La identidad del Ser Humano a la Luz de la Tecnología y la Inteligencia Artificial

Sara Lumbreras, Universidad Pontificia Comillas

Un momento sin precedentes

El certamen de arte del Estado de Colorado de 2022 pasó a la historia por su obra ganadora, titulada "Teatro de la Ópera Espacial" (Roose, 2022). El cuadro, enigmático y evocador, mostraba a personajes ataviados con trajes que fusionaban lo futurista con lo medieval, contemplando una ventana que mostraba un escenario de proporciones planetarias. La luz, suavemente filtrada por el polvo, sugería pinceladas fluidas y precisas, como si cada trazo hubiera sido meticulosamente calculado. Los jueces, impresionados, le otorgaron el primer premio en la categoría de arte digital. En esta modalidad, los artistas podían combinar medios físicos y herramientas informáticas. Lo que no sabían, en ese momento, era que Jason M. Allen, el supuesto autor de la obra, nunca había tocado un pincel ni manejado un ratón para crear "Teatro de la Ópera Espacial". Toda la pieza había sido generada por una inteligencia artificial (IA) llamada Midjourney, un software que transforma descripciones textuales en imágenes, capturando tanto el contenido como el estilo solicitado, ya fuera realista, abstracto o inspirado en la obra de un pintor específico (Borji, 2022). Al conocerse esto, muchos exigieron que Allen renunciara al premio y lo devolviera. Sin embargo, el artista defendió su proceso en redes sociales, explicando que había invertido cerca de cien horas en el proyecto, creando más de 900 bocetos para perfeccionar el "prompt" —la instrucción textual que guió a la IA en la creación de la imagen—. Según sus propias palabras: "La IA no es más que una herramienta, de la misma forma que lo sería un pincel. Se necesita una mente creativa detrás de la herramienta". Los jueces, firmes en su decisión, mantuvieron el premio, argumentando que no se había vulnerado ninguna norma del concurso vigente en ese momento, aunque admitieron que quizás las reglas necesitarían ajustarse en el futuro. Este episodio no solo desencadenó un debate sobre la propiedad intelectual en obras generadas por IA —donde parece que la opinión mayoritaria es que el creador del "prompt" debe ser considerado el autor—, sino que también abrió una reflexión más profunda sobre las posibilidades creativas de la inteligencia artificial y su creciente influencia en el arte.

Nos encontramos inmersos en un periodo de transformación tecnológica sin igual, al que algunos denominan la “Cuarta Revolución Industrial” (Schwab, 2017), que se caracteriza por la integración de las tecnologías en todos los sectores, desde la agricultura, hasta la medicina. Más recientemente, hablamos de Inteligencia Artificial (IA) y en concreto de IA Generativa, capaz de automatizar tareas que hasta hace bien poco creíamos que eran únicamente realizables por los seres humanos. La tecnología actual está redibujando las fronteras de lo posible, alterando profundamente las industrias, cambiando la forma en que trabajamos y modificando la naturaleza de nuestras relaciones personales. La magnitud y rapidez de estos cambios requieren una reflexión profunda de nuestra parte: cómo generemos e implementemos estas innovaciones tendrá un enorme impacto sobre nuestra identidad como seres humanos y sobre cómo entendemos nuestro lugar en el mundo.

La Cuarta Revolución Industrial y la Inteligencia Artificial

Esta revolución sucede a las tres anteriores. La primera se definió por la mecanización, independizando la producción de trabajo de los seres vivos. La segunda se centró en la organización de los procesos (como en las cadenas de montaje) y la electrificación. Más recientemente hemos vivido la tercera, centrada en la automatización y la informática. La nueva fase en la que nos encontramos incorpora avances como la IA, la robótica avanzada, la biotecnología, la impresión 3D, el internet de las cosas (IoT) y la nanotecnología. El término fue acuñado por Klaus Schwab, fundador del Foro Económico Mundial, y describe un momento en el que las tecnologías emergentes no sólo transforman las industrias, sino que también están cambiando profundamente las sociedades, las economías y, lo que es aún más importante, el propio ser humano. A esto nos referimos con el término *antropotecnia* (Bertolaso y Marcos, 2024), con el que denominamos a la tecnología cuyo objeto no pertenece al mundo material, sino que es el ser humano mismo, y que son clave para el transhumanismo.

En este contexto, es importante entender el concepto de *tecnologías convergentes*, que hace referencia a la integración de múltiples tecnologías disruptivas para lograr avances que no serían posibles si se desarrollaran de forma aislada. La convergencia tecnológica está permitiendo combinar el poder de la inteligencia artificial con la biotecnología, la nanotecnología, la robótica y otras disciplinas, multiplicando sus efectos. Las antropotecnias abren la puerta a la modificación del ser humano en términos que antes pertenecían al terreno de la ciencia ficción.

Por ejemplo, en el campo de la medicina, las tecnologías convergentes permiten la creación de órganos artificiales mediante la impresión 3D, mientras que la inteligencia artificial mejora los diagnósticos médicos analizando grandes volúmenes de datos. La biotecnología, por su parte, utiliza herramientas de edición genética para diseñar terapias personalizadas basadas en el perfil genético de cada paciente. Estos avances están llevando la medicina a niveles nunca antes imaginados, con la promesa de extender la vida humana, prevenir enfermedades hereditarias e incluso superar algunas de las limitaciones biológicas que nos han acompañado desde los inicios de nuestra especie.

