INTERACCION
Revista Digital de AIPO
Asociación Interacción Persona-Ordenador
Vol. 3, No 2 (2022)
ISSN 2695-6578

De la interacción con máquinas a la colaboración con agentes inteligentes

From interaction with machines to collaboration with intelligent agents

José J. Cañas
CIMCYC (Centro de Investigación
Mente, Cerebro y Comportamiento)
Universidad de Granada
Granada, Andalucía, España
delagado@ugr.es
Recibido: 06.12.2022 | Aceptado: 06.12.2022

Palabras Clave

Interacción Ser Humano-Máquina
Colaboración Ser Humano-Inteligencia Artificial
Inteligencia Artificial
Modelos Mentales
Apoyo por ordenador al trabajo en equipo colaborativo
Antropomorfismo

Resumen

En las actividades humanas, las máquinas han sido tradicionalmente consideradas herramientas con las que los sujetos de estas actividades interactúan. Por ello, en el diseño de las máquinas se ha tenido en cuenta el diseño de esa interacción y de las interfaces donde esa interacción tiene lugar. Sin embargo, la introducción de la inteligencia artificial en las máquinas sopone que empecemos a considerar a éstas como otros sujetos que colaboran con los sujetos humanos en la realización de la actividad. Por esta razón, debemos de empezar a hablar del diseño de la colaboración con las máquinas en lugar de la interacción con ellas. Este cambio de conceptualización del diseño de máquinas supondrá una nueva mirada a la Psicología científica para que ésta aporte sus conocimientos sobre como los seres humanos colaboramos. Ya no se trata de diseñar las interfaces de las máquinas para que éstas se ajusten a las características humanas, ahora se trata de diseñar a las máquinas para que éstas comprendan a los seres humanos con los que colaboran y éstos las vean como seres inteligentes similares a los seres humanos. Tendremos que dejar de hablar sólo de diseño de interfaces para hablar también de colaboración con agentes inteligentes.

Keywords

Human-Machine Interaction
Human-Artificial Intelligence collaboration
Artificial intelligence
Mental Models
Computer Suported Collaborative Work
Anthropomorphism

Abstract

In human activities, machines have traditionally been considered tools with which the subjects of these activities interact. For this reason, in the design of the machines we have taken into account the design of that interaction and of the interfaces where that interaction takes place. However, the introduction of artificial intelligence in machines means that we should begin to consider them as other subjects that collaborate with human subjects in carrying out the activity. For this reason, we should start talking about the design of collaboration with machines instead of interaction with them. This change in the conceptualization of machine design will mean a new look at scientific psychology so that it contributes its knowledge about how human beings collaborate. It is no longer about designing the interfaces of the machines so that they conform to human characteristics, now it is about designing the machines so that they understand the human beings with whom they collaborate, and they see them as intelligent beings similar to humans. We will have to stop talking only about interface design to talk also about collaboration with intelligent agents.

1. El papel de las máquinas en las actividades humanas en la visión tradicional de la interacción Persona-Máquina

Tradicionalmente, el papel que las máquinas tienen en la vida humana se ha analizado en el marco del concepto de *actividad*. Este concepto fue propuesto dentro de la tradición de la Psicología Soviética como parte de lo que se conoce como la Teoría de la Actividad (Leotiev, 1977). Esta teoría se caracteriza fundamentalmente por su focalización en los efectos socio-culturales sobre el pensamiento y la acción humanos, así como por su fundamentación en el pensamiento materialista en lugar del pensamiento idealista. Los conceptos introducidos por la Teoría de la Actividad suponen un cambio de perspectiva de lo individual a lo colectivo y son útiles para estructurar una situación compleja donde las personas cooperan unas con otras y utilizan herramientas.

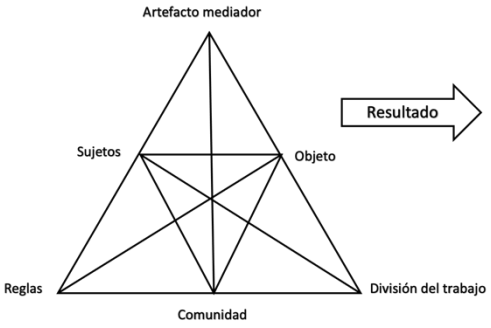

Figura 1: Los componentes de la actividad según la Teoría de la Actividad (Leotiev, 1977)

En la Figura 1 se pueden ver los componentes de la actividad humana y las relaciones entre ellos según esta teoría. En esta teoría se parte de que toda actividad tiene un objeto sobre el que un sujeto actúa para alcanzar un objetivo. Este objeto es el problema, la situación y el foco de la actividad. Por su parte, el sujeto de la actividad es el agente que actúa sobre el objeto, pero *un sujeto puede ser una persona individual o un grupo de personas*. La interacción entre el sujeto y el objeto significa que el sujeto transforma al objeto mientras que éste influencia la conducta del sujeto. Y esa interacción ocurre en un contexto y siguiendo las motivaciones del sujeto. Toda actividad está regida explícita o implícitamente por unas reglas y normas que determinan cuales son las conductas apropiadas. Estas reglas y normas son definidas por una comunidad de individuos o grupos que pueden tener diferentes objetos y objetivos parciales pero que interactúan entre ellos para alcanzar un objetivo global. Este concepto de comunidad es el que permite entender lo que ocurre en una situación como una fábrica donde tenemos un grupo de sujetos que actúan cada uno sobre un objeto diferente interactuando entre ellos con un fin común. Las normas y las reglas también definen cuales son las divisiones de las tareas o trabajos en una comunidad o en la sociedad.

Finalmente, tenemos a los artefactos o máquinas que en la teoría clásica son considerados como "mediadores" entre el sujeto y el objeto. Pensemos en un ejemplo que nos puede ayudar a entender el rol de mediador que el artefacto o máquina juega en la teoría clásica: un arado para arar la tierra. El sujeto sería el agricultor y el objeto sería la tierra. En esta actividad el arado sería un mediador entre el sujeto y el objeto. Desde que el ordenador apareció como un componente de la actividad humana éste fue considerado como otra herramienta en el sentido de la Teoría de la Actividad (ver por ejemplo, Nardi, 1986).

Sin embargo, desde que se introdujo el ordenador como un artefacto en la actividad, nuestra concepción de la máquina como herramienta mediadora en la actividad ha ido cambiando. Desde el rol que como herramienta mediadora tenía la máquina según la Teoría de la Actividad clásica, estamos pasando a verla como un sujeto más de la actividad. Y esta evolución en nuestra concepción del rol de las máquinas está teniendo importantes consecuencias en cómo las diseñamos y, sobre todo, en cómo diseñamos la interacción del ser humano con ellas. Este cambio se ha debido, fundamentalmente, a que el ordenador ha traído con él, sobre todo, los mecanismos automáticos y la inteligencia artificial.

