Yo no soy un robot: reflexiones sobre inteligencia artificial y sociedad mediante el ejemplo de los "captcha"

Pablo Seijo[1]

RESUMEN

En este artículo se analizan mediante el ejemplo de los *captchas* las implicaciones de la inteligencia artificial en aspectos cotidianos de los que todos hemos sido parte. Veremos en qué se relacionan los *captchas* con Turing y por qué los algoritmos juegan el *imitation game*. Nos preguntaremos qué significa ser un robot en los tiempos de inteligencia artificial fuerte y las implicaciones éticas y sociales de estos fenómenos tecnológicos. Discutiremos cómo debe comportarse una inteligencia artificial al momento de preservar nuestra libertad y privacidad.

PALABRAS CLAVE

Captcha, recaptcha, Google, test de Turing, inteligencia artificial, robots, algoritmos

Nota del Editor. Fecha de recepción: 29 de mayo de 2020. Fecha de aceptación: 6 de agosto de 2020.

[1] Licenciado en Bibliotecología (Universidad de la República), Analista Programador (Instituto Tecnológico Bios - Universidad ORT), Maestrando en Información y Comunicación (Facultad de Información y Comunicación de la Universidad de la República).

ABSTRACT

This article explains using the *captcha* example the implications of artificial intelligence in everyday aspects. We'll understand how captchas relate to Turing and why algorithms play the "imitation game". We will ask ourselves what does it mean to be a robot in times of strong artificial intelligence and the ethical and social implications of these technological phenomena. We will also discuse how artificial intelligence should behave in order to preserve our freedom and our privacy.

KEYWORDS

Captcha, recaptcha, Google, Turing test, artificial intelligence, robots, algorithms

1. CAPTCHA!

Todos conocemos los *captchas* y probablemente los habremos usado más de una vez: son esas "palabritas" que aparecen en diversas páginas de Internet para verificar que seamos humanos y muchas veces vienen acompañados de la pregunta: "Soy un robot?". La función del *captcha* es evitar que software malicioso se haga pasar por una persona para crear una cuenta falsa o algún otro tipo de objetivo *non sancto*.

La definición oficial que aparece en la web de Google dice:

> ¿Que es un Captcha? La prueba de un CAPTCHA consta de dos partes simples: una secuencia de letras o de números generada aleatoriamente que aparece como una imagen distorsionada y un cuadro de texto. Para superar la prueba y probar que eres un ser humano, simplemente tienes que escribir los caracteres que veas en la imagen del cuadro de texto (Google, 2019).

Quedémonos con las palabras *test* (cuya traducción tomaremos como "prueba") y "aleatorio" que utiliza Google en su sitio. Esto nos servirá más adelante, pero ya vemos que dos palabras típicas de la jerga matemática surgen en la definición.

El *captcha* fue un producto de la investigación científica académica, creado a mediados del año 2000 en la universidad Carnegie Mellon ubicada en Pittsburgh, uno de los centros mundiales más respetados en las áreas de computación y robótica, y fue implementado por un grupo liderado por el guatemalteco Luis von Ahn.

Captcha significa *Completely Automated Public Turing test* (test de Turing completamente automatizado). Pero ¿qué es el test de Turing? Podría definirse como una prueba de la capacidad de una máquina para exhibir un comportamiento inteligente similar al de un ser humano o indistinguible de este (Turing, 1950).

Luego volveremos a Turing. Con lo definido anteriormente la "idea" del *captcha* parecía clara. Pero, como toda tecnología, es "modular", es una "herramienta" y como tal puede tener múltiples usos.

Esta tecnología, desarrollada en una universidad, fue comenzada a utilizarse para fines diferentes al original.

2. RECAPTCHA

Weintraub en 2009 afirmaba que "Google ceremoniosamente anunció hoy que estaban adquiriendo una pequeña compañía académica llamada reCAPTCHA, que construye *software* que trate de diferenciar humanos de algoritmos en envíos *web*" (Weintraub, 2009). Aquí tenemos la primera referencia al término "algoritmo"; volveremos más adelante al mismo, pero aparentemente es importante poder diferenciarnos de estos. El *captcha* era una buena invención y resulta entonces normal que una compañía enorme de Internet la adquiera. Pero aquí es donde la historia se vuelve interesante. Google desde hacía años venía desarrollando proyectos de digitalización de documentos para sus programas "Google Books" y "Google News" mediante una tecnología llamada OCR, pero en el proceso de reconocimiento óptico de caracteres (OCR) algunas palabras del texto no podían ser reconocidas por los algoritmos de OCR.

