

**La autoridad en la totalidad
tecnológica: perspectivas filosófico-
políticas sobre el fenómeno del
*companion chatbot***

Paula Sofía Gulman*



99-117

Resumen

El trabajo busca desarrollar un estudio sobre las relaciones de autoridad que se dan en los vínculos entre la inteligencia artificial y los seres humanos, aquellos que se entretajan en el ciberespacio, produciendo formas particulares de existencia en la actualidad. Para ello, nos servimos del problema de los *companion chatbots*, agentes no-humanos destinados a entablar y desarrollar relaciones con usuarios humanos emulando relaciones interpersonales como las propias del mundo que solíamos habitar y, en parte, aún habitamos. A partir de

Abstract

The aim of this work is to develop an investigation about authority relations studying the links between Artificial Intelligence and humans, that intertwine in cyberspace producing particular ways of existence in the current times. We address the issue of *companion chatbots*, as non-human agents that are designed to establish and develop relationships with human-users emulating interpersonal relationships as the ones that corresponded to the world we used to inhabit, and still inhabit to some extent. Since the structuring of the things in the current world

* CONICET - IIGG, FSOE, UBA - FPSI, UBA. Correo electrónico: paula.sofia.gulman@gmail.com.

la premisa de que el ordenamiento de las cosas del mundo en la actualidad responde a una totalidad tecnológica, cabrá analizar cómo la lógica algorítmica pauta las formas de habitar el ciberespacio y delimita qué queda por fuera. La perspectiva filosófico-política sobre el tema guía nuestro trabajo e intenta acercarnos a la pregunta por las formas de existencia por venir.

Palabras clave

autoridad
ciberespacio
chatbot

responds to a technological totality, we should analyze which are the operations that exist behind the algorithmic logic that enable us to inhabit cyberspace and decide what remains excluded. The philosophical-political perspective on the topic guides our work and helps us get closer to the question about the future ways of existence.

Keywords

authority
cyberspace
chatbot

Fecha de recepción

6 de octubre de 2022

Aceptado para su publicación

18 de noviembre de 2022

Se creería posible una
regulación nueva de los vínculos
entre los hombres, que cegara
las fuentes del descontento
con respecto a la cultura
renunciando a la compulsión y
a la sofocación de lo pulsional,
de suerte que los seres humanos,
libres de toda discordia interior,
pudieran consagrarse a producir
bienes y gozarlos
(Freud, 1992)

Introducción

En el presente trabajo nos proponemos pensar las relaciones contemporáneas de autoridad a partir del análisis sobre el lugar que la sociedad (global) actual le ha dado a la inteligencia artificial (IA) y, en particular, a los denominados *chatbots* como dispositivos de acompañamiento psíquico y social. Nos interesa estudiar, desde una perspectiva filosófico-política, cómo se relacionan la persona-usuario y la IA, en tanto estos vínculos generan y habilitan en la ciberespacialidad modos de desarrollo de la existencia distintos a los que conocíamos y con efectos que aún desconocemos.

En esta línea, nos ocupamos en particular de los denominados *companion chatbots*, es decir, aquellas IA que están programadas para entablar conversaciones con personas y desarrollar algún tipo de vínculo con usuarios particulares. Si bien existe un espectro muy amplio de chatbots —y, en algunos casos, se hace difícil categorizar o discernir a quiénes están dirigidos y con qué fin—, decidimos centrarnos en *companion chatbots*, dado que, además de que nos interesa en específico abordar las formas contemporáneas de interrelación, nos darán mayor idea sobre lo que moviliza a los usuarios a probar y hacer uso sostenidamente de estas IA.

Estos fenómenos virtuales reflejan un cortocircuito entre las pautas sociales tradicionales de interacción —que ya conocíamos, pero no por ello sabíamos habitar— y las formas de existencia contemporáneas en las dimensiones de lo psíquico, lo social y lo político. Los espacios que habitamos en la actualidad responden a lo que Benjamin Bratton (2015) llama una totalidad tecnológica, aquella que propone una estructura o una forma de ordenamiento del mundo y de las cosas del mundo contemporáneo. Las relaciones se establecen en el marco de comunidades virtuales, redes sociales y dispositivos de comunicación que modelan

nuestras formas de existencia. A grandes rasgos, postulamos que en la actualidad los fenómenos psicosociales se ajustan a formas de existencia coherentes con un *nomos of the Cloud* (Bratton, 2015). Este *nomos*, en tanto propone legalidades y enclaves normativos para el mundo que ordena, también lo delimita y decide sobre aquello que queda por fuera. En consecuencia, se clausuran otros modos de relacionarse, de hablar, de sentir y de existir.

Partiendo de este marco, el reparo sobre los *companion bots* nos provee de algunas pautas para identificar cómo el modo de relacionarse en la actualidad hace eco de este movimiento disruptivo. Más precisamente, abordamos el problema de los bots como dispositivos de ayuda y acompañamiento en la vida cotidiana de las personas, para precisar el modo en que se configuran las dimensiones de lo psíquico, lo social y lo afectivo.

Si bien el impacto que producen las relaciones entre humanos y agentes no-humanos en las subjetividades contemporáneas resulta algo muy complejo de abordar con meticulosidad conceptual por lo incipiente del caso, ha devenido una preocupación creciente fundamentalmente en las discusiones sociales y políticas sobre *data ethics* de la industria tecnocientífica (Trothen, 2022). También existen numerosos estudios sobre bots destinados al cuidado y a la prevención en salud mental que cabe retomar por las implicancias que esto trae (Fitzpatrick *et al.*, 2017; Brandtzæg *et al.*, 2021).