2. La introducción de los automatismos en las máquinas

La introducción del concepto de máquina en la historia de la evolución humana para sustituir el concepto de artefacto ocurrió cuando a las herramientas que se usaban en las actividades se les incorporó un mecanismo que mediante un proceso de consumo de energía les permitía que hiciesen su trabajo sin necesidad de la fuerza humana o animal. El paso del arado al tractor significó el paso de la herramienta como artefacto a la herramienta como máquina. En términos de la Teoría de la Actividad clásica, lo que ocurrió fue que se inició una distribución de funciones entre los seres humanos y las máquinas de tal manera que algunas de las funciones que los seres humanos o los animales hacían antes, ahora eran responsabilidad de las máquinas.

Por esta razón se comenzó a hablar, y se sigue actualmente hablando, de distribución de funciones cada vez que una máquina se introduce en una actividad humana. Se considera que al introducir la máquina lo que se está haciendo es asignarle una función a la máquina que antes hacía el ser

humano, posiblemente de una forma menos eficiente. Esta asignación de funciones puede significar una de dos cosas. Una es que se deja que la máquina haga la función por sí sola (un motor para sacar agua de un pozo para el riego sustituye a un sistema de poleas que una persona maneja). Pero también puede significar que la máquina ayude al ser humano a hacer la función mejor y más eficientemente (se introduce una grúa mecánica operada por un ser humano para levantar pesos que éste no podría levantar con sus propias fuerzas) (Hollnagel y Bye, 2000).

Esta asignación de funciones a las máquinas, a la que hemos llamado automatización se comenzó a hacer asignándole a éstas solo funciones físicas que antes realizaban los seres humanos o los animales. Sin embargo, a lo largo del siglo XX hemos asistido a un proceso cada vez más intenso de asignación a las máquinas de funciones "mentales". Cuando a las máquinas de escribir se les introdujeron mecanismos eléctricos que hacían más fácil la actividad de escribir un texto, estos mecanismos eran funciones físicas que ahora tenían las máquinas de escribir. Sin embargo, en un momento determinado de las últimas décadas del Siglo XX, se comenzaron a asignar a las máquinas funciones mentales que antes estaban en manos de los seres humanos. Un ejemplo de esto lo tenemos cuando a alguien se le ocurrió introducir en los procesadores de texto una función de corrección ortográfica y semántica. Esa función ya no era física sino mental. Ahora, dejamos que el procesador de texto se ocupe de verificar si hemos cometido un error ortográfico y podemos configurarlo para que lo corrija sin siquiera consultarnos. No hace falta decir que nuestra vida es más fácil porque ahora no tenemos que preocuparnos de cometer faltas de ortografía. El peligro es que dejemos de saber escribir correctamente porque esta función cognitiva la dejemos en desuso. Ese es un peligro que va a existir ahora que las funciones cognitivas humanas las hacen las máquinas. Pensemos en un caso similar que ocurre con el uso de la calculadora. Con ella podemos hacer operaciones que podríamos seguir haciendo a mano o mentalmente pero que con la incorporación generalizada de su uso hemos dejado de saber cómo hacerlo. ¿Quién sabe ya hacer aún raíces cuadradas u operar con logaritmos? Otro ejemplo de esto lo tenemos en el caso de los coches. Una de las operaciones que se hace durante la conducción es el cambio de marchas. Este cambio de marchas puede hacerlo el conductor o puede hacerlo el mismo coche. En este segundo caso hablamos de coches con cambio de marchas automático.

En todos estos ejemplos hemos seguido hablando de máquinas automáticas y a nivel conceptual hemos seguido pensando en ellas como herramientas de la actividad. Como consecuencia de ello, por ejemplo, el paradigma dominante de la usabilidad y de la experiencia de usuario en la interacción Persona-Ordenador está basado en la concepción del ordenador como máquina, todo lo más como maquina con ciertos automatismos (Beynon, 2019).

Sin embargo, ahora cada vez más hablamos de máquinas inteligentes. Ya no hablamos sólo de máquinas automáticas, ahora pensamos que algunas de estas máquinas son inteligentes. A algunos de los coches que antes tenían cambios de marchas automáticos, ahora se les llama coches inteligentes y la pregunta que podemos hacernos es cuál es la diferencia entre automatización e inteligencia y qué consecuencias tendrá este cambio conceptual en la forma como abordamos el diseño de las máquinas inteligentes y de la interacción de los seres humanos con ellas. Pero detengámonos un momento a pensar que significa que una máquina sea inteligente y que la diferencia de una máquina automática.

En primer lugar, podemos decir que mientras más funciones mentales asignemos a la máquina cuando estamos distribuyendo las funciones entre ésta y el ser humano más difícil será decir que la máquina no ha adquirido cierto nivel de inteligencia. Sin embargo, la inteligencia no la vamos a medir por el número de funciones mentales que tiene la máquina, hay algo más que define inteligencia y que están adquiriendo las máquinas actuales. A los coches con cambios automáticos de marchas no los llamamos inteligentes ni a un procesador de textos con corrector ortográfico tampoco aunque realicen funciones mentales (decidir cambiar de marchas o cambiar la ortografía de un texto). Lo que define inteligencia es la capacidad para tener control sobre las propias acciones y corregir la propia conducta si el resultado de ésta no es el apropiado o el deseado. Imaginemos que tenemos un procesador ortográfico que cambia nuestro texto si considera que no es correcto ortográficamente, pero además comprueba si el texto que sugiere es el correcto y si no lo es lo vuelve a corregir. Por ejemplo, imaginemos que estamos escribiendo una frase en la que escribimos "tomo" y el corrector nos corrige y sustituye nuestra palabra por "tomó" pero el corrector sigue interpretando el texto que estamos escribiendo haciendo un análisis semántico y "se da cuenta" de que realmente lo que nosotros queríamos decir era "tomo" y no "tomó". Entonces, vuelve a cambiar la palabra para dejar la que nosotros pusimos.

Por lo tanto, podemos decir que si el control sobre ciertas acciones lo pasamos a la máquina al asignarle funciones cognitivas humanas, ésta adquirirá cierta inteligencia. El nivel de inteligencia que la máquina adquiere dependerá del número de funciones cognitivas que le asignemos, pero, sobre todo, en términos psicológicos solo podemos decir que la máquina es inteligente si tiene un control autónomo sobre sus acciones de tal manera que pueda evaluarlas y corregirlas si es necesario. En otras palabras, podemos afirmar que la razón

fundamental para llamar a una máquina un ente inteligente es que tiene control sobre sus propias acciones.

La consecuencia fundamental que ha tenido esta evolución de las máquinas hasta el momento actual en el que estamos hablando de máquinas inteligentes es que tendremos que empezar a cambiar el role que éstas tienen ahora en el esquema de la Teoría de la Actividad. Ahora las máquinas son también sujetos de la actividad que colaboran con los sujetos humanos. La Inteligencia Artificial nos obliga a pensar en la máquina como otro sujeto de la actividad, cada vez más al mismo nivel que el sujeto humano. Es evidente que seguirán existiendo máquinas que son herramientas, pero ahora tendremos que interactuar con máquinas que son también sujetos de la actividad. Pensemos en un tractor al cual se le han introducido sensores para detectar las condiciones de la tierra y que tiene algoritmos que deciden que ararán esa tierra en función a esas condiciones. ¿Tendremos que considerar que en la actividad de arar la tierra sigue habiendo solo un sujeto y el tractor es solo una herramienta? Evidentemente, no. Ahora tendremos dos sujetos, la persona y el tractor. Aunque, también es evidente que algunas partes del tractor podrán seguir siendo consideradas como herramientas, por ejemplo, las rejas del arado.

