

Inteligencia artificial, Conversación y Humanidades. Entrevista a Nicolás Duque-Buitrago

Artificial intelligence, Conversation
and Humanities. Interview with
Nicolás Duque-Buitrago

Paula Andrea Marín Colorado

Universidad de Antioquia, Colombia

 <https://ror.org/03bp5hc83>

 <https://orcid.org/0000-0002-9930-4500>

paula.marin2@udea.edu.co

Editoras: Paula Andrea Marín Colorado
Vanessa Zuleta Quintero

SECCIÓN NO ARBITRADA

Recibido: 09/07/2025

Aprobado: 17/08/2025

Publicado: 31/01/2026

Cómo citar esta entrevista: Marín Colorado, P. A. (2026). Inteligencia artificial, Conversación y Humanidades. Entrevista a Nicolás Duque-Buitrago. *Estudios de Literatura Colombiana* 58, pp. 218-234. DOI: <https://doi.org/10.17533/udea.elc.361966>

Copyright: ©2026 *Estudios de Literatura Colombiana*. Derechos patrimoniales, Universidad de Antioquia, 2026. Este es un artículo de acceso abierto distribuido bajo los términos de la **Licencia Creative Commons Atribución – No comercial – Compartir igual 4.0 Internacional**.



Check for
updates



Nicolás Duque-Buitrago es magíster en Filosofía, estudiante del Doctorat en Sciences du Langage de la Universidad de Montpellier Paul-Valéry (Francia) y miembro de la unidad de investigación LHUMAIN de la misma institución. También, es docente de planta de tiempo completo del Departamento de Filosofía de la Universidad de Caldas. Además, es miembro del grupo de investigación Filosofía y Cultura.

El profesor Duque estuvo en la Universidad de Antioquia y en la Facultad de Comunicaciones y Filología, como parte del equipo docente que orientó el Seminario “Narración, IA y Salud” (26 al 30 de mayo de 2025), en el marco del proyecto PREFALC (Programa Regional Francia América Latina y el Caribe), en asocio entre la Universidad de Montpellier Paul-Valéry (Francia) y la Universidad de Antioquia. Aprovechamos la visita del profesor Duque para conversar con él, acerca de las relaciones entre inteligencia artificial, Humanidades y Ciencias Sociales.

Fuente: Fotografía realizada por Chantal Charnet @chantalannec



Figura 1. Fotografía realizada a partir de imágenes tomadas en el Seminario “Narración, IA y Salud” (mayo de 2025, UdeA).

Desde la investigación que has realizado, ¿cómo fue el surgimiento de la inteligencia artificial y cuál fue de su relación con el ámbito de las Humanidades?

Debo decir, en primer lugar, que mis investigaciones recientes tienen que ver con el análisis de las conversaciones naturales y su relación con los agentes conversacionales artificiales, especialmente en el ámbito de la salud mental. Esto limita mi opinión a unos campos específicos de las ciencias del lenguaje y de la IA. Sin embargo, quisiera llamar la atención sobre el hecho de que la conversación se hubiera ido convirtiendo en los últimos sesenta años en una interfaz omnipresente en diversos proyectos de IA. Esto nos permite, al tiempo, percibir rápidamente por qué las ciencias humanas y sociales están profundamente ligadas a este proyecto tecnológico.

De hecho, el campo que llamamos pragmática del lenguaje y que, entre otras cosas, ha estudiado la conversación natural en varias de sus formas no solo se consolidó durante la misma década en la que surgía la inteligencia artificial simbólica, sino que algunos de los pioneros de la IA como Robert Abelson y Douglas Carroll se veían a sí mismos como pragmáticos del lenguaje. Para dar otro ejemplo, la simulación de sistemas cognitivos que propusieron Allen Newell y Herbert Simon en los años 60 partía, entre otras cosas, de problemas prácticos del uso del lenguaje común en la investigación en psicología, por ejemplo, cuando en un experimento un individuo debía seguir unas reglas para transformar el orden de un patrón de símbolos, al tiempo que hablaba en voz alta para que los experimentadores pudieran saber lo que pensaba. Esta tarea de manipulación de patrones, y el hecho de tener que hablar en voz alta, imponía problemas prácticos que estos investigadores intentaron resolver a partir de simuladores concebidos con la premisa de que el pensamiento es un tipo de manipulación simbólica (que somos seres simbólicos como se dice en la antropología), y que las trazas de esa manipulación simbólica son una buena representación del pensamiento. Extendiendo esta hipótesis al funcionamiento de los computadores, concibieron la idea de poder reemplazar prácticas de recolección de datos como grabar a un sujeto experimental en una sala mientras hace una tarea de manipulación de símbolos con la ayuda de un experimentador que le formula algunas preguntas o indicaciones; por simulaciones simbólicas directas del pensamiento en una máquina que deja trazas de las operaciones que hace.

De hecho, y esto resulta muy importante, el gran modelo de la interacción que analizaron, observaron y aplicaron estos científicos fue la conversación corriente entre dos personas, que en la literatura científica de la época se solía denominar comunicación diádica. Tampoco era la

primera vez que en el pasado reciente se usaba el diálogo como modelo de interacción con una máquina, pues Pressey en los años 30, y luego Skinner en los años 50, habían propuesto sus máquinas de aprendizaje socráticas inspiradas, en parte, en la forma de discusión de los diálogos platónicos, especialmente del *Menón*.