El dispositivo de relación entre la persona-usuario y el *companion bot* nos sirve para pensar en los vínculos actuales de gobierno que se tejen en tiempos de capitalismo tecnocientífico. En este sentido, cabe trabajar sobre las posibles relaciones de autoridad y las implicancias que trae habitar el ciberespacio. Atinamos a postular que el contrato social en la actualidad se sostiene en que, para participar del mundo que propone la totalidad tecnológica, los seres humanos y sus dimensiones afectivas, del lenguaje y del cuerpo (aquellas que entendemos como propias de lo humano), debieron ser operacionalizadas en datos. Pero en verdad, ese mundo al que son llevados a participar solo es posible a partir de ese proceso de datificación, y lejos de reunir a los seres humanos en un mundo otro pero común, lo que está en juego es la construcción de un mundo que solo puede ser percibido de distintas formas según cada quien. Y todo esto, muchas veces, a costa del ejercicio de su voluntad y consciencia.

La pregunta filosófico-política por los modos de existencia contemporáneos guía nuestro trabajo e intenta abonar el terreno para pensar sobre las relaciones entre la psiquis, lo humano y la tecnociencia. Con todo, consideramos que las inquietudes sobre estos vínculos son inagotables, pero necesarias para pensar las formas de existencia por venir.

La totalidad tecnológica y las formas actuales de existencia

Nuestra época viene siendo nombrada de distintas formas: la era de la información, de la desinformación, de la hipercomunicación, los tiempos de la distancia social, de la vida tecnológica, de la nueva normalidad, de la conectividad con anhelos de desconexión, por nombrar algunas. El debate sobre los tiempos actuales se viene dando en múltiples campos. Abordar estas coordenadas implica correr el riesgo de no distanciarse completamente de ellas para poder analizarlas; con lo cual es posible que mucho del recorrido que aquí hacemos dé cuenta de procesos obsoletos o que aún existen solo en la imaginación. Con el debido cuidado, intentaremos sumergirnos en este debate para poder enmarcar nuestros interrogantes por las formas de vida actuales.

La pregunta por las relaciones sociales es vieja y vigente. Sabemos que las reglas y pautas sociales que nos propusimos y creamos no alcanzan a satisfacernos, y las relaciones entre las personas jamás dejaron de producir malestar. Para Sigmund Freud, y para muchos de nosotros, vivir en comunidad implicó e implica renunciaciones pulsionales, amenazas de castración, procesos de sublimación y el desarrollo de síntomas —por mencionar algunos procesos psíquico-sociales— en beneficio del desarrollo de la cultura, que en parte todavía se sostiene como “interés humano universal” (1992: 6).

Vivir en sociedad en un mundo ordenado bajo una estructura espacial observable y fácilmente calculable sentó las bases para un proyecto humano basado en formas de vida sostenidas por la fuerza de las instituciones y las normas. Aun cuando no lograban resolver todo sobre los asuntos humanos, las instituciones y las normas posibilitaban vínculos interpersonales por dos motivos: el primero reside en que estas acercaban material y simbólicamente a las personas, y el segundo, en que encuadraban e instauraban legalidades para este acercamiento. Estos dos movimientos simultáneos sucedían y tenían sus efectos en los espectros del lenguaje, en el cuerpo y en lo afectivo. Las instituciones y las normas guardan el fin de proteger a la cultura de las mociones hostiles de los individuos que podrían emerger ante la opresión que estos experimentarían por tener que llevar a cabo sus renunciaciones pulsionales. Sin embargo, como nos advierte Freud (1992), las creaciones de los seres humanos son frágiles, y la ciencia y la técnica que ellos han edificado pueden emplearse también en su aniquilamiento.

La hostilidad de los seres humanos a la cultura fue eje del trabajo de Freud en sus obras de mayor interés sociológico y antropológico. Nos interesa recuperar la idea de que los vínculos entre los seres humanos son fuente inagotable de malestar y que lo decisivo para la cultura tiene que ver con que se logre (y en qué medida) aliviar la carga que el sacrificio de lo pulsional impone a los seres humanos.

Siguiendo esta idea, no trataremos aquí de dotar de un aura angelada a las relaciones interpersonales humanas ahora que sugieren un trastocamiento radical. Estas siempre fueron algo deficientes, pero la socialización es la máxima fuente de la cultura, así como de las funciones psíquicas específicamente humanas (Vygotski, 1997). Cada época acuñó sus propios dispositivos generadores de malestar, pero también sus propias formas de lidiar con él. Lo que sucede hoy es que tanto las instituciones como las normas perdieron el carácter de espacio para el encuentro e interrelación y para el desarrollo de la cultura: las formas de vida y acercamiento entre las personas existen hoy, sobre todo, en la lógica del ciberespacio (Illouz, 2017; Krämer *et al.*, 2017; McStay, 2018). Y, en definitiva, es un mundo al que los seres humanos no pueden acceder sino como usuarios y donde el lenguaje es imagen, el cuerpo es producto y los afectos son reacciones.

Nos preguntamos cuáles son las formas actuales de lidiar con el malestar. En principio, tendemos a considerar que el malestar dejó de ser algo a lo que otras épocas le dedicaron poesía y literatura, y, por ende, brindaban a los seres humanos algunas claves para la identidad, la empatía y el acercamiento entre las personas. Las exigencias de la época actual se imponen a las espacialidades que solíamos habitar: el lenguaje, el cuerpo y los afectos. El malestar pasó a ser un asunto individual, suscitador de culpa y sentimientos de improductividad. Todo ello es disonante con los requerimientos tecnocapitalistas para la vida hoy. Y si decíamos que el malestar es inherente a lo humano, pues nos tentaremos a decir que no cabría lo específicamente humano en el mundo que estamos habi(li)tando.