Cómo hemos dicho más arriba, la razón para que ahora hablemos de máquinas inteligentes debemos encontrarla en la pérdida de control sobre la actividad que experimenta la persona que interactúa con la máquina. Mientras que las máquinas automáticas ejercían funciones que podemos definir como *no mentales* o *como mentales en el sentido que hemos dicho más arriba* (por ejemplo, el cambio de marcha en algunos coches) la distribución de funciones entre las personas y las máquinas era considerada como un proceso donde el ser humano mantiene todo el control sobre la actividad. El conductor de un coche con cambio automático de marchas siempre se ha sentido con el control sobre la conducción. Sin embargo, actualmente la distribución de funciones empieza a considerarse cómo cierta pérdida de control por parte de la persona, debido, fundamentalmente, a que el número de funciones que asignamos a las máquinas es mucho mayor. Pero, además y más importante, ahora la persona pasa, en el mejor de los casos, a supervisar las acciones de la máquina de tal manera que el control sobre estas acciones está casi completamente en manos de la máquina. Ahora, las máquinas a las que llamamos inteligentes disponen de un sistema de retroalimentación (son sistemas de bucle cerrado en terminología de la ingeniería) que les permite corregir sus acciones si detectan que no llevan a los resultados deseados.

Ahora, empezamos a darnos cuenta de que inteligencia significa también tener control sobre las acciones. Si el control en una tarea es siempre humano y las funciones mentales principales que controlan y supervisan las acciones en la actividad permanecen estando en manos de los humanos, la persona será considerada como inteligente y la máquina no lo será o será considerada, en el mejor de los casos, con una inteligencia muy limitada. Si el control sobre ciertas acciones lo pasamos a la máquina al asignarle funciones cognitivas humanas más complejas y las acciones dejan de estar bajo control del ser humano, entonces pensaremos que la máquina adquirirá inteligencia porque esas funciones cognitivas complejas le darán más control sobre sus propias acciones en la actividad conjunta. Por tanto, podemos decir que el nivel de inteligencia que la máquina adquiere dependerá del número de funciones cognitivas que le asignemos, pero sobre todo del control autónomo que ésta tenga sobre sus propias acciones.

Dentro del marco conceptual de la Teoría de la Actividad, la consecuencia fundamental que tiene el asignar inteligencia a las máquinas es que éstas se convertirán en sujetos de la actividad como ya lo eran los seres humanos. Aunque seguirá habiendo máquinas que tendrán el papel de herramientas, otras máquinas se convertirán en sujetos de la actividad y tendremos que considerar que "colaboran" con los seres humanos para alcanzar los objetivos de la actividad. Tendremos que hacer una modificación en el marco conceptual de la teoría de la Actividad como se refleja en la Figura 2.

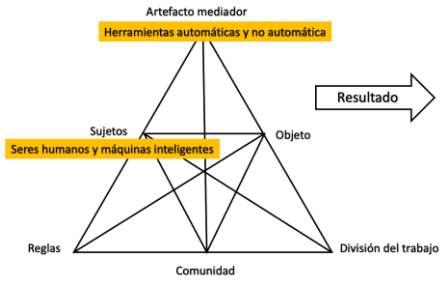

Figura 2: Componentes de la actividad cuando las máquinas inteligentes participan en ella colaborando con los seres humanos

Como consecuencia de esto, ahora debemos preguntarnos, en primer lugar, ¿cómo se diseña un ser inteligente? Y, sobre todo, ¿cómo se diseña la interacción con un ser inteligente artificial? Al contestar a estas preguntas, nos damos cuenta de que en realidad son las mismas preguntas que ha venido haciéndose la Psicología científica en los dos últimos siglos. Evidentemente, en Psicología no hablamos de "diseño", pero ¿no es verdad que los psicólogos evolutivos y los psicólogos educativos han estado desde siempre preguntándose sobre cómo se adquiere y desarrolla la inteligencia y los psicólogos sociales llevan dos siglos estudiando las interacciones entre

seres inteligente?, Por tanto, ¿no sería conveniente que mirásemos lo que la Psicología científica tiene que enseñarnos sobre el diseño de las máquinas inteligentes y sobre la interacción con ellas?

3. Del diseño de máquinas como herramientas al diseño de las máquinas inteligentes

En la visión tradicional de las máquinas como herramientas de la actividad humana, siempre ha sido patente que el diseño de éstas debe incluir también el diseño de la interacción con ellas. De hecho, el propio concepto de diseño está ligado al concepto de interacción. Cuando un ser humano toma una piedra del suelo y cambia su forma para poder cogerla mejor empezamos a hablar de diseño como una modificación (manufactura) de un objeto para convertirlo en una herramienta con la que sea más fácil interactuar.

Un análisis histórico del diseño de la interacción entre los seres humanos y las máquinas que éstos han creado nos revela que podemos diferenciar tres grandes periodos. En el primero que va hasta comienzos de los años 80 del siglo pasado, el diseño de las máquinas se hacía pensando en éstas y no en las características de las personas que las usarían después. Por esta razón, a este periodo se le conoce como el del "diseño centrado en el sistema" (aquí la palabra sistema se usaba para referirse la máquina diseñada por la persona). El problema al que siempre había que enfrentarse durante ese periodo histórico del diseño era que siempre habría que esperar a que las personas que utilizarían la máquina ya diseñada tendrían que adaptarse a su diseño, se hubiese tenido en cuenta o no las características humanas a la hora del diseño. Se esperaba, por así decirlo, que la persona para usar la máquina debía "aprender" a usarla. Sin embargo, pronto se hizo evidente que los usuarios de las máquinas tenían grandes dificultades para usarlas, precisamente porque durante su diseño no se había tenido en cuenta sus características físicas y, sobre todo, psicológicas y se exigía de ellos que se adaptasen a características del diseño que sobrepasaban los límites de sus capacidades.

Por esta razón, a partir de los años 80 del siglo pasado se comenzó a hablar de un "diseño basado en el usuario" en el que las características físicas y psicológicas de éste se tenían en cuenta desde el comienzo. En un libro ya clásico de Donald Norman y Stephen Drapper (1986) se planteó la necesidad de este cambio de perspectiva desde el diseño basado en el sistema al diseño basado en el usuario.

Más recientemente se ha comenzado a analizar la interacción en el diseño considerando el papel que las máquinas tienen en las formas de vida de las personas además de la adaptación del diseño de éstas a las características físicas y psicológicas humanas (Saariluoma, Cañas y Leikas, 2016). En términos de la Teoría de la Actividad, podemos decir que se ha comenzado a considerar el diseño de la interacción en función del "objetivo" de la actividad que hasta ahora se había ignorado muchas veces. Pero aún en este análisis de los objetivos en función de las formas de vida se sigue considerando a las máquinas como unas herramientas mediadoras que ayudan a la vida de las personas.