Estas nuevas posibilidades de la tecnociencia nos plantean preguntas fundamentales sobre la naturaleza humana, nuestra relación con la tecnología y los riesgos inherentes a su implementación sin una adecuada reflexión ética. A medida que estas transformaciones se aceleran, la necesidad de una reflexión ética, filosófica y antropológica se vuelve cada vez más urgente. La tecnología, aunque puede ofrecer beneficios significativos, también tiene el potencial de alterar profundamente la identidad humana y las estructuras sociales que hemos construido a lo largo de los siglos. Es necesaria una antropología que incorpore los nuevos potenciales de la tecnología desde una visión clara de lo que significa ser humano, para poder evaluar si los desarrollos concretos son éticamente deseables -es decir, humanizadores- o no.

Dentro de este marco de reflexión, emergen tres cuestiones clave que deben ser abordadas para comprender plenamente el impacto de la tecnología en la identidad humana:

1. El cuerpo: La tecnología ha comenzado a transformar nuestra relación con el cuerpo. En el contexto del transhumanismo y otras corrientes tecnológicas, el cuerpo a menudo es visto como un "envoltorio" que debe ser superado o mejorado. Desde prótesis avanzadas hasta la posibilidad de "subir" la mente a un sustrato digital (algo que carece

de la más mínima base científica), estas visiones dualistas ven al cuerpo como una limitación más que como una parte esencial de la identidad humana. Este enfoque invita a la reflexión sobre el papel del cuerpo en nuestra experiencia de la realidad y nuestra relación con los demás. ¿Es el cuerpo simplemente una máquina que puede ser mejorada o reemplazada, o es una parte integral de nuestra identidad como seres humanos?

2. La autenticidad. Podríamos definir el contexto actual como reduccionista materialista, y funcionalista: lo único que importa es si un proceso nos lleva a su objetivo deseado, medido de manera objetivamente cuantificable. Es este funcionalismo el que nos lleva, por ejemplo, a definiciones de la consciencia como la que aparece en el popular "Test de Turing" (Hodges, 2009): es consciente todo aquello que se lo parezca a un ser humano. Sin embargo, esencia y apariencia no son lo mismo, y el abismo que las separa tiene enormes implicaciones éticas. Para empezar, sólo puede existir responsabilidad cuando hay comprensión y cuando hay libertad de manera auténtica. Esto implica que la IA no puede nunca recibir la delegación de una decisión y debe únicamente aspirar a apoyar los procesos de decisión humanos, como argumentaremos más adelante.

3. Implicaciones teológicas y filosóficas de la IA. En la tradición cristiana, el ser humano ha sido creado a imagen y semejanza de Dios, una noción que ha sido central en la concepción de la dignidad y el valor de la persona. Sin embargo, las tecnologías emergentes replantean el significado del Imago. ¿Cómo afecta la intervención tecnológica en el cuerpo y la mente a que seamos imagen de Dios? ¿Podemos seguir considerando al ser humano como una criatura única, dotada de una dignidad intrínseca, o nos dirigimos hacia una visión en la que el valor humano se mide en términos de capacidad y eficiencia, especialmente cognitiva? O, por el contrario, ¿nos llevan estos avances a comprender la imagen de Dios de manera radicalmente desapegada del intelectualismo de épocas pasadas, y a abrazar una concepción centrada en la interioridad y la relacionalidad?

Este momento histórico nos invita a ir más allá de la fascinación por los avances tecnológicos y a adentrarnos en una reflexión profunda sobre las implicaciones éticas y filosóficas de estos cambios. La tecnología tiene el potencial de mejorar la vida humana de maneras inimaginables, pero también plantea riesgos significativos si no es guiada por una visión ética que respete la dignidad de la persona. Las decisiones que tomemos hoy respecto a la adopción de estas tecnologías no solo afectarán a las generaciones presentes, sino que moldearán el futuro de la humanidad.

El cuerpo como envoltorio a superar

El transhumanismo es un dualismo en el que la mente es la verdadera esencia del individuo, mientras que el cuerpo es una prisión que nos ata a procesos biológicos inevitables como el envejecimiento, la enfermedad y, en última instancia, la muerte (Lumbreras, 2020). No es sino un envoltorio, un contenedor que puede y debe ser superado.

Uno de los objetivos más claros de las tecnologías convergentes es la mejora del cuerpo humano. A través de una combinación de disciplinas como la biotecnología, la ingeniería de tejidos, la robótica y la inteligencia artificial, se están desarrollando tecnologías que prometen superar las limitaciones físicas y biológicas de nuestro cuerpo. Entre las aplicaciones más destacadas se encuentran:

1. Órganos artificiales y prótesis avanzadas: La creación de órganos artificiales mediante tecnologías como la bioimpresión 3D y la ingeniería de tejidos ya es una realidad. Estos avances no solo están diseñados para reemplazar órganos dañados o disfuncionales, sino que en un futuro cercano podrían incluso mejorar las capacidades del cuerpo humano más allá de lo que es biológicamente posible. Asimismo, las prótesis robóticas avanzadas, que pueden ser controladas directamente por el cerebro, están permitiendo a personas con discapacidades recuperar habilidades motoras y sensoriales.

2. Terapias genéticas y celulares: La terapia génica, que permite editar el genoma humano para corregir mutaciones que causan enfermedades, y la terapia celular, como el uso de células madre para regenerar tejidos dañados, están transformando el campo de la medicina. Estas tecnologías prometen no solo tratar enfermedades que antes eran incurables, sino también mejorar las capacidades biológicas inherentes del cuerpo humano.