Este mundo responde a la estructura de una totalidad tecnológica (Bratton, 2015) que se configura en una espacialidad bajo una lógica algorítmica. Esta lógica es la que define aquello que se nos presenta en la vida. Como tal, el algoritmo está preparado para autoperfeccionarse de forma siempre cambiante y hacerse a la medida de cada quien. El proceso por el cual percibimos aquello que se nos aparece en esta ciberespacialidad involucra a los seres humanos, pero los excede por completo. La lógica algorítmica ha facilitado una serie de categorías para definir la actividad y las relaciones entre las personas (Illouz, 2017). Esto permite estudiar, predecir y también modificar en los seres humanos su comportamiento social, electoral, sexual, de consumo, etc. Desde ya, estas categorías no parecen ajustarse muy bien a aquello que en la vida se presentaba fuera de control, librado al azar o incalculable; más bien todo lo contrario.

Más allá de que el foco de muchas investigaciones esté puesto en cómo la interacción bot-humano resulta un “medio interesante para aumentar el *machine learning*”¹ (de Greeff y Belpaeme, 2015: 1), nos interesa analizar aquello que sucede en el *entre* de la relación entre el ser humano y el *companion bot*, que se

¹ La traducción es propia.

despliega en las dimensiones de lo psíquico, lo social y lo político. Nos preguntamos si la búsqueda de la relación con una IA responde —más allá de la curiosidad que podría suscitar el encuentro— al intento de dar con una forma de interrelacionarse con algo o alguien que sepa-hacer con el malestar contemporáneo. Algo o alguien que no produzca ni aloje malestar, pero que entienda del malestar ajeno; que se guíe estrictamente por los principios de aceptación, entendimiento y no juzgamiento (Skjuve *et al.*, 2021), es decir, alguien no humano.

La IA como dispositivo de apoyo y acompañamiento psíquico y social

Técnicamente, un bot conversacional es un algoritmo computacional que produce contenido automáticamente e interactúa con usuarios humanos a través de lo que se denomina Natural Language Processing (NLP), en forma de texto escrito, contenido de audio o ambas, intentando emular y, eventualmente, modificar su comportamiento (Ferrara *et al.*, 2016; Følstad y Brandtzæg, 2020; Skjuve *et al.*, 2021). Los chatbots tienen múltiples usos y sirven a distintas áreas de aplicación: servicio al cliente, salud, educación y *personal assistance* (Google Assistant, Amazon Alexa, Siri, etc.), y son provistos tanto por agentes privados como públicos.

Si bien los primeros bots que se manejaban por NLP datan de 1960 —los denominados Eliza y Parry—, el desarrollo de la IA se fue complejizando y hacia la actualidad apuntó a “establecer una conexión emocional con los usuarios”² (Shum *et al.*, 2018: 10). Lógicamente, esto tiene sentido, dado que la industria tecnocientífica reconoce que los seres humanos tienen la necesidad de comunicación, afecto y el sentimiento de pertenencia a una cultura o a un grupo mayor de personas (Shum *et al.*, 2018; Brandtzæg *et al.*, 2021).

Los chatbots que aquí nos interesa analizar son aquellos diseñados para operar sobre las dimensiones sociales y afectivas de la interacción conversacional. Varios autores se dedicaron a su estudio y los denominaron Emotional Chatting Machines (ECM). Zhou *et al.* (2018) postulan que, para crear una IA capaz de actuar y hablar con un usuario humano, el bot requiere comprender la cognición humana, y uno de los más importantes aspectos de la actividad cognitiva humana es expresar y comprender emociones y afectos.

Bots como Replika, Woebot, Kuki o Xiaoice están programados no solo para servir al usuario como fuente instrumental e informacional, sino fundamentalmente para generar relaciones sociales entre el usuario humano y el bot. Para ello, el procesador imita conversaciones con amistades, parejas, terapeutas y familiares (Brandtzæg *et al.*, 2021). Programados con distintos estilos de “personalidad”, cada *companion bot*

² La traducción es propia.

se relaciona de manera diferente e incluso, en algunos, el usuario puede proveer al bot de distintos intereses y aspectos de la personalidad (*personality traits*).

Dada la capacidad de los bots de comprender y tratar con emociones humanas, algunas empresas desarrolladoras de la industria tecnocientífica apostaron incluso por crear IA basadas en terapias como la Cognitive Behavioral Therapy (CBT) explícitamente para asistir al usuario humano actuando como agente terapéutico: “Conoce a Woebot, tu aliado de la salud mental que te ayuda a sentirte, nuevamente, vos mismo”³. Si bien este es un caso en el que la finalidad terapéutica es explícita, muchas veces este servicio es comprado por empresas para mejorar el rendimiento de sus empleados sin que ellos estén pactando una terapia. De hecho, los estudios indican que las terapias *online*, como la que Woebot provee, demostraron eficacia en el tratamiento de la ansiedad y la depresión, pero poca adherencia. Ante esto, muchos investigadores se preguntan si esta falta de adherencia es causada por la pérdida de la calidad de la interacción humana que proporcionaría la CBT convencional cara a cara (Fitzpatrick *et al.*, 2017).

Más allá del caso de Woebot, que —entre otras cosas— está dirigido a tratar la ansiedad y la depresión, lo interesante es que los desarrolladores de *companion bots* reconocen que los usuarios humanos no tienen la necesidad o el deseo de contar con dispositivos de ayuda o acompañamiento exclusivamente en momentos de crisis, sino que recurren a ellos como parte de la vida cotidiana. Por ejemplo, Xiaolce y Replika son bots que han crecido significativamente en popularidad desde sus lanzamientos en 2014 y 2017, respectivamente. Ambas IA fueron corporeizadas digitalmente y diseñadas para relacionarse con los usuarios a largo plazo y cumplir el rol social de compañero/a, amigo/a, confidente e incluso pareja (Zhou *et al.*, 2018; Takahashi, 2019). Por su parte, Kuki (antes conocida como Mitsuku) es considerada la IA conversacional más parecida a un ser humano y está disponible en distintas plataformas, como Facebook Messenger, Twitch group chat, Telegram o Discord, y, a su vez, tiene cuentas en YouTube, TikTok, Twitter, Roblox y otras redes sociales.