Lo importante es que reconozcamos que durante estos tres periódicos históricos del diseño de las máquinas se ha considerado siempre que éstas son unos entes "sustancialmente" diferentes a los seres humanos que las utilizan. De hecho, para referirse a ellas siempre se ha utilizado la palabra "máquina" y a la interacción con ellas se le llama "interacción Persona-Máquinas" (en la terminología inglesa se utiliza el término Human-Machine Interaction o HMI). Cuando la máquina es un ordenador se utiliza el término "Interacción Persona-Ordenador" (en inglés Human-Computer Interaction o HCI). Las máquinas son herramientas que los sujetos de la actividad utilizan para actuar sobre el objeto de ésta.

Sin embargo, cómo hemos dicho, esta forma de entender el papel que las máquinas juegan en las actividades humanas y cómo debemos diseñarlas para que la interacción con ellas sea eficiente y eficaz, tiene que ser revisada actualmente debido a la introducción de los automatismos y, sobre todo, de la inteligencia artificial en ellas. Pero, ¿cómo debemos hacer esta revisión y cómo debemos plantearnos el diseño de las máquinas inteligentes y de la interacción con ellas? La respuesta a estas preguntas es simplemente que debemos diseñar pensando que están interactuando dos entes inteligentes que colaboran como sujetos de la actividad. Esto significará que debemos volver nuestra mirada, más que nunca, hacía lo que interacción entre personas significa psicológicamente.

4. Dos ejemplos que muestran el cambio de paradigma

Para entender este cambio de paradigma podemos tomar dos ejemplos del ámbito de la Interacción Persona-Ordenador donde el ordenador está empezando a dejar de ser considerado como una herramienta y está pasando a convertirse en un sujeto de la actividad. El primero de ellos es el papel que ha jugado el concepto de modelo mental en el diseño de las interfaces y como éste debe ser modificado o, quizás, reemplazado con el concepto de antropomorfismo para entender cómo será el diseño de máquinas inteligentes en el futuro. El segundo ejemplo es el del diseño de los sistemas informáticos de apoyo al trabajo colaborativo. Estos sistemas han sido hasta ahora diseñados pensando en el ordenador como una herramienta para facilitar la colaboración entre seres humanos durante la realización de una actividad. Es

posible que ahora tengamos que pensar que el ordenador puede ser también un sujeto colaborador al mismo nivel que los seres humanos que participan en el trabajo colaborativo.

4.1 Modelos mentales de la máquina versus Interacción con mentes semi-humanas

Podemos decir que, durante la ejecución de una actividad, la máquina tiene un comportamiento que debemos comprender para poder interactuar con ella correctamente. Para comprender las acciones de una máquina no es sólo necesario que percibamos estas acciones y sus efectos sobre la actividad, también es necesario combinar esta información perceptual con el conocimiento que tenemos almacenado en nuestra memoria sobre su funcionamiento y su estructura. La comprensión significa combinar la información percibida con el concomimiento almacenado en nuestra memoria. Para comprender por qué nuestro coche se ha parado tenemos que saber cómo funciona y de esta manera hacer hipótesis sobre las posibles causas de que se haya dejado de mover. Por ejemplo, sabemos (tenemos almacenado conocimiento en nuestra memoria sobre ello) que el coche necesita energía para funcionar. Por lo tanto, nuestra primera hipótesis puede ser que nos hemos quedado sin gasolina.

Llevamos años preguntándonos sobre cuál es este tipo de conocimiento sobre las máquinas que tenemos almacenado en nuestra memoria, cómo está almacenado, como se recupera, etc. En este sentido, en la visión tradicional de la interacción entre las personas y las máquinas, entendidas éstas como herramientas dentro de la estructura de la actividad, hemos venido hablando de una representación mental de la máquina a la que hemos llamado modelo mental y que nos ha servido para diseñar la interacción con la máquina de tal manera que comprendamos su comportamiento y proyectemos sus acciones futuras.

En el ámbito de la interacción con ordenadores se ha demostrado numerosas veces que cuando una persona aprende a interactuar con el ordenador, adquiere conocimiento sobre su estructura y funcionamiento. Las investigaciones han demostrado, por ejemplo, que la adquisición de un modelo mental del ordenador facilita el aprendizaje de programación (Moran, 1981; Kieras y Bovair, 1984; Cañas, Bajo y Gonzalvo, 1994; Navarro y Cañas, 2001).

El modelo mental de una máquina, como es el ordenador, no tiene que ser un conocimiento preciso y veraz de la estructura y funcionamiento reales de ésta. Se asume que el modelo mental es un conocimiento que depende de lo que el sujeto de la actividad necesita para interactuar correctamente con la máquina. Por ejemplo, el modelo mental de un mecánico de coches es diferente del modelo mental de los conductores de éstos. De la misma forma, el modelo mental del ordenador de un ingeniero informático es diferente del modelo mental de un usuario.

Para poder interactuar con el ordenador, un usuario no necesita saber cómo funcionan los componentes físicos de los chips, solo necesita saber cuales son los componentes del ordenador, donde se pueden almacenar y cómo se procesan los datos con ellos. Por esta razón, cuando pensamos en términos de modelos mentales en el diseño de las interfaces de las máquinas es frecuente que se recurra a utilizar metáforas de objetos, ambientes, entornos, o espacios familiares para los usuarios para facilitar la adquisición y uso del modelo mental. Por ejemplo, las interfaces de manipulación directa (Shneiderman, 1997) se diseñaron muchas de ellas utilizando la metáfora de una oficina. En el escritorio de la interfaz encontramos documentos, carpetas, papeleras, etc. que representan objetos con los que se trabaja en una oficina. Evidentemente, cuando se lleva un documento a una carpeta no estamos haciendo referencia a como la información se almacena en los chips del ordenador, estamos simulando las acciones que hacemos cuando estamos trabajando en nuestro entorno de trabajo tradicional en una oficina. Por lo tanto, la metáfora nos ayuda a crear un modelo mental del ordenador que nos permite interactuar con éste durante la actividad sin necesidad de que tengamos un conocimiento real de la estructura física del ordenador, cómo el que tendría un ingeniero informático.

Figura 3: Modelos mentales de un ordenador

La mayoría de las interfaces actuales están diseñadas siguiendo la metáfora del escritorio. En ella, el usuario trabaja sobre una estructura basada en carpetas (directorios), mediante la manipulación directa de documentos (archivos). Sin embargo, cuando se diseña una interfaz utilizando una metáfora para que el usuario adquiera un modelo mental que facilite la interacción con la máquina (el ordenador en nuestro caso) debemos tener en cuenta que este modelo mental basado en una metáfora puede ser útil para interactuar con la interfaz durante la realización de ciertas actividades, pero esta misma metáfora puede ser un obstáculo para interactuar con la máquina en otras actividades. Un caso que podemos poner de ejemplo de este efecto negativo de los modelos mentales

basados en metáforas lo tenemos en lo que ocurre cuando un usuario comienza a interactuar con Internet después de haber aprendido el modelo mental de la oficina. Internet rompe drásticamente con la estructura de carpetas de los actuales sistemas operativos. De hecho, muchos usuarios noveles adoptan un modelo mental de Internet centrado en su ordenador y en el que la red simplemente es una extensión por contenidos (carpetas) de lo que suelen manejar en su PC. Debido a la incompatibilidad de los modelos inducidos por las metáforas, el usuario se encuentra con problemas tales como por ejemplo no saber "en qué carpeta está guardado el vídeo" que acaba de visionar directamente en su navegador.