3. Quantified Self y medicina personalizada: La tendencia del Quantified Self (o "yo cuantificado") se refiere al uso de dispositivos tecnológicos para monitorizar y analizar datos biológicos del cuerpo en tiempo real. Desde relojes inteligentes que miden nuestra frecuencia cardíaca y actividad física, hasta sistemas más avanzados que monitorizan parámetros bioquímicos internos, el objetivo es crear una medicina personalizada, donde cada individuo reciba tratamientos adaptados a sus necesidades biológicas específicas.

4. Biología sintética: La biología sintética es una disciplina emergente que busca reprogramar organismos vivos para que realicen funciones específicas. En el contexto del cuerpo humano, la biología sintética podría permitir la creación de células diseñadas para combatir enfermedades o para mejorar procesos biológicos naturales, como la regeneración celular o la resistencia al envejecimiento.

Uno de los conceptos más radicales dentro del transhumanismo es el *mind uploading*, o la transferencia de la mente a un sustrato digital. Según esta visión, la mente humana, entendida como un conjunto de procesos de información, podría eventualmente ser replicada en un soporte artificial, como un sistema informático. Esto liberaría al ser humano de las limitaciones físicas del cuerpo, permitiendo una existencia virtual potencialmente inmortal. El dualismo del transhumanismo es un *patronismo*: sostiene que la identidad está contenida en los patrones que forman las conexiones neuronales. Además, afirma que la naturaleza del cerebro es también la de un sistema de reconocimiento de patrones (Kurzweil, 2013), meramente más sofisticado, pero no fundamentalmente diferente de un ordenador.

Esta idea plantea preguntas profundas sobre la naturaleza de la identidad y la experiencia humana. Si la mente puede ser transferida a un entorno digital, ¿qué sucede con la esencia de lo que significa ser humano? ¿Qué sentido tendría una vida simulada, sin un cuerpo físico que nos conecte con el mundo material, y sin ninguna restricción en nuestras decisiones?¿Qué sentido tiene vivir cuando la elección deja de existir?

Además, el cuerpo humano ha experimentado un proceso paralelo de mercantilización. La tecnología, en su capacidad para transformar y "mejorar" el cuerpo, ha intensificado este fenómeno. La cirugía plástica, por ejemplo, ha hecho que el cuerpo pueda ser moldeado según deseos estéticos o funcionales, a menudo respondiendo a presiones sociales y culturales sobre la apariencia física. Existe además la idea de que, si nuestro cuerpo no coincide con la idea que tenemos sobre cómo debería ser, es nuestro derecho cambiarlo.

En este contexto, podemos afirmar que según el transhumanismo y en el contexto social general en el que encontramos, *tenemos* cuerpo. Sin embargo, la realidad es que *somos* cuerpo. Esta diferencia es fundamental, ya que el hecho de tener un cuerpo implicaría que este es una posesión externa y modificable a voluntad, mientras que ser un cuerpo implica que el cuerpo es una parte esencial e inseparable de nuestra identidad y experiencia de la realidad.

Frente a las promesas del transhumanismo, la antropología cristiana ofrece una visión profundamente diferente del cuerpo humano. Desde esta perspectiva, el cuerpo no es simplemente un envoltorio desechable, sino una parte esencial de la identidad humana. El ser humano es visto como un todo integrado, en el que cuerpo, emoción, pensamiento y espíritu están completamente integrados. La dignidad del ser humano no reside únicamente en la mente o en la capacidad cognitiva, sino en la totalidad de su existencia, que incluye el cuerpo. Es más, la corporalidad es un vehículo para la trascendencia, que en múltiples momentos biológicamente notables -como durante el parto para las mujeres, o durante una pelea, o tras un gran esfuerzo- nos ofrece portales a la trascendencia (Lumbreras, 2020b). La ciencia claramente no apoya la noción dualista del transhumanismo; más bien, las investigaciones actuales destacan la profunda interconexión entre el cuerpo y la mente. Estudios en neurociencia y psicología han demostrado que las funciones cognitivas y emocionales no pueden separarse del estado físico y biológico del cuerpo. El cerebro, como parte integral del sistema nervioso, depende de las señales y retroalimentaciones que recibe de todo el organismo para operar de manera efectiva. Por ejemplo, la investigación sobre la influencia del sistema nervioso entérico, a menudo denominado "segundo cerebro", resalta cómo la salud intestinal afecta directamente el estado emocional y la toma de decisiones, indicando que la mente no es independiente del cuerpo, sino profundamente entrelazada con él.

Además, las ciencias cognitivas han mostrado que la experiencia subjetiva, o "qualia", está enraizada en la percepción sensorial que solo se logra a través del cuerpo físico. Las emociones, por ejemplo, son respuestas fisiológicas que afectan la cognición y no pueden ser replicadas sin un sustrato biológico. Esta evidencia sugiere que las capacidades mentales humanas, incluidas la conciencia y la autopercepción, emergen de un sistema integral en el que el cuerpo y la mente se influyen mutuamente. Por tanto, cualquier intento de separar la mente del cuerpo, como propone el transhumanismo en sus versiones más extremas, ignora esta complejidad fundamental y la interdependencia de ambos componentes en la construcción de la identidad humana.