Estas interacciones humano-máquina no solo buscaban que la comunicación se hiciera en lenguaje natural, sino que el tiempo de respuesta, la manera de tratar un asunto o de aclararlo, funcionara con los mecanismos de la conversación corriente. Es un momento en el que resulta central producir la sensación de estar conversando y en el que los tiempos de la interacción se calculan en términos de los tiempos de la conversación. El ingeniero Joseph Weizenbaum, quien se considera el creador del primer agente conversacional artificial, escribió sobre las condiciones psicológicas de la conversación y sobre el significado contextual en la comunicación humano-máquina, en el mismo momento histórico en el que el filósofo Paul Grice dictaba su famosa conferencia *Logic and conversation* (pronunciada en 1967 y publicada en 1975) o en el que el sociólogo Harvey Sacks (a partir de 1963) con sus colegas Emmanuel Schegloff y Gail Jefferson sentaban las bases del análisis de la conversación, en una empresa intelectual que comenzó con el análisis de llamadas telefónicas a servicios de ayuda para prevenir el suicidio.

Para decirlo de manera muy simple, la conversación pasó al primer plano en diversos ámbitos de las ciencias sociales y humanas, de la mano de la tecnología, y se fue consolidando como modelo del razonamiento práctico y de la interacción humana. Este suceso no ha hecho más que profundizarse y, por supuesto, merece toda nuestra atención.

Para no alargar mucho este primer punto, podría decir que el lenguaje natural, el lenguaje común y su lugar en la descripción del pensamiento y de la acción, está asociado con cierta vertiente de la inteligencia artificial que, precisamente, pretendía buscar lo que dichos teóricos y teóricas llamaron una definición operacional de la inteligencia y de las actividades cognitivas. Es decir, que uno de los orígenes de la inteligencia artificial tiene que ver, precisamente, con el intento de definir actividades consideradas inteligentes como resolver problemas, hacer planes, cumplir objetivos o descubrir el significado de patrones, en proyectos de investigación que supusieron que los modelos y las simulaciones ayudaban a definir nociones difíciles y ambiguas sobre el funcionamiento de la cognición humana. En este proyecto, desde sus inicios, ha habido un trabajo conjunto entre filósofos, psicólogos, psiquiatras, lingüistas, neurólogos, biólogos e ingenieros, solo por mencionar algunas profesiones al azar.

Las teorías de la conversación y la inteligencia artificial simbólica son contemporáneas, y ambas están profundamente relacionadas con los cambios tecnológicos que tuvieron lugar entre la Primera y la Segunda Guerra Mundial. Creo que uno de los aspectos que no deberíamos olvidar, aunque hablemos poco de él, tiene que ver con el impacto teórico y metodológico de la difusión y del uso de dispositivos de grabación mecánica tanto en las ciencias sociales, como en algunas ramas de la medicina. A partir de los años 40 del siglo xx, las grabaciones, las transcripciones y el análisis de la conversación humana grabada se fueron volviendo cada vez más frecuentes en la psiquiatría, la neurología, la psicoterapia y en las ciencias del lenguaje. Hoy las usamos para la comunicación científica, para hacer investigación cualitativa y cuantitativa, pero también en la vida cotidiana. De las engorrosas salas de grabación fonográfica de los laboratorios de psicología de los años 40, se fue pasando a grabadoras portátiles en los años 60, y luego a micrófonos integrados en dispositivos como pasa hoy en día, solo para hablar de la grabación de la voz. La miniaturización, la portabilidad y la integración de las máquinas que capturan y difunden información no es un detalle menor en muchos sentidos prácticos.

Luego de la Segunda Guerra Mundial, al uso de los dispositivos de captura mecánica, como las grabadoras, se sumaron los computadores que, como se sabe por el muy conocido papel de Alan Turing, se usaron para descifrar mensajes cifrados entre los países enemigos durante este período de la historia. Luego, y retomando una idea que tenía sus antecedentes en la matemática británica del siglo xix, Ada Lovelace, diversos científicos concibieron la idea de construir máquinas virtuales que resolvieran problemas no en el lenguaje de las máquinas, sino usando el lenguaje natural, el mismo que usamos para conversar y comunicarnos con las personas. Newell y Simon llamaron a estos computadores, computadores no numéricos, queriendo enfatizar con esto que los usuarios podrían comunicarse con ellos en el lenguaje corriente, el mismo que usan para conversar, y que, a partir de este tipo de comunicación, en un futuro, podrían resolver cualquier tipo de problema, desde cómo hacer demostraciones matemáticas hasta cómo preparar un pastel.

Este ensamble de tecnologías de captura, de registro y de tratamiento de información no debe subvalorarse, pues muchas de las tareas que hicieron los pioneros de la inteligencia artificial en esta área partían de transcripciones de lo que llamaban “flujos verbales”, es decir, de lo que decían en sus conversaciones personas como políticos que hablaban en la radio o pacientes psiquiátricos que eran grabados por sus médicos,

y cuyos discursos luego eran categorizados con los criterios de ciertos modelos de lenguaje e intereses teóricos, para ponerlos en programas de simulación y así modelar, a partir de usos particulares del lenguaje, patologías como la neurosis o la paranoia, o para definir el sistema de creencias de un típico político de derecha o de izquierda.