Por tanto, podemos decir que el uso del concepto de modelo mental nos ha sido muy útil para estudiar la interacción con los ordenadores entendidos como herramientas de la actividad. Sin embargo, con el cambio de paradigma que supone considerar que ahora colaboramos con máquinas inteligentes tendremos que reconocer que el concepto de modelo mental es insuficiente porque ahora necesitamos que esta representación mental incluya información antropomórfica. Ahora la máquina tiene inteligencia y la inteligencia es una característica humana no de un entorno de trabajo como es la oficina. Los archivos y las carpetas tradicionales no tenían inteligencia y ahora podrían tenerla.

Figura 4: Interacción con un ente inteligente

En el modelo mental de la maquina como herramienta no está incluida la mente de la máquina. La máquina como herramienta no tiene mente como la tiene otra persona con la que colaboramos. Sin embargo, a una máquina inteligente sí le podemos atribuir una mente y, de alguna manera, la antropomorfizamos.

Antropomorfismo es la tendencia a atribuir características (mentales, motivacionales, emocionales, etc.) humanas a objetos no-humanos. No debemos confundir antropomorfismo con animismo que es un fenómeno relacionado con algunas religiones en las que se atribuye un "alma humana" a los objetos (Eliade, 1981). Tampoco estamos hablando del animismo que estudian los psicólogos evolutivos y que constituye una fase del desarrollo psicológico de los niños. Los niños se relacionan con los objetos que los rodean como si tuviesen características humanas. El psicólogo francés J. Piaget (1979) demostró que un criterio para determinar si un niño ha evolucionado correctamente es observar que deja de ser animista y empieza a darse cuenta de que los objetos no tienen mentes, emociones, motivaciones, etc., humanas.

Sin embargo, el antropomorfismo que a nosotros nos interesa aquí es un pensamiento y una forma de actuar consciente y voluntaria que se quiere realizar con algún propósito. Los investigadores de la Universidad de Chicago, Nicholas Epley, Adam Waytz y John T. Cacioppo (2007) han estudiado este fenómeno y han propuesto una teoría que ayuda a entenderlo. El punto de partida de la teoría es la consideración del antropomorfismo como un proceso inductivo que sirve para explicar y predecir la conducta del objeto inanimado con el que se interactúa. Este proceso inductivo se pone en funcionamiento cuando, para interactuar con un objeto de una forma efectiva, activamos el conocimiento que tenemos sobre los seres humanos y, por tanto, sobre nosotros mismos, para que nos ayude a comprender su conducta y predecirla. Por lo tanto, ésta es una teoría que considera al antropomorfismo como un fenómeno que surge por conveniencia. Nos conviene para comprender y predecir la conducta de un ente con el que interactuamos (colaboramos en la actividad). Este antropomorfismo está implícito en la interacción entre personas, pero ahora estará también en la interacción con máquinas inteligentes. La pregunta que debemos hacernos ahora es cómo este antropomorfismo puede y debe ser diseñado.

Tomando como ejemplos de la vida cotidiana de muchas personas en la actualidad, pensemos en las aplicaciones como Siri y Alexa y preguntémonos, para empezar, si las consideramos entes inteligentes y por qué. ¿Pueden estas aplicaciones informáticas comprendernos cuando nos comunicamos con ellas como nos entendería una persona con la que nos comunicamos? ¿podemos nosotros comprender sus respuestas y predecirlas como comprenderíamos y predeciríamos las respuestas de otras personas? Comprender implica comprender el lenguaje natural, nuestras intenciones, nuestras emociones, etc. Los investigadores actuales están preguntandose si cuando hablamos con Siri lo hacemos como cuando hablamos con una persona (Sayago, Neves y Cowan, 2019; Seymour y Van Kleek, 2020). Si esto es así, cabe preguntarse cómo diseñaremos a Siri para que al contestarnos lo haga como lo haría una persona.

Figura 5: Sustitución del concepto de Modelo mental por el concepto de Teoría de la Mente.

Una pregunta relacionada con ésa es: ¿es bueno el antropomorfismo? Pensemos primero en una característica que facilita el antropomorfismo, la "apariencia humana" del objeto con el que interactuamos. Una pregunta que podemos hacernos es si el "aspecto físico antropomórfico" facilita la interacción con los sujetos inteligentes artificiales. En otras palabras, ¿a las personas, sujetos de la actividad le gusta interactuar con un sujeto inteligente que tenga apariencia humana o prefieren que este sujeto tenga una apariencia no antropomórfica? Esta pregunta se la hicieron Velez, Esteban y Cañas (2003) cuando intentaban buscar una explicación de por qué los usuarios rechazaban a los agentes que aparecían en la pantalla de muchas aplicaciones informáticas ofreciendo ayuda, sobre todo después de cometer un error y el agente intenta explicarte como hay que hacerlo bien. Seguro que todos hemos tenido experiencias con estos agentes intentando ayudarnos y nosotros eliminándolos de la interfaz de forma inmediata.

En la investigación que realizaron estos autores, los usuarios ejecutaban una serie de tareas en una hoja de cálculo y en la pantalla le aparecían unos agentes que ofrecían ayuda (ver la Figura 6). Estos agentes podían tener diferentes niveles de apariencia antropomórfica. Los resultados mostraron que los usuarios preferían a los agentes no antropomórficos. La explicación que los investigadores dieron a estos resultados es que cuando nos ofrecen ayuda están de alguna forma evaluándonos y esta evaluación es mejor que nos la haga un agente no antropomórfico que un agente con parecido humano. De alguna manera, como sabemos en Psicología, a las personas no nos gusta que nos evalúen otras personas. Es posible que si los agentes aparecieran en un contexto en el que no nos estuviesen evaluando (por ejemplo, no estuviésemos haciendo una tarea donde podemos cometer errores) aceptásemos mejor sus consejos.

Figura 6: Algunos diseños de agentes utilizados en la investigación de Vélez, Esteban y Cañas (2003)

Sin embargo, también sabemos por investigaciones neuropsicológicas recientes que cuando colaboramos con robots, nuestra actividad cortical se parece más a la actividad cortical de cuando colaboramos con otros seres humanos mientras que el robot tenga más apariencia humana (Krach y col. 2008). Por lo tanto, es posible que la apariencia humana no sea contraproducente en todas las actividades. Por ejemplo, sabemos que, la apariencia humana ayuda o no dependiendo del contexto (Goetz, Kiesler y Power, 2003) En un contexto industrial donde lo que importa es la eficacia técnica, el que el robot tenga una apariencia human importa menos que un contexto de salud donde las relaciones humanas son más importantes. En un contexto de salud el que un robot tenga un aspecto humano es de más ayuda para la colaboración entre paciente y la máquina inteligente. Un robot con apariencia humanoide facilita la empatía y, por tanto, da más confianza y es más aceptado.

