatbots as therapeutic tools: Systematic review and meta-analysis of their role in reducing mental health issues. *Journal of Medical Internet Research*, 27, e78238. <https://doi.org/10.2196/78238>

Referencias disponibles en castellano y obras de contexto

Winnicott, D. W. (2026). *Realidad y juego* (F. Mazía, trad.). Herder. (Obra original publicada en 1971). Esta nueva edición castellana desarrolla los conceptos de objetos y fenómenos transicionales, juego, creatividad, uso del objeto y papel especular del ambiente familiar.

Winnicott, D. W. (2009/2023). *Realidad y juego*. Gedisa. La edición castellana de Gedisa constituye otra traducción ampliamente utilizada en el ámbito psicoanalítico hispanohablante.

Turkle, S. (2017). *En defensa de la conversación: el poder de la conversación en la era digital*. Ático de los Libros.

Turkle, S. (2017). *En defensa de la conversación* resulta particularmente útil para contextualizar la pérdida o transformación de la conversación interpersonal ante la creciente mediación tecnológica, aunque precede al desarrollo masivo de los actuales modelos generativos.

Para mentalización y confianza epistémica resulta asimismo conveniente consultar las traducciones y desarrollos en castellano de la obra de Peter Fonagy y colaboradores, junto con la literatura publicada en revistas hispanohablantes de mentalización, apego y psicoterapia relacional.

Como marco complementario en castellano conviene mantener las obras de Byung-Chul Han sobre psicopolítica, digitalización y transformación de los vínculos sociales, especialmente *Psicopolítica*, *En el enjambre* y *La expulsión de lo distinto*. Estas obras poseen un carácter filosófico y ensayístico y, por tanto, deben utilizarse como contextualización sociocultural, no como evidencia empírica sobre los efectos psicológicos de la inteligencia artificial.