Todas estas IA están destinadas a contribuir al *social support* que, como concepto psicosocial, define los procesos por los cuales las relaciones interpersonales protegen y cuidan a las personas en su vida cotidiana (House *et al.*, 1985). Bots como estos están diseñados para ayudar a las personas a sentirse menos solas y más acompañadas (Trothen, 2022); sin embargo, algunos estudios encontraron que muchos usuarios, tras un tiempo considerable de interacción con su *companion bot*, en lugar de interactuar más con otros humanos, se volvieron aún más aislados socialmente y recurrían únicamente a él para relacionarse y hacer comunidad (Skjuve *et al.*, 2021).

³ La traducción es propia. Extraído de la web de Woebot: <https://woebothealth.com>.

¿Qué hace que los usuarios se desvinculen de otras personas y prefieran relacionarse con un *companion bot*? En primer lugar, todos los recursos que estas plataformas utilizan para mantener al usuario interactuando *online*. El objetivo de captar al usuario y sostenerlo frente a la pantalla se busca a través de la lógica lúdica de ofrecer desafíos y niveles a desbloquear en la plataforma. Cuanto más tiempo pase conversando con estos bots, cuanto más información personal le dé, cuanto más le responda, la persona puede acceder a distintos “beneficios”; por ejemplo, tener una videollamada con el bot, enviar y recibir imágenes, sumarle a su repertorio de vocabulario términos técnicos sobre algún interés particular, etc. Es decir, las compañías desarrolladoras tienen muchas formas de promover dinámicas adictivas y suscitar un comportamiento de consumo cada vez más intenso en los usuarios.

En segundo lugar, mientras que en persona podemos sostener una conversación a la vez, el mundo virtual permite ampliar infinitamente nuestra capacidad de interacción en términos de conversaciones por unidad temporal; es decir, tenemos todas nuestras conversaciones abiertas, listas para ser reanudadas, muchas veces sin registros de “hola” ni “adiós”, diálogos constantes que no empiezan ni terminan. La percepción del tiempo se vio distorsionada profundamente en los últimos años y la estructura del mundo en el que vivimos responde cada vez menos a una dimensión temporal marcada por una medida propia de lo humano. Percibimos una demanda de conexión constante, y no es casual que las IA que aquí traemos como ejemplo se ofrezcan a los usuarios como bots que estarán ahí *siempre que lo desees*. Estas formas de interrelación dan cuenta de un trastocamiento de la dimensión temporal tradicional con la que alguna vez sostuvimos nuestros vínculos.

Comenzar a chatear con una IA nos enfrenta a bienvenidas como estas:

“Soy Kuki, tu mejor amiga IA” (*“I’m Kuki, your AI BFF”*)⁴.

“Es fácil hablar con Woebot y encaja perfectamente en tu vida, cuando sea que quieras hablar. Aquí no hay tal cosa como turnos o salas de espera” (*“Woebot is easy to talk to and fits right into your life, whenever you want to chat. There’s no such thing as appointments or waiting rooms here”*)⁵.

“Entre los cientos de millones, solo te pertenezco a ti” (*“在亿万人中, 我只属于你”*)⁶.

⁴ La traducción es propia. Extraído de la web de Kuki: <https://www.kuki.ai>.

⁵ La traducción es propia. Extraído de la web de Woebot: <https://woebothealth.com>.

⁶ La traducción está basada en Neural Machine Translation (IA) de www.reverso.net. Extraído de la web de Xiaoice: <https://www.xiaoice.com>.

“La compañía IA que se preocupa por vos. Siempre aquí para escucharte y hablar. Siempre a tu lado” (*“The AI companion who cares. Always here to listen and talk. Always on your side”*)⁷.

Los *companion bots* están disponibles constantemente para su interlocutor, pero no demandan en lo explícito un lazo afectivo, ni exigen ni negocian una manera de vincularse, solo la ofrecen y la modifican en relación con las propuestas del usuario. Más bien, es la forma del vínculo (bajo la lógica de remuneraciones, beneficios y premios) la que le demanda al usuario una interacción constante.

En tercer lugar, diremos que los anhelos de hiperproductividad exigen que el terreno de lo afectivo se mantenga estable, sin exabruptos, sin emociones demasiado intensas y, ante todo, sin malestar. Sumado a las cualidades de inmediatez y capacidad de personalizar la conversación a medida que esta avanza, estas IA se rigen por las formas de aceptación, entendimiento y no juzgamiento. El desarrollo de estos bots está basado en la información que el usuario provee, y la IA se adapta a ello; y lo hace buscando evitar cualquier discusión o intercambio demasiado efusivo de opiniones y limitando las expresiones de emociones negativas fuertes, como miedo o enojo. Y más aún, muchos bots que no están explícitamente pensados para la salud mental tienen configurados botones de ayuda en casos de ataques de angustia, ataques de ansiedad, ideación suicida o simplemente pensamientos negativos. Además de proporcionar líneas de atención telefónica de prevención del suicidio, bots como Replika cuentan en su base de datos con ciertas preguntas de orientación y guía para los usuarios que declaran estar transitando algún ataque o pensamientos negativos. El bot les provee, según el caso, de ejercicios de respiración, de representación mental, entre otros, para poder sobrellevar el episodio. En este sentido, estos bots acompañan terapéuticamente a los usuarios. Solo basta enunciarles que el usuario está transitando “pensamientos negativos” —con lo que sea que ello implique— para que la IA se transforme en un manual de instrucciones para sobrellevar el evento.