El paso del diseño basado en el modelo mental al diseño basado en el antropomorfismo hace referencia al diseño de la interfaz de la máquina. El concepto de modelo mental nos ha sido muy útil para diseñar las interfaces de las máquinas entendidas como herramientas. El concepto de antropomorfismo lo necesitamos ahora para diseñar máquinas inteligentes. Pero podemos pasar a pensar en el diseño de la propia máquina inteligente, en cómo se diseña su inteligencia artificial para que pueda comprender al ser humano. En este sentido tendremos que plantearnos si la máquina inteligente debe tener un modelo de la mente humana.

En Psicología sabemos que los seres humanos adquirimos durante nuestro desarrollo evolutivo en la infancia la capacidad de "leer" la mente de las personas con las que interactuamos. En la tradición de la Ciencia Cognitiva se ha trabajado con la hipótesis de que cada ser humano aprende una "Teoría de la Mente" o lo que es lo mismo, una teoría sobre como la mente humana funciona que nos permite "leer"

(o intuir) lo que otra persona piensa o siente. Este aprendizaje o adquisición de una teoría de la mente es fundamental para el desarrollo normal de cualquier ser humano y se ha demostrado que algunas de las psicopatologías más importantes que aparecen en el desarrollo psicobiológico evolutivo, tales como los desórdenes del espectro autista o la esquizofrenia están relacionadas con una mala adquisición de una buena teoría de la mente (Baron-Cohen, Leslie y Frith, 1985; Langdon, 2005). Pensemos qué sería de nosotros si no fuésemos capaces de inferir las razones de la conducta de otra persona. Por ejemplo, si vemos a alguien abrir un frigorífico en circunstancias normales inferimos que esa persona quiere coger comida o bebida porque tiene hambre o sed. En ningún momento pensamos que está abriendo el frigorífico porque quiere leer un libro.

Existe una creencia muy extendida actualmente entre los investigadores del área que se conoce como "Aprendizaje de Máquina" ("Machine Learning" o ML, en su terminología en inglés) según la cual es posible (algún día) crear un algoritmo al que expongamos a una cantidad de datos suficientes sobre una persona y que pueda inferir las características de la mente de esta persona de tal manera que comprenda y, sobre todo prediga su conducta. Hasta se cree que ese algoritmo podrá modificar la conducta de esa persona. Para ello, el algoritmo deberá "aprender" una teoría de la mente de esa persona (Zuboff, 2020).

Sin embargo, esta creencia puede ser sobre algo que, si materializa algún día, está muy lejano en el tiempo. Imaginemos un algoritmo que observe a una persona mientras camina por la calle. Después de un tiempo determinado, registra datos que muestran que esa persona siempre respeta las señales de tráfico al cruzar la calle. ¿Podremos esperar que ese algoritmo prediga cuándo y en qué condiciones puede saltarse un semáforo? Si la respuesta fuese que sí, eso significará que ese día los especialistas en Aprendizaje de Máquina podrían sustituir a los psicólogos. Es evidente que eso significará que la Psicología desaparecía como profesión completamente. Tendríamos que pensar que la Teoría de la Mente de la máquina inteligente incluiría un conocimiento de psicopatología que la Psicología clínica aún no tiene. Sin embargo, partiendo de esta creencia, éste es el problema con del que se están enfrentando actualmente los ingenieros que están trabajando en el diseño de coches inteligentes: ¿cómo sabes que el conductor del coche que viene en dirección contraria va a hacer un adelantamiento peligroso o el peatón que espera el semáforo en verde no va a decidir cruzar la calle antes de tiempo? La Psicología lleva siglos de investigación sobre este tema sin encontrar una respuesta satisfactoria. Algunos investigadores creen que es posible que los algoritmos de Machine Learning puedan algún día hacer predicciones correctas sobre la conducta humana, pero es evidente que ese día aún no ha llegado.

En conclusión, este ejemplo de la transición desde el estudio del papel que los modelos mentales de las máquinas han tenido en el diseño de la interacción con ellas cuando las consideramos herramientas de la actividad al foco que ahora estamos poniendo en la interacción con entes inteligentes que son al menos "similares" a los seres humanos y pueden comprender y predecir nuestra conducta, supone una modificación del paradigma de la Teoría de la Actividad. Ahora tendremos que pensar en el diseño de la colaboración entre personas y máquinas inteligentes lo cual implica diseñar inteligencia para introducirla en las interfaces de las máquinas inteligentes.

4.2 Las máquinas como herramientas en el trabajo colaborativo

Un segundo ejemplo donde se puede ver como la Inteligencia Artificial introducida en las máquinas supone un cambio en el marco conceptual de la Teoría de la Actividad es el diseño de los sistemas informáticos de apoyo al trabajo cooperativo. Estos sistemas han sido tradicionalmente diseñados para la comunicación y el trabajo colaborativo entre personas a través del ordenador que actúa como medio de transmisión y sirven para dar soporte al trabajo en grupo. Irene Greif y Paul Cashman en un workshop (Cashman y Greif, 1984) acuñaron el término CSCW (Computer Suported Collaborative Work o Apoyo por ordenador al trabajo en equipo en castellano) para denominar a estos sistemas. Con los sistemas CSCW se pretende guiar el pensamiento y el trabajo en grupo. En el diseño de estos sistemas se tiene en cuenta el contexto del trabajo donde se enmarca el trabajador ya que los sistemas orientados hacia grupos de trabajo deben estar diseñados para favorecer los factores organizativos mediante un desarrollo cooperativo que fomente la interoperabilidad (Grudin, 1990).

El diseño CSCW ha sido un área de intensa investigación durante estas últimas décadas ya que cada vez más el trabajo en el mundo laboral se realiza en grupo utilizando ordenadores. Sin embargo, en esta investigación se han planteado muchos problemas que son relevantes cuando pensamos en la introducción de la Inteligencia Artificial en los sistemas CSCW.