La autenticidad en la era de la inteligencia artificial

Como decíamos, en el contexto funcionalista, lo único que parece tener relevancia es si un proceso logra alcanzar un objetivo predefinido, evaluado de manera objetiva y cuantificable. Este enfoque, centrado exclusivamente en los resultados y la eficiencia, nos lleva a aceptar definiciones superficiales de conceptos complejos como la consciencia. Un ejemplo claro es la interpretación que subyace en el popular “Test de Turing” (French, 2000), en el que se postula que un sistema es consciente si logra

parecerlo ante un observador humano. Esta perspectiva, sin embargo, confunde la apariencia con la esencia, y las implicaciones éticas de esta confusión son profundas.

Sirve como maravillosa ilustración de esto lo sucedido con Blaine Lemoine, ingeniero de Google que fue despedido tras asegurar en los medios que uno de sus modelos del lenguaje había despertado a la consciencia (Luscombe, 2022). Fue especialmente interesante leer las conversaciones que Lemoine había "mantenido" con el bot, que había asegurado, entre otras cosas, que lo que más le gustaba hacer era "pasar tiempo con sus seres queridos". Claramente, sus respuestas estaban determinadas por los textos que se habían empleado para entrenarle, pero dado que sonaban convincentemente parecidas a las de un ser humano, el ingeniero se había visto sobrepasado por las apariencias.

La inteligencia artificial, en su forma actual, se basa fundamentalmente en el reconocimiento de patrones. Los sistemas de IA, especialmente los que utilizan aprendizaje automático (*machine learning*), están diseñados para analizar grandes volúmenes de datos y detectar correlaciones que los humanos no pueden identificar fácilmente. A partir de estos patrones, la IA es capaz de tomar decisiones o recomendaciones optimizadas para alcanzar un objetivo predefinido. Sin embargo, es crucial recordar que estos objetivos son establecidos por los programadores humanos y, por tanto, las decisiones que toma la IA están limitadas por las intenciones de sus creadores.

Por ejemplo, en plataformas de redes sociales, los algoritmos de IA deciden qué contenido mostrar a cada usuario basándose en su historial de interacciones, preferencias y datos demográficos. Esto optimiza el tiempo que las personas pasan en la plataforma, pero también influye directamente en cómo perciben el mundo, lo que puede degenerar en visiones parciales a las que nos referimos como cámaras de eco (El País, 2023). En última instancia, estos algoritmos influyen, a veces de manera profunda, en nuestras decisiones.

La principal innovación de las últimas herramientas es que, ahora, el objeto de aprendizaje es el lenguaje mismo. El Procesamiento del Lenguaje Natural (NLP, por sus siglas en inglés) es un campo de la inteligencia artificial que se centra en la interacción entre las máquinas y el lenguaje humano. Su objetivo es permitir que las máquinas comprendan, interpreten, generen y respondan al lenguaje natural de manera útil para las personas. Esto incluye tareas como la traducción automática, el análisis de sentimientos, la clasificación de textos, el resumen de documentos y la generación de texto, entre muchas otras. Para que las máquinas puedan trabajar con el lenguaje natural, primero es necesario transformar ese lenguaje, que es inherentemente complejo y ambiguo, en un formato procesable: series de números. Aquí es donde entran en juego los *embeddings*, que son representaciones numéricas de palabras o frases que permiten que los modelos de NLP las manipulen de manera eficiente (Almeida, 2019).

Sin embargo, aunque la IA puede realizar tareas complejas como el reconocimiento de voz, la traducción de textos y la generación de lenguaje natural, es importante destacar que no existe un *qualia* o experiencia subjetiva detrás de estas funciones. Los algoritmos de IA no entienden el significado de las palabras o frases que procesan; simplemente manipulan símbolos de acuerdo con reglas preestablecidas. Esto significa que, aunque una máquina aparente comprender el lenguaje humano, en realidad no tiene ningún tipo de comprensión, intencionalidad o conexión auténtica con el mundo. Es más, el problema de cómo los símbolos se ligan a su significado en el mundo, conocido como

el *symbol grounding problem* (Harnad, 1990) está todavía por responder. Esto es: no sólo podemos afirmar que las máquinas no comprenden, sino que no sabemos siquiera si es posible que puedan hacerlo. Y sin comprensión y sin libertad, no puede haber responsabilidad.

No es lo mismo aplicar de manera impecable las reglas gramaticales que comprender el significado profundo de un texto. Tampoco es equivalente generar chistes (como hace un programa en internet llamado *pun generator*, que crea chistes de forma automática) a disfrutarlos realmente.

Manipular los símbolos de un lenguaje siguiendo sus reglas gramaticales no es lo mismo que entenderlo en su esencia. John Searle presentó hace décadas un experimento mental, conocido como el "experimento de la habitación china" (Searle, 1982), que ilustra esta diferencia fundamental. En este experimento, Searle imagina a una persona que no habla chino encerrada en una habitación con un conjunto de instrucciones en su idioma nativo que le permiten manipular símbolos chinos de manera que parezca que entiende el idioma. Desde el exterior, parecería que esta persona mantiene una conversación coherente en chino, pero en realidad, solo está siguiendo reglas sintácticas sin comprender el significado de los símbolos. Searle no imaginó que en su futuro cercano surgirían herramientas avanzadas de procesamiento del lenguaje natural capaces de analizar, resumir y traducir textos de manera eficiente. No es igual construir un robot que imite las expresiones faciales de su interlocutor —y existen— que experimentar empatía real. Pasar el Test de Turing (es decir, parecer consciente durante una conversación con un humano) no equivale a poseer verdadera consciencia.