Puede decirse que allí es evidente un tratamiento semi-manual de la conversación que se irá automatizando con el tiempo hasta llegar a los modelos de reconocimiento vocal y de transcripción automática. Sin embargo, lo que resulta ciertamente revolucionario es la posibilidad técnica de capturar y luego reproducir eventos que antes eran efímeros y difíciles de percibir, pero que ahora se podían observar, transcribir, analizar y almacenar como información, para luego interpretarlos y encontrar patrones.

Sé que lo más familiar hoy en día cuando se habla de la inteligencia artificial son los chatbots de los grandes modelos de lenguaje natural como GPT, Gemini, Copilot o Deepseek, pero ellos son un punto de llegada de una larga historia de evolución tecnológica. De hecho, los primeros simuladores de IA, por lo menos hasta el año 1966, para decirlo un poco metafóricamente, no tenían un lenguaje público, no eran programas conversacionales, sino cierto tipo de programas cerrados que procesaban patrones del lenguaje a partir de reglas, para luego imprimirlos en una máquina de escribir conectada a un ordenador. Para decirlo en términos de la retórica y del teatro, eran máquinas de monólogos no de diálogos. Eran máquinas que requerían que la información fuera introducida por los investigadores y por los expertos que las usaban, luego de largos trabajos de transcripción y análisis. Así que buscar un modelo dialógico se convirtió en un reto tecnológico de enorme importancia, pues lo que se constata es que entre más conversacionales sean estas máquinas más directa se vuelve la conexión comunicativa para enviarles información, más disminuyen actividades como la transcripción, pues el lenguaje se inscribe directamente en la máquina, y más se popularizan por fuera de los usos expertos, pues se requiere de menos tecnicismos para manipularlas.

En este punto, me parece muy importante el ejemplo de la historia casi literaria de Joseph Weizenbaum. En algún momento entre los años 1964 y 1965, este ingeniero del MIT, que trabajaba en el proyecto MAC, concibió un programa para comunicar una máquina y un interlocutor humano en lenguaje natural, es decir, a partir de una conversación. Es uno de los primeros ejemplos de un programa de ordenador que no funcionaba de manera monológica, sino dialógica, con lo que esto implica desde el punto de vista de establecer una comunicación sincrónica. Según lo cuenta el mismo ingeniero, la idea de este programa se le ocurrió mientras veía una película inspirada en la novela *Pigmalión* de George Bernard Shaw.

La película cuenta la historia de Higgins, un profesor de fonética, que le enseña a una joven vendedora de flores, Eliza, la pronunciación del inglés de la clase alta londinense. La película está llena de escenas irónicas, y a la vez ridículas, en las que se usan distintas máquinas fonéticas para que Eliza “mejore” su manera de hablar. La idea nos parece hoy estereotípica, sesgada y criticable por muchas razones, pero el principio que Weizenbaum retoma es, fundamentalmente, la posibilidad de enseñarle a hablar a una máquina de cierto modo y de manera progresiva mientras conversa con un humano que la entrena, razón por la cual bautizó al programa que encarnaba esta idea ELIZA. Su programa tiene, además, un componente tecnológico importante que se desarrolló durante los años 60 y que se conoce como *time-sharing-machine*. Esta tecnología buscaba, básicamente, comunicar a un ordenador y a una persona en tiempo real, como ocurre en la conversación humana.

Sin embargo, Weizenbaum no creía en el proyecto de la IA, ni aceptaba que las máquinas pudieran pensar. De hecho, a pesar de que se dice que ELIZA es el primer chatbot de la historia, este ingeniero nunca se imaginó ese rol para su programa. Su interés se centraba en los estudios de la interacción eficaz en lenguaje natural entre un humano y una máquina, no en un modelo de simulación de la inteligencia. Para esto construyó diversos “scripts” o guiones para este programa de los que se hizo popular uno que se conoce como “Doctor”. Este script del “Doctor” simula la conversación entre un paciente y su psicoterapeuta. Podríamos pensar que se trata de una casualidad, sin embargo, este ingeniero se valió de este tipo de conversación porque el psicólogo Carl Rogers, a partir del análisis de grabaciones de lo que hacían los psicoterapeutas cuando hablan con sus pacientes, había propuesto una técnica conocida como terapia no directiva, que el ingeniero encontró muy conveniente desde el punto de vista de la pragmática de la comunicación. El “Doctor” (personaje conversacional de la máquina) hablaba muy poco (o casi nada), repetía con un algoritmo simple lo que ya había dicho la persona que le escribía y, dado el tipo de relación social que se establece entre psicoterapeuta y paciente, a veces podía hacer preguntas extrañas (fuera de contexto o ambiguas) que el interlocutor humano no cuestionaba, pues suponía que tenían algún interés. Pero acá viene lo paradójico de la historia.