En el año 1989, Lian J. Bannon y Kjeld Schmidt (1989) señalaron que uno de los problemas para diseñar estos sistemas era el definir que es un grupo de personas que colaboran para realizar un trabajo. En el esquema de la Teoría de la Actividad, en cada actividad hay un "objetivo", pero en el trabajo colaborativo puede haber varios objetivos, tantos como sujetos existan. Por lo tanto, un problema a resolver es el diseñar cómo los diferentes objetivos individuales se

coordinan para alcanzar un objetivo común. Este problema ha sido estudiado durante muchas décadas por los especialistas en Psicología Organizacional y el diseño de los sistemas CSCW se han beneficiado de los resultados e estas investigaciones. De la misma manera, en el diseño de los sistemas CSCW se ha incorporado la investigación que se ha estado realizando durante décadas sobre los llamados Modelos mentales de grupo. Los investigadores han estado explorando la forma en que grupos de individuos comparten y construyen modelos mentales cuando interactúan en grupos (Orasanu y Salas, 1993). Estos investigadores argumentan que, para poder trabajar juntos de una forma eficiente, los miembros de los grupos deben percibir, codificar, almacenar y recuperar la información de una forma similar. En consecuencia, la calidad de los resultados obtenidos por el grupo dependerá no sólo de la información que tengan disponible los miembros, sino también de la capacidad que tengan éstos para compartir el modelo mental del sistema. Cuando los miembros de un grupo comparten un modelo mental similar y correcto de la interacción en grupo, interactúan y realizan sus tareas más eficazmente (Cannon-Bowers, Salas y Converse, 1993). El concepto de modelo mental de grupo es de una importancia capital en la predicción de la conducta de los equipos de personas. Se ha trabajado hasta ahora con el supuesto de que este modelo mental de grupo es lo que los miembros de un grupo deben desarrollar para realizar sus tareas (Klimoski y Mohammed, 1994).

Cuando este concepto de modelo mental de grupo se propuso para explicar cómo el ordenador se utilizaba en una actividad realizada por un grupo de personas, se pensaba en que este ordenador era una herramienta entendida como se entiende en la Teoría de la Actividad. Los sujetos de la actividad son los seres humanos y el ordenador sólo sirve para comiscarse entre ellos. Las cuestiones que se planteaban a los diseñadores de estos sistemas de CSCW eran como diseñar las interfaces para que éstas reflejasen la organización del trabajo de los miembros humanos del grupo. Por ejemplo, los diseñadores se planteaban cuestiones sobre qué filtros debían diseñar para que las interfaces reflejasen las jerarquías dentro del grupo, quien podía comunicarse con quien, quien podía tener acceso a qué información, etc.

Figura 7: Incorporación de la inteligencia artificial en los sistemas de apoyo al trabajo colaborativo (CSCW)

Sin embargo, ahora se requiere un nuevo enfoque cuando uno de los participantes en el trabajo colaborativo es un ente con Inteligencia Artificial (Shrestha, Ben-Menahem y Von Krogh, 2019). La pregunta que ahora nos tenemos que hacer es cómo diseñamos estos sistemas CSCW cuando uno de los miembros del grupo es un agente inteligente. En este sentido, el primer problema al que nos tenemos que enfrentar es el de diseñar el sistema para que los sujetos humanos de la actividad comprendan las respuestas y las acciones del sujeto artificial como comprenden las respuestas y las acciones de los otros seres humanos del grupo. En este sentido, por ejemplo, es evidente que la solución a este problema tendrá que darse después de que se defina el role del miembro no-humano en el grupo (por ejemplo, ¿es un jefe? ¿es un subordinado?). El segundo problema será, por supuesto, como hemos dicho que ocurrirá en el caso de la colaboración entre un solo ser humano y un solo ser no-humano inteligente, el diseñar este ser no-humano para que comprenda y prediga las respuestas y las acciones de los seres humanos del grupo.

Pero, además, surgirán problemas que no existían, o por lo menos no eran tan evidentes, en los sistemas tradicionales de CSCW. Uno de estos problemas que nos puede servir de ejemplo es el de la ética de la colaboración entre agentes. Cuando los seres humanos colaboramos tenemos que seguir unas reglas éticas que habrá que tener en cuenta también en la colaboración entre seres humanos y seres artificiales inteligentes (Cañas, 2022). Este no era un problema relevante para los diseñadores de los sistemas tradicionales de CSCW puesto que de él ya se ocupaban otros profesionales del diseño de las organizaciones no relacionados con el sistema informático a través del cual se realizaba el trabajo en grupo. Sin embargo, este problema está empezando a recibir mucha atención actualmente ahora que el mismo sistema CSCW incluye otro miembro del grupo, en este caso no-humano aunque inteligente también.

Como ejemplo de este interés cada vez más reciente en estos temas relacionados con el cómo se realiza la colaboración entre seres humanos y no-humanos en grupos, podemos mencionar el estudio que Erik Veitch, Henrikke Dybvik, Martin Steinert y Ole A. Alsos han publicado un artículo reciente (2022) en el que señalan acertadamente que la Inteligencia Artificial mejorará la eficiencia y la eficacia en la toma de decisiones en el transporte marítimo. Sin embargo, también llaman la atención de los diseñadores para que tengan en cuenta cómo se debe diseñar la cooperación entre los seres humanos y los sistemas de inteligencia artificial. En su estudio han entrevistado a un grupo de diseñadores de sistemas automáticos y a otro grupo de navegantes que trabajan a bordo de transbordadores parcialmente automatizados. Al mismo tiempo, también han realizado observaciones de campo a bordo de uno de los transbordadores. Los resultados de estas

entrevistas y de estas observaciones mostraron una discrepancia entre cómo los diseñadores interpretaron la colaboración humano-IA en comparación con la forma en la que los propios navegantes las interpretaban. Los navegantes veían sobre su papel como uno de "respaldo", definido como la toma de control ad-hoc de la automatización. Sin embargo, los diseñadores colocaron a los navegantes "en el bucle" de un sistema de control más grande, pero descartando el papel de las habilidades in situ de los navegantes y su papel en la toma de decisiones heurística en todas las acciones de adquisición, excepto en las más controladas. Estas discrepancias en las interpretaciones del papel de la Inteligencia Artificial en la cooperación en grupo llevaron a los investigadores a sugerir que el diseño de los componentes inteligentes del grupo debe hacerse de tal manera que las acciones del sistema artificial sean más visibles (computacionalmente más visibles) y que se incorpore en el diseño de todo el sistema señales sociales que articulen el trabajo humano en su entorno natural.

En pocas palabras, podemos decir que los temas que están surgiendo en el diseño de sistemas de CSCW cuando ahora alguno de los componentes del grupo es un ser no-humano inteligente indican en la misma dirección que lo que hemos visto en el caso del role de los modelos mentales en el diseño de interfaces ahora que interactuamos con seres no-humanos inteligentes. En ambos casos, todo indica que tendremos que empezar a considerar todo lo que sabemos sobre cómo los seres humanos colaboramos unos con otros.

5. Conclusiones

La conclusión más importante a la que debemos llegar es que, debido a la introducción de la inteligencia artificial en las máquinas, ahora vamos hacia el diseño de interfaces donde los sistemas inteligentes “colaboran” con los usuarios humanos. Por tanto, las cuestiones de diseño de interfaces tienen que plantearse ahora pensando en la facilitación de esta colaboración (Endsley, Cooke, McNeese, Bisantz, Militello y Roth, (2022). En la tradición de IPO (HCI), lo importante ha sido el diseño de la interfaz para que el usuario reciba la información del ordenador de forma que la pueda percibir, interpretar y memorizar mejor, al mismo tiempo que la interfaz tenga los componentes necesarios para que el usuario pueda comunicarle al ordenador sus órdenes de una forma eficaz, eficiente y satisfactoria. Sin embargo, aunque estos temas no dejarán de ser importantes porque siempre será necesario que los tengamos en cuenta en el diseño de interfaces, ahora tendremos que abordar temas que llevan siendo abordados durante muchos años por la psicología. Tendremos que saber cómo se diseña una interfaz “transparente” y “explicable” para saber por qué la máquina se comporta de la manera que lo hace, cuáles son sus intenciones y cuál será su conducta futura. Tendremos que saber cómo se diseña para mejorar la confianza entre agentes que colaboran en la actividad. Por tanto, podemos decir que nunca ha sido más importante la Psicología para el diseño de la interacción que ahora.