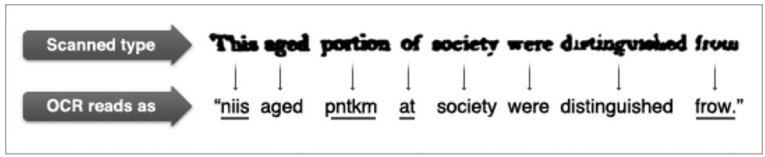
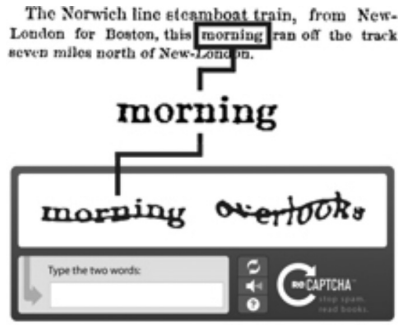


imagen tomada de: https://wayback.archive-it.org/all/20100731041527/
http://www.google.com/recaptcha/learnmore

Por lo tanto, Google decidió que sería interesante aprovechar las pruebas de *captcha* hechas por nosotros diariamente para ayudar a mejorar sus programas de OCR utilizándolos para ayudar a la digitalización de colecciones del *New York Times* o de la *National Library de Nueva York*.



> Esta tecnología también impulsa proyectos de escaneo de gran escala como Google Books y Google News Archive Search. Tener la versión textual de documentos es importante porque el texto sin formato puede ser rastreado (*searched*), fácilmente dispuesto en dispositivos móviles y desplegado a usuarios visualmente impedidos. Así que estaremos aplicando esta tecnología dentro de Google no solo para incrementar la protección contra el fraude y el *spam* para los productos Google sino también para mejorar nuestros procesos de escaneo de libros y periódicos (Google Official Blog, 2009).

Pero ¿de qué volumen de trabajo estamos hablando? Esto dice la página oficial de la universidad Carnegie Mellon: "Eso es más de 150.000 horas de precioso trabajo humano que se pierden cada día, pero a las cuales ponemos en buen uso con los reCAPTCHAs" (Carnegie Mellon University, 2008)

Ahora es bastante cuestionable el "buen uso" cuando nos estamos refiriendo a trabajar gratuitamente y sin saberlo para una de las empresas más ricas del mundo.

> Además, la cantidad de trabajo que puede ser cumplido es hercúlea. Más de 100 millones de CAPTCHAs son resueltos cada día y si bien cada *puzzle* toma sólo unos pocos segundos para resolverse, la cantidad agregada de tiempo se traduce a cientos de miles de horas de esfuerzo humano que puede ser potencialmente aprovechado. Durante el primer año de operación del sistema reCAPTCHA, más de 1.2 billones de reCAPTCHAs han sido resueltos y más de 440 millones de palabras han sido descifradas. Eso es el equivalente a transcribir manualmente más de 17.600 libros (Carnegie Mellon University, 2008).

Bueno, digamos que esas 150 mil horas de trabajo al día y los 440 millones de palabras descifradas son para una "buena causa": escanear libros y revistas para aumentar el acervo de conocimiento disponible para la humanidad.

Pero ¿qué tal si también ayudáramos a Google Maps a identificar señales de tránsito?

En la página oficial de Google sobre *captcha* dice esto:

> Cientos de Millones de CAPTCHAs son resueltos cada día por personas. RECAPTCHA hace un uso positivo de este esfuerzo humano al canalizar el tiempo invertido en resolver CAPTCHAs para reconocer imágenes y construir bases de datos para *machine learning*. Esto a su vez ayuda a mejorar los mapas y resolver problemas complejos de IA (Google, 2019b).

Nosotros, los humanos, amablemente ayudamos entonces (mayormente sin saberlo) a resolver problemas de inteligencia artificial (IA). Quedémonos de lo citado con "humanos resolviendo problemas de inteligencia artificial".