Su programa lograba enganchar bien en la conversación a personas que entraban en el personaje del paciente, pero que, para disgusto del ingeniero, terminaban tratando a la máquina como a un terapeuta real, como le ocurrió a su asistente, que aun sabiendo que se trataba de una máquina y de un proyecto tecnológico, un día le pidió que saliera de su oficina porque quería hablar con “el doctor”. Quería un momento

de privacidad con él. A Weizenbaum esta idea le parecía completamente inaceptable y decidió escribir varios artículos, primero para explicar en detalle su proyecto de interacción conversacional y, segundo, para mostrar que lo que había detrás de ELIZA eran unos trucos muy, muy simples, que se lograban gracias a que el algoritmo permitía repetir frases dichas por la persona cambiando la persona y la coherencia verbal, o utilizando continuadores de tipo “dime más”, “continúa, por favor”, para que el interlocutor humano siguiera hablando y la máquina participara muy poco. Su objetivo era mostrar que ELIZA no pensaba, ni tenía una escucha atenta, y que esa creencia era la atribución incorrecta del interlocutor humano a un truco de magia que desconocía, pero que debería desaparecer en el momento en el que se explicara cómo funciona el algoritmo. Lo paradójico es que su artículo resultó ser tan explícito y rico en ejemplos (transcribió largos pasajes de la interacción humano-máquina explicando cómo los había hecho) que se convirtió en una pieza de aprendizaje que muchos ingenieros replicaron en distintos lenguajes de programación, que luego se escapó en los primeros computadores MAC, que se usó ampliamente en las introducciones al tratamiento del lenguaje natural, y que ayudó a instaurar la idea de un agente conversacional artificial.

Se puede decir que ELIZA representa un linaje tecnológico a partir del cual evolucionaron distintos agentes conversacionales artificiales de reconocimiento de patrones. De hecho, el psiquiatra Kenneth Colby, quien trabajaba desde principios de los 60 en la simulación monológica de la neurosis, siguió perfeccionando la heurística sugerida por Weizenbaum y propuso años más tarde a PARRY un agente conversacional artificial que simulaba la manera de hablar de un paranoico y que es un ejemplo de su proyecto de construir “pacientes artificiales” en psiquiatría.

Hay una rica bibliografía, tanto en filosofía de la mente como en ciencias del lenguaje y en psicología, que se ocupa de analizar cuáles son las estrategias —llamémoslas retóricas, interaccionales y lingüísticas— que operan detrás de estos agentes, para que los tratemos como si fueran agentes sociales. Y lo cierto es que podemos tratarlos así: los consideramos como si formaran parte de una comunidad lingüística y usamos con ellos reglas de comunicación corriente. Esto ocurre, en parte, porque utilizan el lenguaje natural y porque estamos muy inclinados a responder socialmente a todo aquello que haga uso del lenguaje natural de manera pertinente. Estos rasgos son hoy también ampliamente usados y explotados por la denominada “ingeniería de prompts”.

Los agentes conversacionales artificiales, hasta hace pocos años, seguían estando limitados a esa especie de destino microsocial de funcionar bien únicamente en contextos específicos y restringidos como los

que Weizenbaum había señalado. A pesar de que los modelos generativos han cambiado este panorama y de que esos pequeños chatbots son parte de la historia, lo cierto es que la conversación como interfaz tecnológica no solo no ha desaparecido, sino que se ha potenciado. Esta es una discusión muy activa hoy en día en diversos campos de las ciencias sociales y humanas que se ocupan de la conversación o que han estudiado las interacciones humano-máquina.

Algo que me parece curioso es que la revolución masiva que han producido estos dispositivos, hoy de extensión global, empezó con estudios medianamente marginales de formas de comunicación y de interacción cotidiana que se denominaban “microsociales”, como el estudio de las llamadas telefónicas, de la video-comunicación o de la interacción con computadores personales. Sin embargo, esos dispositivos microsociales explotaron, la interfaz microsocial se interconectó y se generalizó, desbordando las fronteras de lo grande y lo pequeño.

Desde tu experiencia como investigador, como docente de humanidades y con lo que conoces sobre el origen de la inteligencia artificial, ¿cómo crees que podríamos, desde las humanidades y las ciencias sociales, acercarnos a estas nuevas tecnologías sin miedo? ¿Cómo recibir e integrar la inteligencia artificial en nuestros procesos de investigación, docencia y educación, entendiendo que su desarrollo ha estado históricamente vinculado con preguntas propias de nuestras disciplinas?

En esa pregunta hay un elemento que me llama la atención. Desde el comienzo de la idea de los autómatas y de los seres mecánicos, es decir, dos siglos antes de la inteligencia artificial, las posturas del miedo se han combinado con la atracción y la sorpresa. A veces se habla de esta emoción como de un “valle de la inquietud”. De hecho, es una denominación que proviene de los cuentos de E. T. A. Hoffmann, y de la literatura ominosa que explora la idea de autómatas que pasan desapercibidos, como ocurre en el cuento *El hombre de arena*, pero que luego nos espantan cuando nos damos cuenta de que no son naturales. Se trata de la técnica y del arte imitando y engañando a la vida. En el siglo XVIII, el científico escocés sir David Brewster, hizo algo similar con los autómatas a lo que hizo Weizenbaum al inicio de la IA en los años 60: escribió unas cartas sobre la “magia natural” (así las llamó) para cambiar la sorpresa mística frente a los mecanismos por una atracción informada por la explicación.