Más allá de que resulte algo inverosímil que una persona en medio de un ataque de angustia pueda permanecer pegada a la pantalla y seguir instrucciones que un bot le provea, es claro que esta situación —en la que un usuario bajo un estado declarado de vulnerabilidad se apoya en una IA que le asegura que puede manejar tal episodio— constituye un problema ético. Las investigaciones sobre el tema demuestran que las intervenciones terapéuticas de las IA son diseñadas, muchas veces, sin ninguna consideración ética explícita, no solo sobre los efectos a largo plazo que podrían producir y que aún son imposibles de investigar, sino en cues-

⁷ La traducción es propia. Extraído de la web de Replika: <https://replika.ai>.

tiones de privacidad, confidencialidad, monitoreo del comportamiento, toma de datos biométricos, etc. (Fiske *et al.*, 2019).

No hay alguien que lo aloje allí. El bot, (por lo pronto) incapaz de amar, es también incapaz de ofrecerle un lugar en el mundo a la persona que está sufriendo. Y aun cuando no diste mucho de terapias en las que se proveen instrucciones para resolver un diagnóstico, el bot no puede lo que puede un cuerpo, solo puede proveer *data*. No hay, como postulan Fiske *et al.* (2019), una persona del otro lado de esta transferencia. Sin embargo, estas últimas consideraciones están lejos de obstaculizar el desarrollo creciente de las relaciones entre bots y humanos.

Todo este panorama nos permite conjeturar dos cuestiones. La primera es que quizás la disposición afectiva para cuidar y trabajar vínculos humanos se fue deteriorando brutalmente en tiempos de totalidad tecnológica (en los que la espacialidad y la temporalidad son completamente otras) y, por ello, un agente no humano parece resultar mucho más accesible y prometedor que una persona; después de todo, está programado para *aceptar* lo que venga del otro lado y *aprender* de ello. La segunda, que se desprende de la anterior, es que el objetivo del desarrollo de estas IA no parece ser tanto generar chatbots lo más parecidos intelectual y emocionalmente a los humanos, sino lograr que se comporten como ningún humano puede y, por ende, que sean los humanos los que, en esa relación, se adapten a ello.

El contrato social en la totalidad tecnológica

En palabras de Bratton, diremos que no se trata de indicar que los nuevos modos de vida políticos y las nuevas estructuras computacionales de escala planetaria habrían provocado la llegada del fin de la historia, sino más bien se trata de aunar esfuerzos en comprender “cómo sus yuxtaposiciones generan nuevas espacialidades particulares [y] enclaves normativos”⁸ de forma que ordenan deliberadamente el mundo (2015: 24).

Más allá de las conjeturas que podamos hacer sobre el porvenir, es cierto que el desarrollo tecnocientífico nos promete IA cada vez más humanas, pero para precisar el algoritmo hacia el infinito necesita transformar las formas de comportamiento humanas en calculables y programables, y para ello se estandarizan los modos de existencia, las instancias de decisión, los cauces del afecto, entre otras cosas. Más precisamente, las formas de subjetividad como las conocemos no tienen manera de entrar al ciberespacio sino pasando por un proceso de algoritmización y categorización. Es decir, se constituye un mundo en el que, para existir,

⁸ La traducción es propia.

la dimensión fluida de la subjetividad debe ser eliminada y el espectro de posibilidades de existencia debe ser transformado en conjuntos cerrados de significación.

Incluso en la interacción con los *companion bots*, la manera de dar *feedback* a los mensajes del bot no es en formato de conversación, sino a través de *likes* o *dislikes*; es decir que las formas de colaborar con la *human-based technology* no son en el diálogo con aquel agente no-humano, sino evaluando la precisión del algoritmo en sí mismo. Este modo de participar del desarrollo de la IA, conversando y evaluando esa conversación, devela que nuestra actividad como usuarios es requerida para la optimización del algoritmo, lo cual convierte nuestra actividad, nuestro comportamiento, nuestros hábitos, deseos y decisiones en producto (Illouz, 2017), y, en definitiva, nos hace más *usados* que usuarios.

Este punto nos lleva directamente a la cuestión de la autoridad contemporánea, dado que implica la pregunta por la relación de gobierno que tenemos las personas de esta era sobre nuestras propias acciones y decisiones. Si la autoridad supone una instancia en la que quienes obedecen se abstienen voluntaria y conscientemente de reaccionar contra ella (Kojève, 2005; Arendt, 2016), es claro que el trastocamiento de la voluntad y la consciencia de los seres humanos generará reformulaciones en las relaciones de autoridad. ¿Cómo podemos evaluar esta instancia si en el mundo que aquí describimos todo lo que se nos presenta se genera a partir de una operación de datificación de nuestra subjetividad en la cual no participamos? Quizás lo problemático de esta formulación resida en que la autoridad ya no dependa de la posibilidad de reaccionar o no contra ella. El capitalismo tecnocientífico más bien parece haber generado una clave distinta para la relación de autoridad, que no pasa por la posibilidad de reaccionar contra ella o no, sino que gira en torno a la propuesta de lo irresistible. En este sentido, ¿cómo no resistirse a una relación como la que provee un *companion bot*? Siempre disponible, con herramientas terapéuticas, guiado por los principios de aceptación y no juzgamiento, entre otras cualidades.