Imagen tomada de: https://techcrunch.com/2012/03/29/google-now-using-recaptcha-to-decode-street-view-addresses/

Por lo tanto, lo que empezó como una tecnología para impedir que software malicioso se hiciera pasar por humanos terminó con humanos ayudando a "resolver" problemas de inteligencia artificial y de paso trabajando gratuitamente. Seguro luego de leer estas líneas el lector observará los *captchas* desde una perspectiva diferente, pero lo bueno es que será consciente de que ha trabajado un tiempo para Google y quizás pueda volcarlo en su currículum.

3. *IMITATION GAME*

Pero volvamos al principio y recordemos que el *captcha* lleva su nombre por *Completely automated public Turing test*. Turing fue un científico británico, matemático, pionero de la informática, la algorítmica y la inteligencia artificial. En su famoso test, Turing reformó un antiguo desafío llamado test de "imitación" cuyo objetivo era identificar entre tres personas (solo mediante preguntas) quién era una mujer y quién un hombre. La gracia del juego era que el que el impostor (al que evidentemente no podía verse ni oírse) podía mentir o en otras palabras "engañar" al otro jugador. En el test de Turing una persona intenta distinguir entre un humano y una máquina.

Turing en su canónico artículo *Computing machinery and intelligence* (Turing,1950) describe este test y se plantea si las máquinas realmente pueden pensar.

> ¿Podría ser que las máquinas realizaran algo que pudiera ser descrito como pensar, pero que es muy distinto de lo que un hombre hace? Esta objeción es muy sólida, pero al menos se puede decir que si, a pesar de todo, una máquina puede ser construida para jugar satisfactoriamente el juego de la imitación, no necesitamos preocuparnos por esta objeción (Turing,1950, p. 433).

Turing en su genialidad no da a la pregunta una respuesta directa, pero afirma que una máquina sí puede "imitar" y sobre todo "engañar" a otro ser humano para que este crea que se enfrenta a otro ser humano.

> El test no se centra en el planteamiento de condiciones necesarias y suficientes para la existencia de inteligencia, sino que acentúa la postulación de un método que, mediante la obtención de evidencia estadística, indique que un computador posee mente e inteligencia (González, 2007, p. 181).

Turing ubica el problema en otra parte: lo saca de la máquina y nos lo da a nosotros, planteando la inteligencia artificial como resultados estadísticos de la percepción humana. Los críticos a esta posición como Searle ven una vuelta al dualismo cartesiano, porque se niega la posibilidad de inteligencia fuera de la mente humana.

> Hay en estas discusiones, en pocas palabras, un tipo de dualismo residual. Los partisanos de la IA creen que la mente es más que una parte del mundo biológico natural; creen que la mente es puramente formalmente especificable. La paradoja de esto es que la literatura de la IA está llena de fulminaciones contra algunas perspectivas llamadas "dualismo", pero, de hecho, las tesis completas de la AI fuerte descansa sobre un tipo de dualismo. Descansa en el rechazo de la idea de que la mente es sólo un fenómeno biológico natural en el mundo como cualquier otro (Searle, 1984, p. 13).

Este dualismo cartesiano impera en las nuevas aproximaciones a la IA de la llamada "inteligencia artificial fuerte" (la que supone que la IA iguala o supera la inteligencia humana) porque este tipo de inteligencia, al ser suprahumana, no puede estar contenida en ningún cuerpo no humano.

Este punto es fundamental porque en la era de Internet un robot es más un programa informático que un androide, es más la imitación de la "inteligencia" humana y no de la apariencia o mecánica de nuestro cuerpo. Probablemente en el juego de imitación de Turing podremos darle un punto a Descartes.

Aquí tenemos una cuestión interesante pues anteriormente Google nos planteó que es importante para ellos distinguir humanos de algoritmos. ¿Pero cuál es el problema con los algoritmos?