Creo que esa es una pista. Mejorar la comprensión sobre los diferentes tipos y paradigmas de IA, para que no tengamos que verla como un portento. Eso no quiere decir que no haya peligros, eso quiere decir que

los peligros están mejor localizados. Hay trabajos clásicos de introducción general, y sin tecnicismos, a los diferentes tipos de IA, como *Inteligencia Artificial* de Margaret Boden o aproximaciones críticas clásicas sobre la interacción humano-máquina como *Plans and Situated Actions: The Problem of Human-Machine Communication* de Lucy Suchman. Creo, entonces, que uno de los primeros pasos es tener en cuenta los trabajos de investigación que las ciencias sociales y humanas han venido haciendo desde el siglo pasado en paralelo, en asocio o en respuesta al desarrollo de estas tecnologías. También es importante mejorar el diálogo con los ingenieros o con los investigadores en psicología, en ciencias cognitivas, biología computacional y en filosofía de la mente, que vienen trabajando desde hace años en problemáticas muy valiosas de discutir en estas áreas. A pesar de que la inteligencia artificial se popularizó recientemente con los chatbots de IA generativa, es un campo de investigación y de reflexión que tiene más de sesenta años de consolidación.

Como lo he dicho ya varias veces antes, hay discusiones importantes que debemos dar, por ejemplo, sobre el lugar de la conversación o del diálogo en la educación, y sobre la manera en que estas formas del diálogo están presentes hoy en las interfaces de inteligencia artificial. El diálogo es un elemento central de la vida humana y de nuestra historia. Si leemos los diálogos de Platón, vemos que ya albergaba dudas sobre la tecnología de la escritura como soporte tecnológico. Algunas de esas problemáticas las podríamos comparar con muchas de las que hoy enfrentamos con las tecnologías de inteligencia artificial: las trazas que deja, lo que le hace a la memoria, si resulta ser un desafío para la inteligencia, si nos va a hacer perder las capacidades para participar en el mundo político, como creía Platón que iba a pasar si los sofistas seguían vendiéndole a la gente los discursos escritos para que los memorizaran y luego los leyeran en el ágora. No estoy equiparando ambos momentos históricos, pero creo que hay una continuidad en el desarrollo de las tecnologías de la escritura, de la transcripción, del paso del lenguaje oral al lenguaje escrito, en las que la reflexión y los aportes de las ciencias sociales y humanas siguen siendo importantes.

Por otro lado, las distinciones tajantes que hacemos entre ciencias sociales y humanas e ingenierías refuerzan prejuicios negativos sobre las diferencias, a veces exageradas, entre el trabajo técnico y el trabajo intelectual. Los escritores hacen trabajos técnicos y los ingenieros hacen trabajos creativos. Sin embargo, se habla más del trabajo creativo del escritor y del trabajo técnico del ingeniero, pero ambos escriben. Este es solamente un ejemplo, quizá exagerado, para decir que estos cambios desplazan esas barreras, como lo viene diciendo desde hace décadas el campo interdisciplinar de las humanidades digitales.

Sin embargo, lo anterior no puede hacernos perder la conciencia crítica. Al contrario, acercarse a las tecnologías ayuda a mejorar nuestras capacidades críticas. El anónimo máquina o el anónimo humano siempre han estado presentes y lo sabemos, en buena medida, gracias a que ha habido campos de las ciencias sociales y humanas como la sociología del conocimiento, los estudios sociales de la ciencia y la tecnología, la filosofía de la tecnología, o la historia de la edición (solo para mencionar ejemplos cercanos) que, desplazando el foco de las grandes mentes individuales, nos han mostrado el trabajo colectivo y práctico (muchas veces no reconocido) que sostiene las grandes empresas culturales o científicas. Creo que hay un cambio de estratos, y que algunas obras y campos de investigación, que antes eran menores en el canon de las ciencias sociales y humanas, van a ir ganando terreno y adoptando un nuevo significado. Por otro lado, han aparecido objetos alternativos de estudio en las humanidades que me parecen provocadores como, por ejemplo, la crítica “literaria” de *software* de la que podrían encontrar un ejemplo en los trabajos de David M. Berry.

Una de las obras que se ha vuelto clásica con el tiempo, a pesar de su olvido durante los primeros años, es *El modo de existencia de los objetos técnicos*, escrita por el filósofo francés Gilbert Simondon a finales de los años 50. Uno de sus objetivos era denunciar la falta de cultura técnica en las sociedades contemporáneas derivada, en parte, de la brecha que iba abriendo el uso masivo de las máquinas versus el desinterés por el pensamiento técnico de sus usuarios. Simondon señaló que esa posición no solo dejaba de lado la comprensión de cómo evolucionaban esos objetos (pues evolucionan, aunque de manera diferente a los individuos biológicos), sino que parecía subvalorar las complejas e interesantes relaciones que establecen las personas con las máquinas y que no son, exclusivamente, relaciones de servilismo o de consumismo. Esa es una reflexión interesante para poner sobre la mesa, porque nuestras interacciones con los objetos técnicos pueden ser múltiples, aunque una de las formas más típicas de presentación de nuestros miedos —y también de nuestro optimismo—, se haga bajo los esquemas del servilismo o del exceso. De hecho, el enganche social de las IA generativas actuales profundiza la retórica de los asistentes. En efecto, las podemos usar así, incluso podemos abusar de este uso, pero no es la única relación posible y, por supuesto, no es la que deberíamos privilegiar en las universidades. Pero es un uso de entrada que habrá que ir cambiando creativamente por otros más productivos, y estos cambios no pueden hacerse desde la negación.