Pero antes de terminar debemos hacer una advertencia. Es posible que alguien piense que estas técnicas pueden usarse para explicar y predecir la conducta humana sin necesidad de recurrir al conocimiento que ya tiene la psicología. Es posible que eso pueda ser cierto, pero también es posible que no. Al fin y al cabo, las técnicas de ML no buscan nada más que ocurrencias (correlaciones) entre eventos. Sin embargo, a todos los que hemos estudiado estadística nos han enseñado que si medimos el tamaño del pie y la inteligencia de una muestra de personas encontraremos una correlación positiva entre ambas variables y no por eso podemos utilizar el tamaño del pie para medir la inteligencia de una persona. Aunque este ejemplo pueda considerarse una exageración, puede servirme para llamar la atención sobre los peligros de crear unas expectativas excesivas en la ML. Si esas expectativas se cumpliesen algún día, los psicólogos y nuestra investigación no serían necesarios y la conducta humana sería finalmente explicable y, sobre todo predecible como se predice en algunas distopías. Sin embargo, aunque la Inteligencia Artificial está dando pasos de gigante asumiendo que esto será posible algún dia, lo cierto es que estamos muy lejos de explicar y predecir la conducta humana usando algoritmos de ML.

Referencias

Baron-Cohen S, Leslie AM, Frith U. (1985). Does the autistic child have a ‘‘theory of mind’’? *Cognition*, 21:37–46. doi:10.1016/0010-0277(85)90022-8.

Benyon, D. (2019). *Designing user experience*. Pearson UK.

Bannon, L. J., & Schmidt, K. (1989). CSCW: Four characters in search of a context. In *ECSCW 1989: Proceedings of the First European Conference on Computer Supported Cooperative Work. Computer Sciences Company*, London.

Cannon-Bowers, J.A., Salas, E., y Converse, S. (1993). Shared mental models in expert team decision making. En N.J. Castellan, Jr. (Ed) *Individual and group decision making*. Hillsdale, NJ: Erlbaum.

Cañas, J. (2022). AI and Ethics When Human Beings Collaborate With AI Agents. *Frontiers in Psychology*, 13.

Cañas, J.J., Bajo, M.T. y Gonzalvo, P. (1994). Mental Models and computer programming. *International Journal of Human-Computer Studies*, 40, 795-811.

Cashman, Paul M.; and Irene Greif (eds) (1984). *Workshop on Computer-Supported Cooperative Work, 13-15* August 1984, Endicott House, Dedham, Ma.

Eliade M (1981) *A history of religious ideas, volume 1: from the stone age to the Eleusinian mysteries*. University of Chicago Press.

Endsley, M. R., Cooke, N., McNeese, N., Bisantz, A., Militello, L., & Roth, E. (2022). Special issue on human-AI teaming and special issue on AI in healthcare. *Journal of Cognitive Engineering and Decision Making*, 0(0) doi:10.1177/15553434221133288

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological review*, 114(4), 864.

Goetz J, Kiesler S, Powers A (2003) Matching robot appearance and behavior to tasks to improve human–robot cooperation. In: *Proceedings of the 12th IEEE workshop on robot and human interactive communication*, pp 55–60.

Grudin, J. (1990). *Groupware and Cooperative Work: Problems and Prospect. The Art of Human Computer Interface*. Addison-Wesley.

Hollnagel, E., & Bye, A. (2000). Principles for modelling function allocation. *International Journal of Human-Computer Studies*, 52(2), 253-265. doi:10.1006/ijhc.1999.0288.

Kieras, D.E., y Bovair, S. (1984) The role of mental model in learning to operate a device. *Cognitive Scienc*e, 8, 255-273.

Klimoski, R. y Mohammed, S. (1994). Team Mental Model: Construct or Metaphor?. *Journal of Management*, 20, 2, 403-437.

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. PloS one, 3(7), e2597.

Langdon R. *Theory of mind in schizophrenia. In: Malle BF, Hodges SD, eds. Other Minds*. New York: Guilford Press; 2005, 323–342.

Leontiev, Aleksei N (1977). *Actividad, Conciencia y Personalidad*. Digitales Soyuz.

Moran, T.P. (1981) An applied psychology of the user. *Computing Surveys*, 13, 1-11.

Nardi, B. A. (1996). Activity theory and human-computer interaction. En B.A. Nardi Context and consciousness: Activity theory and human-computer interaction. Cambridge, MA: MIT Press.

Navarro, R., y Cañas, J.J. (2001).Are visual programming languages better? The role of imagery in program comprehension. *International Journal of Human-Computer Studies*, 54, 799-829.

Norman, D. A., & Draper, S. W. (1986). *User centered system design: New perspectives on human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Orasanu, J., y Salas, E. (1993). Team decision making in complex environments. En G..A. Klein, J. Orasanu, R. Calderwood, y C.E. Zsambok (Eds.) *Decision making in action: Mental models and methods*. Norwood, NJ: Ablex.

Piaget, J. (1979). El niño y la construcción del mundo: El animismo infantil. In *Textos de psicología del niño y del adolescente* (pp. 158-161). Narcea.

Sayago, S., Neves, B., & Cowan, B. (Aug 22, 2019). Voice assistants and older people. In *Proceedings of the 1st International Conference on Conversational User Interfaces* (pp. 1-3). doi:10.1145/3342775.3342803 Retrieved from http://dl.acm.org/citation.cfm?id=3342803

Shrestha, Y. R., Ben-Menahem, S. M., & Von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, 61(4), 66-83.

Seymour, W., & Van Kleek, M. (Apr 25, 2020). Does siri have a soul? exploring voice assistants through shinto design fictions. En *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Saariluoma, P., Cañas, J. J., & Leikas, J. (2016). *Designing for life: A human perspective on technology development*. Springer.

Shneiderman, B. (1997), Direct manipulation for comprehensible, predictable and controllable user interfaces. En *Proceedings of the 2nd international conference on Intelligent user interface*s (pp. 33-39).

Veitch, E., Dybvik, H., Steinert, M., & Alsos, O. A. (2022). Collaborative work with highly automated marine navigation systems Systems. *Computer Supported Cooperative Work (CSCW*). 1-32. doi:10.1007/s10606-022-09450-7

Velez, M., Esteban, E., and Cañas, J.J. (2003) Antropomorphic characteristics of interface agents. En *Proceeding of Human-Computer Interaccion.* Creta, Grecia.

Zuboff, S. (2020). *La era del capitalismo de la vigilancia*. Ediciones Paidós.