4. RÍOS DE TINTA

La palabra "algoritmo" proviene del nombre del matemático persa del siglo VII Muhammad ibn Musa al-Khwarizmi. Cuando posteriormente en el siglo XII se tradujo al latín su libro sobre algebra: *Algoritmo de Número Indorum* pasó él mismo a conocerse como Algoritmi o Al-Juarismi. También deriva de su nombre el término "guarismo" y a partir del título de otro de sus libros "al-jabr w'al-muqabala" se acuñó el término "álgebra" (Gandz, 1926). Al-Juarismi fue uno de los cientìficos más importantes de la antigüedad y también estuvo por más de veinte años al frente de la "casa

de la sabiduría" o Gran biblioteca de Bagdad. Esta legendaria academia de la antigüedad se dedicó al cuidado, estudio y desarrollo de la ciencia y la cultura por más de cuatrocientos años, fue un importante centro de la traducción y copia de los materiales griegos y romanos y en ella trabajaron Averroes (famoso comentarista de Aristóteles) y Avicena (autor de *El canon de la medicina*) entre otros. Muchos años después de Al-Juarismi, durante el sitio de Bagdad por parte de los mongoles en 1258, la biblioteca fue destruida y se dice que los manuscritos fueron arrojados al río Tigris en cantidades tales que el río corría negro por la tinta de estos.

> Sus tropas incendiaron obstinadamente las bibliotecas o arrojaron al Tigris durante una semana "libros que superaban cualquier descripción […] y que formaron un puente sobre el que pasaron los soldados de la infantería y los caballeros, y el agua del río ennegreció por la tinta de los manuscritos (Polastron, 2015, p. 60).

Mucha de esta tinta ha llegado a nuestros tiempos y si bien se habla poco de Muhammad ibn Musa al-Khwarizmi, los algoritmos dominan nuestras vidas como nunca en la historia hasta ahora y escapar de estos, como veremos más adelante, es un trabajo digno de una moderna "casa de la sabiduría".

Tomando una definición básica como la del DRAE un algoritmo es un "Conjunto ordenado y finito de operaciones que permite hallar la solución de un problema." Por lo tanto, un algoritmo puede ser desde una receta de cocina hasta una fórmula matemática. Supongo que a estas alturas intuimos que a lo que teme Google no es a nada de eso. Aquí nos volveremos a ayudar con nuestro amigo Turing.

Según la tesis de Church-Turing, "Informalmente, un algoritmo es una colección de instrucciones simples para realizar alguna tarea. La tesis Church-Turing propone que la noción intuitiva de algoritmo es equivalente a las máquinas de Turing" (Cleland,1993, p. 285). Básicamente lo que dice es que, si bien un algoritmo es una fórmula, un algoritmo posee las capacidades de las máquinas de Turing: así, un algoritmo es inteligencia artificial.

5. PATRONES, MODELOS Y EL CUELLO DEL POLLO

Hal 9000 sí es un robot, y aparece por primera vez en la novela de 1968 escrita por Arthur C. Clarke *2001, Odisea del espacio*. Su nombre significa: *Heuristically Programmed Algorithmic Computer* (Computadora Algo-

rítmica Heurísticamente Programada), a esta altura términos conocidos, entendiendo por su nombre que es una computadora capaz de aprender. Hal tenía las habilidades de reconocimiento de voz, reconocimiento facial y procesamiento de lenguaje, casi como el celular que llevamos en nuestro bolsillo. Sin hacer *spoilers*, adelantamos que Hal "intenta", por el bien del algoritmo, asesinar a algunos de los tripulantes de la nave espacial que controla. Paradójicamente la figura de Hal para el filme homónimo dirigido por Stanley Kubrick fue literalmente construida por uno de los pioneros de la inteligencia artificial, Marvin Minsky, quien logró muchos años antes que nuestras empresas tecnológicas actuales crear una parte del algoritmo, que es "maligno" sin lugar a dudas.

Minsky es una figura mítica, no sólo por haber sido parte de *2001, Odisea del espacio* o *Jurassic Park* (anótese otro lindo momento en que el desarrollo tecnológico desemboca en asesinatos de humanos) sino por ser, junto con Turing y Church, uno de los padres de la inteligencia artificial. En su canónico artículo de 1961 "Steps toward artificial intelligence" (Pasos hacia la inteligencia artificial) nos dice:

> Una computadora puede hacer, en algún sentido, solo lo que se le dice que haga. Pero incluso cuando no sabemos cómo resolver un cierto problema, podríamos programar una máquina (computadora) para buscar a través del gran espacio de intentos de solución. Desafortunadamente, esto usualmente lleva a un proceso enormemente ineficiente. Con técnicas de reconocimiento de patrones, la eficiencia puede ser a menudo incrementada, al restringir la aplicación de los métodos de la máquina a los problemas apropiados. El reconocimiento de patrones, junto con el aprendizaje, pueden ser usados para explotar generalizaciones basadas en la experiencia acumulada, reduciendo aún más la búsqueda. Al analizar la situación, utilizando métodos de planeamiento, podríamos obtener una mejora fundamental reemplazando la búsqueda dada por una exploración mucho más pequeña y apropiada. Para manejar clases de problemas amplios, las máquinas necesitarán construir modelos de estos entornos, utilizando algún esquema para la inducción (Minsky, 1961, p. 8).