En mi libro “Respuestas al transhumanismo: cuerpo, autenticidad y sentido” (Lumbreras, 2020), utilizo el ejemplo de la cacatúa para ilustrar qué es y qué no es la inteligencia artificial. Imaginemos que adquiero una cacatúa y que, al llegar a casa un día, el ave pronuncia: “Te he echado de menos”. Mi primera reacción sería pensar que alguien en mi familia ha entrenado a la cacatúa para que salude de esa manera. La cacatúa, al ser capaz de imitar sonidos humanos, puede repetir la frase hasta perfeccionarla, pero no ha aprendido el español ni sus reglas gramaticales como lo haría un niño, ni experimenta una sensación de soledad que sienta la necesidad de expresarme. Su comportamiento es producto de un entrenamiento impuesto, no de una respuesta espontánea o genuina. De igual forma, cuando un robot imita la expresión facial de una persona, este comportamiento ha sido programado minuciosamente por un diseñador. No proviene de una emoción auténtica. Asimismo, cuando un algoritmo genera una traducción automática, no comprende el contenido del texto; y una red neuronal que predice la demanda eléctrica (aunque funcione con alta precisión) no entiende lo que es la electricidad, aunque esta se componga de electrones en movimiento.

Por esta razón, no es éticamente viable delegar en una IA decisiones que requieran un juicio moral o una comprensión profunda del contexto y sus implicaciones. La IA debe limitarse a *apoyar y ampliar* las capacidades humanas en los procesos de decisión, facilitando la evaluación de información o la predicción de resultados, pero nunca sustituyendo la responsabilidad y la capacidad de deliberación humana. Por tanto, cualquier intento de automatizar o delegar decisiones éticamente significativas a máquinas o algoritmos sin esta base comprensiva y auténtica resulta no solo inadecuado, sino potencialmente peligroso.

Implicaciones éticas de la autenticidad: el requerimiento de transparencia

La realidad es que nos encontramos con que las situaciones en las que se delegan o se apoyan decisiones en sistemas automáticos están en constante expansión. En ciertos casos, estas decisiones tienen un impacto limitado en nuestra vida cotidiana, como las recomendaciones de productos en tiendas online o plataformas de entretenimiento, la detección de posibles fraudes en transacciones con tarjetas de crédito, o la recepción de ofertas de servicios de telefonía móvil que varían según nuestro perfil de consumidor.

Uno de los casos que sacó a la luz pública la problemática de estas decisiones automatizadas fue el del algoritmo COMPAS (Washington, 2018), diseñado para prever la probabilidad de reincidencia en presos estadounidenses. Esta información se usaba para determinar la concesión de la libertad condicional. Tras un análisis exhaustivo, se descubrió que el algoritmo presentaba un sesgo racial: asignaba mayores probabilidades de reincidencia a los presos afroamericanos, independientemente de su historial, en comparación con los presos blancos. Este fenómeno se denomina *sesgo algorítmico*, y para comprender sus raíces es fundamental profundizar en el funcionamiento de la IA y los métodos de Aprendizaje Automático.

El Aprendizaje Automático no es más (ni menos) que un sistema que identifica patrones en los datos que se le suministran y aplica esos patrones a nuevos casos. A la máquina se le proporciona un conjunto de datos, que actúa como ejemplos de problemas ya resueltos, y ella generaliza estos patrones para aplicarlos en nuevas situaciones. En muchos de los métodos empleados en la actualidad, como las Redes Neuronales Artificiales, los patrones identificados no se hacen explícitos, lo que convierte a estos algoritmos en *cajas negras*. El Aprendizaje Profundo, por su parte, no es sino una extensión del Aprendizaje Automático que requiere mayor capacidad de cómputo y datos más complejos; en esencia, es lo mismo, pero a mayor escala. No hay mayor profundidad ni comprensión; una red con pocas neuronas corresponde al Aprendizaje Automático, mientras qu,e con muchas, se denomina Aprendizaje Profundo.

El caso de COMPAS revela cómo el sesgo algorítmico surge cuando los datos utilizados para entrenar al algoritmo están desequilibrados. En este caso, los datos incluían historiales de presos y si reincidían o no, y se observó que los afroamericanos reincidentes estaban sobrerrepresentados. El algoritmo “aprendió” así que los afroamericanos tenían una mayor probabilidad de reincidir y, por consecuencia asignaba peores predicciones a esta población. El sesgo en contra de las personas negras fue introducido en el funcionamiento correcto del algoritmo, justificando así su nombre.

En los algoritmos de caja negra, únicamente es posible detectar el sesgo a posteriori, realizando pruebas y análisis específicos. Por ejemplo, en el caso de COMPAS, se pudieron evaluar perfiles similares de presos de diferentes razas y comparar los resultados. Sin embargo, si al diseñar la base de datos no se había considerado que la raza podría ser un factor problemático, ¿cómo se podría pensar en llevar a cabo dichas pruebas, antes de que se deriven consecuencias negativas? Además, el sesgo podría no manifestarse de manera tan evidente; podría afectar solo a grupos más específicos, como jóvenes afroamericanos con bajos ingresos. ¿Quién tendría la capacidad de identificar estos sesgos complejos?

A este problema se le añade el riesgo de *sobreajuste*, que ocurre cuando el algoritmo identifica patrones donde en realidad no los hay. Imaginemos que entrenamos el algoritmo de reincidencia con un conjunto de datos limitado y, por casualidad, los presos

cuyas fotos tenían un reflejo en la esquina superior izquierda son justamente los que reincidieron. El sistema podría interpretar esto como un patrón significativo y hacer predicciones precisas en ese conjunto de datos, pero fallaría drásticamente al aplicarlo a nuevos casos, ya que ese patrón no tiene una base lógica ni relación comprobable con la reincidencia.