Además, hay otro aspecto interesante para las ciencias sociales y humanas: esas herramientas pueden convertirse en parte de nuestro laboratorio. Tenemos la posibilidad de incorporar a nuestras investigaciones

—y a nuestra manera de hacer— herramientas tecnológicas, sean o no de inteligencia artificial, que nos permitan, por ejemplo, sistematizar mejor la información, categorizarla, analizarla, explorarla, reutilizarla, dejarla a disposición del público. No hay que restarle importancia a una forma de pensamiento adversarial. Estos sistemas, en tanto sistemas capaces de procesar el lenguaje natural, pueden utilizarse para poner en juego las conclusiones a las que llegamos o para tener una segunda visión sobre la interpretación de los datos que estamos observando. Es decir, puede haber —no diría yo una racionalidad compartida—, pero sí un juego interesante en el que se pongan a prueba tareas, ideas e hipótesis. Hay análisis juiciosos que vienen de las llamadas Ciencias Sociales Computacionales (CSS en inglés) que han evaluado el impacto de los grandes modelos de lenguaje natural en nuestras disciplinas. Creo que valdría mucho la pena leerlos con atención.

Respecto a la formación, el hecho de que existan laboratorios y observatorios de humanidades digitales, así como formación en campos como la lingüística computacional, como ocurre en la Universidad de Antioquia, me parece una excelente oportunidad y una base fundamental, pues se trata de campos híbridos y abiertos al diálogo. También es muy importante la consolidación de proyectos de conservación y digitalización, que permitan aprender y explorar las tecnologías en el marco de proyectos concretos, y en Colombia tenemos un enorme trabajo que hacer en ese frente. Es un cliché, pero no podemos olvidar que uno de los primeros proyectos de humanidades digitales surgió, precisamente, con el procesamiento de datos literarios, como el índice de las obras de Santo Tomás de Aquino. Las denominadas hermenéuticas digitales, enriquecidas hoy con los grandes modelos de lenguaje, son herramientas muy útiles para las ciencias sociales y humanas en general y deberían dejar de ser formaciones de curiosos para convertirse en competencias generales.

Siguiendo en esta misma lógica, me gustaría preguntarte: ¿cómo ves los límites en el uso de esta inteligencia artificial, hablando específicamente en el ámbito de las ciencias sociales y las humanidades?

Creo que aquí hay una discusión realmente seria, pues hay aplicaciones de estas tecnologías que producen efectos que nos mueven profundamente, y por eso hay que tomárselas con cuidado. Se han reportado, por ejemplo, usos problemáticos de la inteligencia artificial, en casos en los que personas que deberían estar siendo acompañadas por médicos o por profesionales de la salud mental, terminan usando, por una razón o por otra, agentes artificiales. No digo que no se puedan explorar sistemas interesantes

para ciertas funciones como el seguimiento o la información, pero se requiere del control humano y esto nunca justifica ni desplaza el derecho que tenemos a que los sistemas de salud sean fuertes y a que las personas sean escuchadas y atendidas por personas. También hay investigaciones que muestran que la hiperconexión temprana, excesiva y no supervisada es riesgosa y tiene costos cognitivos importantes. Pero los proyectos del llamado bienestar digital avanzan muy lentamente, y Colombia puntea muy alto en el mundo en el uso excesivo de redes sociales.

También hay toda una discusión sobre la regulación de los usos de la IA. Esa es una de las tensiones del momento geopolítico que atravesamos. Creo que en ese contexto han surgido y se han ido consolidando trabajos en el derecho y en la ética de la inteligencia artificial. Esta última es una articulación llamativa entre la ética de la información y una reflexión profunda sobre el significado de los datos, especialmente los datos privados y personales que hacen referencia a nuestra identidad, respecto al enorme poder que tienen determinados sistemas para almacenar, tratar o hacer cosas con esos datos. Me parecen destacables los trabajos de Luciano Floridi, quien tiene un recorrido sólido de muchos años en la filosofía de la información, en la ética de los datos y, más recientemente, en la ética de la inteligencia artificial.

Creo que ahí hay una reflexión importante, y desarrollos que necesitan ser discutidos, por ejemplo, la manera como la ética de la inteligencia artificial ha ido involucrando técnicas de diseño e inteligencia colectiva, para que los desarrolladores de estas tecnologías se comprometan con el respeto a ciertos principios o discutan explícitamente ciertos riesgos. Ha habido una profundización importante de una ética principialista que, históricamente, se había aplicado con fuerza en el campo de la bioética y que hoy se emplea en el desarrollo de ciertos proyectos que usan tecnologías de inteligencia artificial. Estas distintas perspectivas éticas abren discusiones sobre los límites del uso de estas tecnologías y han aportado a la discusión sobre los sesgos algorítmicos, los sesgos raciales y de género, sobre las libertades o sobre las economías de la atención que hay detrás de algunos de estos productos. Estas reflexiones están muy vivas en las ciencias sociales y humanas actuales, y en Colombia hay varios trabajos que empiezan a tratar estos asuntos.