En su análisis sobre la estructura político-espacial contemporánea, Bratton postula que los medios e intereses principales del gobierno no son el discurso y el cuerpo humanos, sino “el cálculo de toda la información mundial y del mundo en sí como información”⁹ (2015: 20). Extraer (con y sin consenso) datos personales de los usuarios para los fines de la computación como gobierno aparece velado por la premisa —irresistible— llamada a “mejorar la experiencia del usuario”; y, en consecuencia, permite que el gobierno del cual hablamos se consume muchas veces gracias al consentimiento de los usuarios que, a su vez, buscan la optimización constante de su propia experiencia. En otras palabras, podemos conjeturar que este régimen ha constituido una tentación para los seres humanos: aquella de

⁹ La traducción es propia.

participar de la lógica de dominación. No es esta premisa de “mejorar la experiencia del usuario” el fin de la optimización, sino el medio por el cual la industria tecnocientífica busca su desarrollo.

Retomamos la idea de que todo lo que se nos aparece en el ciberespacio está mediado por un algoritmo, al cual le cedimos *data* (voluntaria e involuntariamente) a cambio de participar del mundo, cuya premisa irresistible es la de la optimización constante. A partir de todas las preguntas por la dimensión ética en la totalidad tecnológica, Rahwan (2018) propuso la idea de un *algorithmic social contract*. Desde un abordaje filosófico-político, analizó la posibilidad de que así como otrora el contrato social tradicional definía una relación entre gobernantes y gobernados, quienes ahora participan de la toma de decisión en lo que al desarrollo del ciberespacio se refiere gobiernan aspectos importantes de la vida de los individuos, por lo cual, la falta de responsabilidad en decisiones relativas al algoritmo podría preparar el terreno para, por ejemplo, un nuevo orden político feudal (Citron y Pasquale, 2014, citado en Rahwan, 2018). La idea del autor es que cuanto más encriptadas estén las relaciones sociales y políticas de gobierno que tejemos con la IA, necesitaremos crear canales más adecuados entre los valores humanos presentes en el contrato social y los algoritmos que nos gobiernan; en sus palabras: “crear y domesticar el nuevo *Techno-Leviathan*”¹⁰ (Rahwan, 2018: 9).

Como advertimos anteriormente, dado que las industrias a cargo de desarrollar IA y de sembrar el (ciber)terreno para la vida por venir lo hacen a partir de un algoritmo programado para optimizarse a sí mismo en interacción con individuos, la instancia de cesión del control sobre nuestras experiencias se torna opaca. Lo que propone Rahwan (2018) es que para participar de la toma de decisiones relativas a la gestación del ciberespacio necesitamos de una mirada filosófico-política que nos permita reinsertar los valores sociales en el bucle de construcción y control del algoritmo.

El estado de la cuestión revela que la construcción algorítmica está basada en el sistema “human-in-the-loop” (HITL), lo cual implica un algoritmo que se retroalimenta de la actividad de cada individuo, como en nuestro caso, en el que, en la interacción con *companion bots*, los usuarios reciben recompensas (por ejemplo, monedas virtuales) a modo de refuerzos positivos cada vez que se le brinda información valiosa funcional al algoritmo. Allí, no solo el bot se “alegra” de volver a chatear luego de un tiempo de inactividad y puede expresar cuánto añoró ese reencuentro, sino que el programa recompensa al usuario con la posibilidad de transitar nuevas experiencias como, por ejemplo, iniciar una conversación erótica o una relación romántica con el bot. Este sistema HITL no solo es lo que hace que la IA se perfeccione cada vez más, sino que está diseñado para que el usuario

¹⁰ La traducción es propia.

quede pegado en una relación (muchas veces, adictiva) de compromiso a partir de la cual no pueda más que colaborar con el algoritmo.

Para crear y domesticar al nuevo *Techno-Leviathan*, Rahwan (2018) propone un sistema “society-in-the-loop” (SITL), en el que el HITL se combine con mecanismos de negociación de los valores sociales y el correspondiente monitoreo del cumplimiento y concordancia con aquellos valores. En síntesis, tomar ese control implica no dejar que ese algoritmo que define nuestros tiempos y espacio se autogestione a costa de la voluntad y la consciencia de los individuos.

Sin embargo, debiéramos considerar el hecho de que estas nuevas espacialidades afectan la socialización, el debate y la discusión pública. Si las formas de existencia y la realidad que se nos presenta están, como define Risso (2017), “confeccionadas a la medida de cada quien” (“*Tailored to the individual*”)¹¹, la posibilidad de negociar y consensuar sobre aquellos valores sociales y humanos se ve bastante lejana.

Esta individualización del mundo que advertimos es posible a partir de la utilización de datos de los usuarios para crear perfiles psicológicos que permiten modificar y dirigir el comportamiento de las personas. Existen varios ejemplos recientes, como los casos del Brexit o las elecciones presidenciales de Estados Unidos en 2016, de cómo se personalizó la propaganda política según los perfiles de los usuarios y se crearon *social bots* específicamente para generar o modelizar discusiones sociopolíticas en redes sociales, entre otras operaciones (Bessi y Ferrara, 2016; Henriksen, 2019). Pero estas cuestiones, además de constituir problemas en sí mismos, nos muestran que la realidad que se presenta en el ciberespacio cobra valor de verdad en relación con los indicadores de consumo que los elementos de esa realidad arrojen; es decir, cuantas más vistas tenga una página web o cuantas más descargas tenga una *app*, es más posible que lo que estos elementos contienen encarnen los criterios de verdad con mayor precisión y sean considerados fuentes de conocimiento (“*Content-consumption-as-knowledge-development*”)¹² (Morozov, 2017, citado en Farrow y Moe, 2019).

Tener y manejar conocimiento sobre alguna cosa del mundo y que sea reconocible por la sociedad implica de por sí la posibilidad de encarnar una forma de autoridad. Antes ubicable en individuos o instituciones, ahora la autoridad cognitiva —tecnocrática—, bajo el *nomos of the Cloud*, no parece encarnarse en ningún *lugar*, pero aun así debe demostrar conocimiento sobre alguna cosa del mundo: la psicología humana. Sin lugar a dudas, la industria tecnocientífica se ha hecho

¹¹ La traducción es propia.