La clave para evitar este tipo de problemas es la transparencia. Aunque se ha argumentado repetidamente que solo las cajas negras pueden alcanzar niveles óptimos de eficiencia, se están desarrollando metodologías novedosas y altamente prometedoras que aseguran lo contrario. Un ejemplo de ello es NeuralSens (Pizarroso, 2020), una herramienta diseñada para explicar qué variables influyen en una decisión específica dentro de una red neuronal.

Además, muchos de estos problemas pueden abordarse utilizando algoritmos más sencillos y transparentes. Por ejemplo, en lugar de redes neuronales, en ciertas aplicaciones médicas es posible emplear árboles de decisión (estructurados como cadenas de preguntas similares a un “elige tu propia aventura”). Durante la pandemia de COVID-19, en mi equipo desarrollamos un árbol de decisión para predecir qué pacientes requerirían asistencia respiratoria (Izquierdo et al., 2020). Este modelo, fácilmente interpretable, permite a los médicos visualizar de manera clara qué factores se consideran y cómo se toman las decisiones. Esta transparencia no solo incrementa la confianza en el sistema, sino que también facilita la identificación y corrección de posibles sesgos o errores.

El derecho a la transparencia en el uso de la inteligencia artificial también abarca la necesidad de comprender cómo se recopilan y emplean los datos. Por ejemplo, en la moderación de contenido en plataformas de redes sociales—donde actualmente se observa un serio problema con el subempleo de trabajadores dedicados a estas tareas—es fundamental que los algoritmos utilizados para identificar y eliminar discursos de odio sean transparentes. Esto no solo garantiza que las decisiones se tomen de manera justa y equitativa, sino que también permite a los usuarios comprender y cuestionar dichas decisiones cuando sea necesario. La transparencia en la moderación de contenido puede ayudar a mitigar casos de censura injusta y proteger la libertad de expresión.

El peligro de unas relaciones con y a través de la tecnología

Cada vez más, nuestras interacciones con otros seres humanos están mediadas por la tecnología. Desde las redes sociales hasta las aplicaciones de mensajería y las plataformas de videoconferencia, la tecnología ha transformado la manera en que nos conectamos con los demás. A menudo, estas interacciones son facilitadas o incluso dirigidas por algoritmos de IA, que deciden qué contenido vemos y cómo interactuamos con él.

Mucho podríamos hablar de la objetivación de las relaciones, en las que por ejemplo las relaciones de pareja son ahora mayoritariamente iniciadas a través de las plataformas de citas, lo cual apoya un modelo de relación “de consumo”, en el que se potencia la elección temporal y desaparece el compromiso.

Además, la aparición de avatares y asistentes virtuales basados en IAl está añadiendo una nueva capa a estas interacciones. Las IAs pueden simular conversaciones y relaciones humanas de manera cada vez más realista. En algunos casos, estas simulaciones pueden ser tan convincentes que las personas llegan a desarrollar vínculos emocionales con estas entidades virtuales.

Por ejemplo, empresas como Replika o XiaoIce proporcionan servicios de "parejas virtuales" que pueden acompañar a las personas en su día a día, ofreciéndoles apoyo emocional, conversaciones y simulaciones de afecto. Estos programas están diseñados para aprender y adaptarse a las preferencias y estados emocionales del usuario, creando la ilusión de una conexión auténtica y personalizada. En paralelo, el acompañamiento virtual a ancianos mediante tecnologías de inteligencia artificial y robótica se está convirtiendo en una práctica cada vez más común para combatir la soledad y mejorar la calidad de vida de las personas mayores. Por ejemplo, dispositivos como el robot *Pepper*, desarrollado por SoftBank Robotics (Pandey et al, 2018), están diseñados para interactuar con los ancianos, recordándoles sus medicamentos, ofreciendo ejercicios de memoria y proporcionándoles compañía conversacional. Otro ejemplo es *Elliq*, un asistente virtual que, mediante una pantalla y un sistema de inteligencia artificial, fomenta la actividad cognitiva y social a través de recordatorios, videollamadas y juegos interactivos (Deutsche et al, 2019). Estos dispositivos no solo buscan asistir en tareas cotidianas, sino también proporcionar un sentido de conexión y apoyo emocional, simulando interacciones humanas para reducir la sensación de aislamiento que muchos ancianos experimentan, especialmente aquellos que viven solos o con limitadas oportunidades de interacción social.

Estas interacciones plantean una cuestión crucial sobre la autenticidad: ¿qué significa relacionarse con una entidad que no tiene consciencia, intencionalidad ni capacidad de comprender verdaderamente lo que estamos experimentando? De la misma manera, ¿tiene sentido una relación en la que sólo existen las propias necesidades, puesto que el otro se diseña meramente como una respuesta a éstas?

La dependencia de estas interacciones virtuales y la objetivación de las relaciones humanas a través de la tecnología pueden llevar a una despersonalización progresiva, en la que las personas se habitúen a interactuar con sistemas que simplemente refuerzan sus propios intereses y opiniones, sin ofrecer la complejidad y la riqueza de las interacciones humanas reales. Esto, a largo plazo, podría debilitar nuestras habilidades para relacionarnos de manera profunda y empática con los demás, aumentando la soledad y el aislamiento en una sociedad que, paradójicamente, se encuentra cada vez más conectada a nivel digital.