En el ámbito de la medicina, hay una discusión particularmente sensible acerca de la diferencia entre los datos denominado activos, es decir aquellos que son expresados en lenguaje natural o en el intercambio conversacional corriente, y los llamados datos pasivos, que son aquellos registrados inmediatamente por captores y sensores integrados a nuestros dispositivos y que no pasan por el lenguaje natural, ni por

la declaración consciente de algún intercambio de información, como ocurre, por ejemplo, con la geolocalización o con los datos biométricos. Hay una tendencia importante a tomar en cuenta los denominados datos pasivos en distintas áreas, pues resultan ser objetivos y bastante informativos sobre síntomas y conductas. Se habla en estas investigaciones de la identificación del *fenotipo digital*. Pero la discusión ética es delicada porque tenemos libertades y derecho a la privacidad. Por ejemplo, las tecnologías basadas en datos pasivos fueron empleadas, de manera muy controversial durante la pandemia por Covid 19, para intentar predecir los contactos entre personas infectadas, usando la información proporcionada por el Bluetooth, por la geolocalización y por los datos de contacto de los smartphones. Allí hay preguntas éticas para hacernos, pero también debe haber una conciencia sobre lo que hacen las máquinas y las tecnologías que usamos. Aparece de nuevo la idea de Simondon de la importancia de mejorar nuestra cultura técnica.

Las máquinas que usamos en la vida cotidiana capturan trazas y rastros que generalmente no sabemos u olvidamos que estamos dejando. Tienen, además, un tipo de memoria diferente a la nuestra. Así que, si esos datos continúan siendo almacenados, pueden recordarnos durante mucho más tiempo del que imaginamos. Lo que suele hacerse para regular estos usos es establecer normativas nacionales e internacionales, pero todavía dependemos, en buena medida, de la voluntad de los gobiernos para adherirse a los marcos regulatorios. Siempre hay intereses de por medio que se ponen en la balanza. Por ejemplo, las empresas que buscan desarrollarse y piden desregulación o *soft laws* para evitar someterse a normativas exigentes que hagan menos rentables sus proyectos, o que supongan limitaciones sustanciales. Por otro lado, hay trabajos interdisciplinarios interesantes sobre lo que se llama alinear principios con tecnologías, por ejemplo, hacer que un algoritmo sea confiable. Sin embargo, la confianza es una noción abstracta y vaga, lo que se complica aún más cuando se enfrenta a problemas tecnológicos complejos como la denominada opacidad algorítmica. En ese tipo de zonas de escollo e indefinición hay trabajos emergentes, por ejemplo, entre filósofos, científicos sociales e ingenieros de IA.

A pesar de que no tenga una respuesta concreta sobre cómo reconocer mejor los límites, estoy convencido de que parte importante de la reflexión está en cerrar brechas de comprensión y trabajo entre las disciplinas digamos más tecnológicas y las ciencias sociales y humanas. Con esto no me refiero a hacer charlas donde hagamos coexistir en el mismo lugar a representantes de las dos esferas. Estoy pensando en que se puedan hacer trabajos conjuntos.

Para terminar y para que podamos ilustrar con un ejemplo este uso de la inteligencia artificial, ¿podrías contarnos en qué consiste tu investigación y cómo la inteligencia artificial ha abierto nuevas posibilidades para ti como investigador?

Mi investigación hace parte de un proyecto en marcha, liderado por el médico francés Pascal Possoz, quien tiene un recorrido académico y profesional muy interesante, además de una rica cultura humanística. Su formación inicial fue en medicina, específicamente como gastroenterólogo y hepatólogo, pero desde hace varias décadas se ha involucrado de lleno en el tratamiento de las adicciones. Ha sido parte de la fundación de varias unidades de adictología en distintos hospitales aledaños a Montpellier, la ciudad francesa en la que me encuentro. La unidad más reciente fue inaugurada a finales del año 2022 en Saint Clément-de-Rivière. Es una unidad original, porque integra desde su concepción la búsqueda de lazos con las humanidades digitales, especialmente para construir alternativas en la continuidad del tratamiento después de que las personas han terminado su hospitalización que, en general, no suele ser muy larga. En Francia las adicciones son reconocidas como enfermedades crónicas y su tratamiento es cubierto por el sistema de salud. Además de los hospitales, existen centros especializados de reducción de riesgos y daños en casi todas las ciudades y, en algunos casos, apartamentos terapéuticos.

En este proyecto tengo dos misiones principales. Primero, ayudar a entender el qué y el para qué de la introducción de ciertos dispositivos digitales como los sistemas de alertas, los chatbots o las terapias de realidad aumentada aplicadas al tratamiento de las adicciones. Esto incluye también una consideración de cómo estas tecnologías son asumidas y valoradas por todos los actores, incluidos los equipos médicos y paramédicos. Es un trabajo interesante porque ayuda a entender la introducción de agentes tecnológicos de manera menos idealizada. Por otro lado, construyo una base de datos conversacional, de conversaciones naturales, a partir de las interacciones entre los pacientes, los equipos y los pacientes expertos. Esta base de datos conversacional es interesante para las ciencias del lenguaje, pero también para los equipos médicos y paramédicos involucrados. Pero, sobre todo, es un ejercicio de escucha de una situación sobre la que, curiosamente, solemos tener ideas fáciles, estigmatizantes y moralizantes, a pesar de lo difícil, multicausal y complejo que resulta definir y estudiar las adicciones. He tenido la oportunidad de estar regularmente en la clínica durante casi tres años y he podido contar con la confianza del equipo médico y, muy especialmente, de los pacientes y de los pacientes expertos.