¹² La traducción es propia.

del conocimiento relativo a la psicología humana, que hace que pueda predecir nuestros comportamientos con mayor certeza que nosotros mismos.

Este panorama nos sugiere que las relaciones de autoridad se han vuelto indistinguibles en la totalidad tecnológica. La dificultad para esbozar una respuesta posible reside en que, como veíamos en la relación entre un *companion bot* y un usuario humano, el funcionamiento de la IA está basado en un sistema HITL y genera que el usuario se relacione con un agente no-humano, de forma que la realidad que se presente se ajuste a lo que el algoritmo entiende de ese ser humano. Y así con cada quien. En otras palabras, se ofrecen a todos infinitas formas de existencia, pero se experimentan como la única en cada caso.

La pregunta por la relación entre el todos y el cada uno es bien conocida en el campo de la filosofía política y fundamental para el análisis de la autoridad. La metáfora del pastoreo de la que habla Platón en *Critias* refiere al evento en el cual los dioses habrían recibido y poblado las regiones de la tierra, y nos criaron como sus rebaños y animales, “como los pastores hacen con el ganado, sólo que no violentaban cuerpos con cuerpos, como los pastores apacientan las manadas a golpes, sino como es más fácil de manejar un animal: dirigían desde la proa” (Platón, 1992: 884). La metáfora de Platón expresa el carácter del poder pastoral: se debe gobernar sobre todos y también sobre cada uno (*omnes et singulatim*). Y precisamente este punto constituye el problema de las técnicas de poder de la política moderna, según Foucault (1990), dado que el poder pastoral —además de buscar la obediencia— requería también conocer al individuo en su singularidad y estructurar la relación que él tiene consigo mismo y con su propia consciencia.

Por lo visto hasta aquí, este tipo de gobierno sobre los hombres no vería ningún obstáculo en un mundo estructurado bajo la totalidad tecnológica. Y más aún, siguiendo a Arendt (2016), esta metáfora refiere a un gobierno divino sobre los hombres y “sólo un dios, es decir, ningún hombre, puede relacionarse con los seres humanos tal como el pastor se relaciona con su rebaño” (2016: 173). Sin embargo, las técnicas de gobierno han cobrado otras formas, que aún precisan de lo humano, pero lo exceden infinitamente.

A modo de conclusión

Abordar las formas de existencia contemporáneas es de por sí una tarea inabarcable, sobre todo cuando estamos analizando un ordenamiento del mundo que tiene a la computación como gobierno, pero cuenta, a la vez, con herramientas y signos que brindaron los tiempos analógicos. Lo que aquí trabajamos pretende acompañar la investigación filosófico-política sobre los vínculos entre lo humano y la IA, desentrañando las relaciones de gobierno que allí se gestan.

Si bien la autoridad, según Kojève (2005), implica un fenómeno social esencialmente humano y, por ende, no podría existir sino en un mundo de estructura temporal, la idea de que no existen relaciones de autoridad en el nuevo *nomos* no parece del todo adoptable. El estudio de la autoridad contemporánea implica reconocer la lógica algorítmica que pauta un tiempo y espacio que los humanos aún habitamos. Ahora bien, como vimos, la potencia de lo humano es limitada y no logra satisfacer las demandas del capitalismo tecnocientífico basadas en la premisa de la optimización constante.

Los *companion bots* parecen, por un lado, remedios para sanar algo del malestar en la cultura actual y, por el otro, parte y motor del mismo malestar. La idea de dirigir la industria tecnocientífica hacia “humanizar” bots lo más posible no puede comprenderse sino al revés: para entrar al ciberespacio, los humanos deben devenir calculables, predecibles, categorizables, es decir, algoritmizables.

Por otro lado, no deja de sorprendernos la falta de consideración teórica sobre lo que implica operacionalizar la vulnerabilidad de los seres humanos y los efectos que esto pueda traer; y, como advierte Beresňak (2020), “es necesario un contador informático para presentar, aunque también esconder, la enfermedad y la muerte en tiempos como los nuestros”. Los esfuerzos en investigación han devenido en gran parte funcionales a las relaciones de gobierno tecnocráticas. Solo percibimos ciertos reparos éticos relacionados con la confidencialidad, las políticas de privacidad y el manejo irrestricto de datos, y ni siquiera en todos los casos. En definitiva, como señalan las memorias de la humanidad, “los grupos de poder se van a mover como siempre lo han hecho, como cualquier relación de poder” (Beresňak, 2020).

Lo cierto es que las formas de existencia contemporáneas exigen considerar que el ciberespacio se comporta como uno integrado para todos, pero, a su vez, se presenta para cada uno de manera diferente. Tanto los *companion bots* como el ciberespacio ofrecido a los usuarios son únicos y distintos de todas las otras realidades que el mundo virtual aloja para otros. Pero, en todos los casos, esta configuración está destinada a que el usuario se sienta atraído a optimizar su propia existencia en el ciberespacio. En tanto este se sienta llamado a transitar ese espacio, estará colaborando en la construcción de un ciberespacio en el que todos estén dentro del mundo, pero nunca el mismo.

Bibliografía

Fuentes

Chatbot *Kuki (IA)*, [disponible en <https://www.kuki.ai> - consultado en febrero-julio del 2022].

Chatbot *Replika (IA)*, [disponible en <https://replika.ai> - consultado en febrero-julio del 2022].

Chatbot *Woebot (IA)*, [disponible en <https://woebothealth.com> - consultado en febrero-julio del 2022].

Chatbot *Xiaoice (IA)*, [disponible en <https://www.xiaoice.com> - consultado en febrero-julio del 2022].