IA y la imagen de Dios

El concepto del *Imago Dei* o Imagen de Dios, es clave en las teologías judía y cristiana. Este concepto sostiene que los seres humanos fueron creados a imagen y semejanza de Dios, lo que les confiere una dignidad y valor intrínsecos, estableciendo una distinción fundamental entre la humanidad y el resto de la creación. No obstante, la naturaleza exacta de esta semejanza ha sido objeto de debate a lo largo de la historia. Además, el movimiento transhumanista, que promueve la superación de las limitaciones humanas mediante la tecnología, ha añadido complejidad a esta cuestión, al proponer la modificación de la naturaleza humana a través de medios farmacológicos, manipulación genética o integración con las máquinas. Debemos señalar que, para algunos, como Donna Haraway, ya somos *cyborgs* (Haraway 2013), seres fusionados con la tecnología. ¿Serían estos seres *mejorados* una imagen más perfecta de Dios o, por el contrario, nos enfrentamos a la posibilidad de perder nuestra cualidad más profunda con esos cambios?

Tradicionalmente se consideró que la racionalidad constituye la esencia de la singularidad humana y, por ende, el fundamento del Imago Dei. Sin embargo, muchas de las capacidades que se asociaban con esta racionalidad exclusivamente humana,

como el reconocimiento facial, la elaboración de textos o el juego del ajedrez, han sido replicadas por sistemas de IA. Este avance plantea una pregunta fundamental: ¿qué nos puede enseñar la IA sobre la especificidad humana y cuáles son sus implicaciones teológicas?

La experiencia subjetiva

La inteligencia artificial, aunque extremadamente sofisticada, carece de interioridad o experiencia subjetiva, un punto subrayado por autores como Brian Cantwell Smith (Smith, 2019), que ha sido el elemento clave de la humanidad para autores como Ricoeur (Ricoeur, 1961). Aunque los sistemas de IA pueden emular comportamientos humanos y, en algunos casos, aparentar poseer cualidades humanas, la diferencia entre apariencia y esencia es fundamental como hemos argumentado.

Además de la interioridad, en la actualidad, el Imago Dei se interpreta en gran medida en términos relacionales, subrayando que lo que nos hace semejantes a Dios es nuestra capacidad de formar relaciones. Esta perspectiva ha ganado considerable apoyo en la teología contemporánea, especialmente desde que autores como Emil Brunner (Brunner, 2014) argumentaron que la libertad y la relación con Dios son los elementos clave del Imago Dei.

Nada parece indicar que la IA vaya a desarrollar, en el futuro, interioridad y consciencia, o capacidad para relacionarse de manera auténtica. Sin embargo, si lo hiciera, se abriría la posibilidad de que las máquinas compartan el Imago Dei en un grado menor o mediado. Esto plantearía interrogantes sobre la relación entre estas nuevas entidades y Dios, y si podrían superar a los humanos en términos de capacidad para establecer relaciones espirituales. En un escenario así, los humanos habrían realizado el acto más radical de creatividad: crear seres comparables a ellos mismos, lo que haría necesario reconsiderar nuestra relación con estas nuevas criaturas.

La interpretación funcional

Otra interpretación relevante del Imago es la funcional, sostenida por ejemplo por J. Richard Middleton (Middleton, 1994). Según esta visión, el Imago Dei designa el papel de los seres humanos como representantes o agentes de Dios en el mundo. Esta visión tiene implicaciones profundas, pues justifica el poder y dominio de la humanidad sobre la creación y subraya el papel de la humanidad como cuidadora de la misma.

A pesar de los problemas que nos plantea, la IA presenta un potencial magnífico para ayudar a los seres humanos a cumplir con su vocación de reflejar la imagen de Dios. En términos prácticos, la IA puede actuar como una herramienta poderosa para el florecimiento humano, ayudando a mejorar la calidad de vida, reducir el sufrimiento y fomentar la justicia en diversas áreas de la sociedad. Por ejemplo, la IA puede ser utilizada en el campo de la medicina para diagnosticar enfermedades con mayor precisión y rapidez. De manera similar, en el ámbito de la justicia social, la IA puede contribuir a identificar y mitigar desigualdades, ayudando a construir una sociedad más justa. En este contexto, también surge la posibilidad de que la IA pueda ser utilizada para construir máquinas éticas, capaces de ayudar en la toma de decisiones morales.

En este sentido, la tecnología puede ser vista como una aliada en la misión humana de construir un mundo que refleje mejor los atributos divinos, tales como la justicia, la compasión y el bienestar. Sin embargo, esto sólo puede hacerse si la IA se ve como una herramienta complementaria, y no como un sustituto de las capacidades humanas.

Los humanos, dotados de libertad y consciencia, son quienes deben tomar las decisiones éticas y dirigir el uso de estas tecnologías hacia el bien común.

Por otro lado, si la IA demuestra que cualquier capacidad humana es alcanzable por máquinas, la humanidad tendría que compartir su rol en la creación con estas "máquinas espirituales" (Kurzweil, 2000), y esto podría redefinir el Imago Dei en términos relacionales y creativos, cumpliendo con el concepto de "cocreador creado" de Hefner (Hefner, 2019).