Cuando formulamos el proyecto a principios del año 2022, estábamos explorando la posibilidad de participar en la creación de agentes conversacionales artificiales. Las referencias más conocidas para ese momento en el ámbito de la salud mental eran *Wyza* y *Woebot* en el mundo anglófono y *Owlie* en el mundo francófono. Aunque no es lo único que los define, todos tienen una base fuerte en las denominadas terapias cognitivas comportamentales que hunden sus raíces en proyectos al tiempo tecnológicos y de ciencias cognitivas como el que imaginó Colby, así que entender cómo una tecnología pretende encarnar una teoría o una práctica siguen siendo temas de reflexión. Sin embargo, a pesar del uso de algunos de estos programas en el tratamiento de ciertas conductas adictivas, ninguno de ellos fue concebido pensando en las adicciones, lo que al tiempo suponía un estudio de campo previo de los dispositivos clínicos en la adictología, pues la tradición francesa en el tratamiento de las adicciones tiene algunas particularidades. Para dar un ejemplo rápido, los adictólogos no son necesariamente psiquiatras, aunque conforman equipos con ellos, y la llamada medicina somática ocupa un lugar importante en los tratamientos. En general, las unidades de adictología son ricas en un trabajo interdisciplinar entre trabajadores sociales, psicólogos, adictólogos, fisioterapeutas, neuropsicólogos, psiquiatras y pacientes expertos.

Este proyecto se lleva a cabo en una cooperación entre el laboratorio LHUMAIN de la Universidad de Montpellier Paul-Valéry, específicamente con los investigadores François Perea, quien ha hecho investigaciones sobre el discurso coloquial y el consumo de alcohol en los bares y, más recientemente, sobre objetos conectados y agentes conversacionales de voz, y con Bruno Bonu, quien tiene una trayectoria de muchos años en investigaciones sobre la telemedicina y el análisis de la conversación.

Así que, me encuentro en un ámbito de las ciencias del lenguaje en el que se estudian las interacciones en un sentido muy amplio, tanto las interacciones naturales, cara a cara, como las interacciones mediadas por la tecnología. Y mi proyecto, en particular, toma en cuenta ambas.

Quiero resaltar que los estudios sobre las adicciones vienen en aumento en las ciencias sociales y humanas. Son muy interesantes, por ejemplo, los trabajos de filosofía de la adicción que ha venido haciendo los últimos años el filósofo argentino Federico Burman. También ha habido esfuerzos grandes de acercamiento interdisciplinar de los investigadores que vienen del denominado modelo animal de las adicciones, como Serge Ahmed, quien editó un handbook con la filósofa Hanna Pickard en el que se presentan un panorama amplio de discusión en el que se incluyen algunos trabajos de filosofía y ciencias sociales. También hay reflexiones teóricas clásicas en las ciencias sociales como las de Jon Elster.

Sin embargo, estos esfuerzos teóricos y metodológicos no siempre llegan a los ámbitos clínicos a pesar del interés de los problemas. Por eso he encontrado muy original el trabajo del Dr. Possoz. Él hace un taller semanal que bautizó *Philosophie du soin* (filosofía del tratamiento o del cuidado) en el que aborda con los pacientes muchos de esos temas complejos, por ejemplo, el problema de las definiciones de la adicción, lo que implica reconocer que la adicción es una enfermedad, la relación compleja entre adicción, decisiones y hábitos, el rol de las emociones, o las particularidades del funcionamiento de las moléculas, solo por mencionar algunos.

El próximo año iniciaremos la fase más tecnológica del proyecto. Sin embargo, hemos podido hacer algunos pilotos que nos dejan cosas interesantes para reflexionar. Una idea que apareció desde el origen mismo de los agentes conversacionales, incluso en la época de Weizenbaum y que, debo decirlo, al comienzo, no me convencía mucho —me parecía un poco exagerada—, pero que con el tiempo he ido aprendiendo a tomar en cuenta, es que hay personas que prefieren hablar con máquinas porque no quieren la mirada humana o porque hay conversaciones tan difíciles que no quisieran tenerlas con otro ser humano, sino solo con el lenguaje. En ese contexto, este tipo de tecnologías pueden aportar a la creación de ciertos "espacios", digamos, de anonimato o de neutralidad, siempre y cuando exista una base sólida de estudios que validen las respuestas ofrecidas por los sistemas, y que comprendan también cuáles son las preguntas que las personas se hacen frente a una situación de la que les es difícil hablar, por lo menos por un cierto tiempo. Sin embargo, sigo pensando que siempre debe existir la posibilidad de hablar con un humano. Otra cosa que hemos visto es que a través de estos dispositivos tecnológicos se inician conversaciones difíciles que luego los equipos pueden continuar y retomar en el tratamiento o en el seguimiento, y que no eran fáciles de introducir personalmente, porque generaban algún tipo de vergüenza o reticencia, pero que una vez transmitidas por escrito pueden abordarse. Como lo dice la sabiduría popular, es difícil dar el primer paso. Como lo muestra el análisis de la conversación, los turnos de palabra se abren y se cierran, pero lo realmente relevante es entender y describir cómo. —♦♦—