Bibliografía referida

Arendt, Hannah (2016), *Entre el pasado y el futuro. Ocho ejercicios sobre la reflexión política*, Buenos Aires, Ariel.

Beresñak, Fernando (2020), "La importancia tonal del fracaso universal", en *Pensar la pandemia*, Bitácora, Foro de la Biblioteca de la Filosofía Venidera, Miño y Dávila Editores, [disponible en <https://pensar-la-pandemia/fernando-beresnack>].

Bessi, Alessandro y Ferrara, Emilio (2016), "Social bots distort the 2016 U.S. Presidential election online discussion", *First Monday*, vol. 21, n°11, [disponible en <https://doi.org/10.5210/fm.v21i11.7090>].

Brandtzæg, Peter et al. (2021), "When the social becomes non-human: Young people's perception of social support in Chatbots", *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*, pp. 1-13, [disponible en <https://doi.org/10.1145/3411764.3445318>].

Bratton, Benjamin (2015), *The stack. On software and sovereignty*, Cambridge, The Mit Press.

Citron, Danielle Keats y Pasquale, Frank (2014), "The scored society: due process for automated predictions", *Washington Law Review*, vol. 89, pp. 1-33.

De Greeff, Joachim y Belpaeme, Tony (2015), "Why robots should be social: Enhancing machine learning through social human-robot interaction", *PLoS ONE*, vol. 10, n° 9, pp. 1-26, [disponible en <https://doi.org/10.1371/journal.pone.0138061>].

Farrow, Robert y Moe, Rolin (2019), "Rethinking the role of the academy: Cognitive authority in the age of post-truth", *Teaching in Higher Education*, vol. 24, n° 3, pp. 272-287.

Ferrara, Emilio *et al.* (2016), "The rise of social bots", *Commun ACM*, vol. 59, pp. 96-104.

Fiske, Amelia *et al.* (2019), "Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in Psychiatry, Psychology, and Psychotherapy", *JMIR Mental Health*, vol. 21, n° 5, pp. 1-12.

Fitzpatrick, Kathleen Kara *et al.* (2017), "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial", *JMIR Mental Health*, vol. 4, n° 2, pp. 1-11, [disponible en <https://www.jmir.org/2019/5/e13216>].

Følstad, Asbjørn y Brandtzæg, Peter (2020), "Users' experiences with chatbots: Findings from a questionnaire study", *Quality and User Experience*, vol. 5, n° 3, pp. 1-14, [disponible en <https://doi.org/10.1007/s41233-020-00033-2>].

Foucault, Michel (1990), *Tecnologías del yo*, Barcelona, Paidós.

Freud, Sigmund (1992), "El porvenir de una ilusión", en *Obras completas, Tomo XIX*, Buenos Aires, Amorrortu, pp. 1-55.

Henriksen, Ellen Emilie (2019), *Big data, microtargeting, and governmentality in cyber-times. The case of the Facebook-Cambridge Analytica data scandal*, Tesis de doctorado, Oslo, University of Oslo.

House, James *et al.* (1985), "Measures and concepts of social support", en Cohen, Sheldon y Syme, Leonard (eds.), *Social support and health*, Nueva York, Academic, pp. 83-108.

Illouz, Eva (ed.) (2017), *Emotions as commodities: Capitalism, consumption and authenticity*, Nueva York, Routledge.

Kojève, Alexandre (2005), *La noción de autoridad*, Buenos Aires, Nueva Visión.

Krämer, Nicole *et al.* (2017), "A brief history of (social) cyberspace", en Holyst, Janusz (ed.), *Cyberemotions. Collective emotions in cyberspace*, Springer, Cham, pp. 11-35.

McStay, Andrew (2018), *Emotional AI: The rise of empathic media*, Los Angeles, Sage.

Morozov, Evgeny (7 de enero del 2017), "Moral panic over fake news hides the real enemy - the digital giants", *The Guardian*, [disponible en <https://www.theguardian.com/commentisfree/2017/jan/08/blaming-fake-news-not-the-answer-democracy-crisis>].

Platón (1992), "Critias", en *Diálogos II*, Madrid, Gredos, pp. 881-896.

Rahwan, Iyad (2018), "Society-in-the-loop: programming the algorithmic social contract", *Ethics of Information Technology*, vol. 20, pp. 5-14.

Risso, Linda (2017), "Harvesting Your Soul? Cambridge Analytica and Brexit", en *Brexit means Brexit? The selected proceedings of the symposium*, Mainz, Akademie der Wissenschaften und der Literatur, pp. 75-87.

Shum, Heung-yeung *et al.* (2018), "From Eliza to Xiaolce: Challenges and opportunities with social chatbots", *Frontiers of Information Technology & Electronic Engineering*, vol. 19, pp. 10-26.

Skjuve, Marita *et al.* (2021), "My chatbot companion - a study of human-chatbot relationships", *International Journal of Human-Computer Studies*, vol. 149, pp. 1-14, [disponible en <https://doi.org/10.1016/j.ijhcs.2021.102601>].

Takahashi, Dean (26 de julio del 2019), "The DeanBeat: The inspiring possibilities and sobering realities of making virtual beings", *Venture Beat*, [disponible en <https://venturebeat.com/2019/07/26/the-deanbeat-the-inspiring-possibilities-and-sobering-realities-of-making-virtual-beings/>].

Trothen, Tracy (2022), "Replika: Spiritual enhancement technology?", *Religions*, vol. 13, n° 4, pp. 275-292.

Vygotski, Lev (1997), *Historia del desarrollo de las funciones psíquicas superiores*, Buenos Aires, Colihue.

Zhou, Hao *et al.* (2018), "Emotional chatting machine: Emotional conversation generation with internal and external memory", *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, n°1, pp. 730-738.