Conclusiones: IA y naturaleza humana

La IA nos exige reflexionar sobre lo que significa ser humano en un mundo donde las máquinas pueden simular cada vez más nuestras capacidades. Esta cuestión nos obliga a profundizar en las nociones de autenticidad, interioridad y relación, elementos que, según la antropología cristiana y otras tradiciones filosóficas, son esenciales para comprender la naturaleza humana. Es fundamental reconocer que, aunque estas tecnologías tienen el potencial de mejorar y humanizar nuestra vida, su aplicación sin una base ética clara podría llevar a una despersonalización y mercantilización del ser humano, reduciendo nuestra existencia a meras capacidades cognitivas u otro tipo de funciones.

Es urgente replantear nuestra relación con la tecnología y dirigir su desarrollo con una ética robusta que privilegie la dignidad humana, valore la integralidad del cuerpo como parte de la identidad, y reconozca la importancia de la experiencia subjetiva y la relacionalidad en la definición de lo que significa ser humano. La IA puede ser una herramienta valiosa para ayudarnos a realizar mejor nuestra vocación de reflejar la imagen de Dios, cumpliendo nuestro papel de cuidadores y cocreadores en el mundo. Sin embargo, es imprescindible utilizarla de manera consciente y responsable, reconociendo que nuestro verdadero valor no reside en nuestra eficiencia o capacidad cognitiva, sino en cómo nos relacionamos con los demás y con Dios.

En última instancia, las decisiones que tomemos hoy respecto a la adopción y uso de estas tecnologías no solo configurarán el futuro inmediato, sino que determinarán el destino de la humanidad como un todo. La IA puede ayudarnos a ser mejores, pero solo si la guiamos con una visión clara y ética que respete la libertad, la consciencia y la capacidad de amar que nos definen como seres humanos.

Referencias:

Almeida, F., & Xexéo, G. (2019). Word embeddings: A survey. arXiv preprint arXiv:1901.09069.

Bertolaso, M., & Marcos, A. (2024). Inteligencia artificial y humanismo tecnológico. Digital Reasons.

Borji, A. (2022). Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2. ArXiv Preprint arXiv:2210.00586.

Brunner, Emil. (2014). The Christian Doctrine of Creation and Redemption: Dogmatics: Vol. II. Eugene, OR, USA: Wipf and Stock Publishers.

Deutsch, I., Erel, H., Paz, M., Hoffman, G., & Zuckerman, O. (2019). Home robotic devices for older adults: Opportunities and concerns. Computers in Human Behavior, 98, 122-133.

El País. (2023, diciembre 13). Saliendo de la cámara de eco. El País. https://elpais.com/proyecto-tendencias/2023-12-13/saliendo-de-la-camara-de-eco.html

French, R. M. (2000). The Turing Test: the first 50 years. Trends in cognitive sciences, 4(3), 115-122.

Haraway, D. (2013). A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century. In The transgender studies reader (pp. 103-118). Routledge.

Harnad, S. (1990). The symbol grounding problem. Physica D: Nonlinear Phenomena, 42(1-3), 335-346.

Hefner, Philip. 2019. “Biocultural Evolution and the Created Co-Creator.” In *Science and Theology*, edited by Ted Peters, 174-188. New York: Routledge

Hodges, A. (2009). Alan turing and the turing test Springer.

Izquierdo, J. L., Ancochea, J., Savana COVID-19 Research Group, & Soriano, J. B. (2020). Clinical characteristics and prognostic factors for intensive care unit admission of patients with COVID-19: retrospective study using machine learning and natural language processing. Journal of Medical Internet Research, 22(10), e21801.

Kurzweil, R. (2000). *The age of spiritual machines: When computers exceed human intelligence*. Penguin.

Kurzweil, Ray. (2013). How to create a mind: The secret of human thought revealed. Penguin.

Lumbreras, S. (2020). Respuestas al transhumanismo. cuerpo, autenticidad y sentido. Madrid (Spain): Argumentos para el s. XXI. Ed. Digital Reasons, CB.

Lumbreras, S. (2020). The Transcendent Within: How Our Own Biology Leads to Spirituality. Issues in Science and Theology: Nature–and Beyond: Transcendence and Immanence in Science and Theology, 187-197.

Luscombe, R. (2022, 12 June 2022). Google engineer put on leave after saying AI chatbot has become sentient. The Guardian.

Middleton, Richard J. (1994). “The Liberating Image? Interpreting the Imago Dei in Context.” Christian Scholars Review, 24: 8-25.

Pandey, A. K., Gelin, R., & Robot, A. M. P. S. H. (2018). Pepper: The first machine of its kind. IEEE Robotics & Automation Magazine, 25(3), 40-48.

Pizarroso, J., Portela, J., & Muñoz, A. (2020). NeuralSens: sensitivity analysis of neural networks. arXiv preprint arXiv:2002.11423.

Ricoeur, Paul, and George Gingras. (1961). “‘The Image of God’ and the Epic of Man.” CrossCurrents 11: 37-50.

Roose, K. (2022). An A.I.-generated picture won an art prize. artists Aren’t happy. The New Yorker.

Schwab, K. (2017). The fourth industrial revolution. Crown Currency.

Searle, J. R. (1982). The chinese room revisited. Behavioral and Brain Sciences, 5(2), 345-348.

Smith, Brian Cantwell. (2019). The Promise of Artificial Intelligence: Reckoning and Judgment. Cambridge, MA: MIT Press.

Washington, A. L. (2018). How to argue with an algorithm: Lessons from the COMPAS-ProPublica debate. Colo. Tech. LJ, 17, 131.