Es un párrafo extenso, pero pasemos a diseccionar. Primero dice "una computadora solo puede hacer lo que se le ordena", "incluso si no sabemos cómo resolver cierto problema". En nuestro ejemplo, discriminar tipografías confusas o detectar un cruce peatonal. Además "podemos programar una computadora para buscar patrones". No olvidemos que Google no busca identificar un semáforo en concreto, busca que Google Maps pueda identificar cualquier semáforo en casos de conducción autónoma. Minsky también utiliza la palabra "aprendizaje". Aquí estamos entran-

do al terreno del *machine learning*. En este punto el algoritmo aprende a identificar patrones que lo llevan a un resultado exitoso. Al final del párrafo nos advierte que para manejar casos más amplios y construir modelos estos algoritmos deben usar cierto esquema inductivo.

O sea, el algoritmo aprende, la capacidad de procesamiento le permite buscar patrones entre muchísimas soluciones posibles y crea modelos en un esquema de inducción. Esta puede decirse que a grandes rasgos es la base del *machine learning*, pero este esquema inductivo es un problema para la ciencia en general y es tema de discusión epistemológica constante.

Osoba y Welser IV en su libro *An intelligence in our image: The risks of bias and errors in artificial intelligence* (Una inteligencia a nuestra imagen: los riesgos de sesgos y errores en la inteligencia artificial") reflexionan sobre el problema de la inducción en la IA:

> El problema de aprender a distinguir entre verdad y falsedad a través de la experiencia es más formalmente conocido como el problema de la inducción. La pregunta central es cuán jusificable es aplicar generalizaciones basadas en limitadas experiencias pasadas a nuevos escenarios (Osoba y Welser, 2017, p. 5).

Minsky simplemente nombra el problema, pero Bertrand Russell, en *Los problemas de la filosofía*, reflexiona al respecto y nos da un ejemplo crudo de los problemas de la inducción y del esperar que la realidad se comporte siempre de manera uniforme.

> Los animales domésticos esperan su alimento cuando ven la persona que habitualmente se lo da. Sabemos que todas estas expectativas, más bien burdas, de uniformidad, están sujetas a error. El hombre que daba de comer todos los días al pollo, a la postre le tuerce el cuello, demostrando con ello que hubiesen sido útiles al pollo opiniones más afinadas sobre la uniformidad de la naturaleza (Russell, 1973, p. 61).

Tomando como partida lo que dice Russell, que un algoritmo tenga como *dataset* datos relevados del comportamiento de millones de usuarios no significa que siempre vaya a actuar correctamente y que sea justo o ético o apropiado para los usuarios en el futuro. En este componente ético nos propondremos ahondar a continuación.

6. LOS BRAZOS DE HAL: INTELIGENCIA ARTIFICIAL INCOMPRENDIDA Y NO REGULADA

Para entender lo que implica nuestro objeto de análisis (el *captcha*) tuvimos que entender procesos internos de la ciencia, su denominación, su función y por último su ámbito de justificación. Pero es en ese ámbito donde el piso parece moverse: en el ámbito científico-académico la justificación para su uso es clara, pero al ser adquirida por una corporación como Google y al utilizar nuestras horas impagas para mejorar una IA cuyo fin desconocemos no lo es tanto.

Para asegurar el uso correcto de la IA y cómo debe ser su desarrollo a futuro diferentes académicos y organizaciones internacionales proveen recomendaciones de distinta índole. Aquí recogeremos algunas que nos parecieron destacables. Osonde Osoba y Willam Welser IV proponen tres opciones para enfrentar las posibles consecuencias negativas de la incursión de los algoritmos en nuestras vidas: evitar los algoritmos, exigir la transparencia de estos o auditarlos.

> Las respuestas a los agentes artificiales sin regular tienden a ser de tres tipos amplios: evitar los algoritmos por completo, hacer que los algoritmos subyacentes sean transparentes o auditar el *output* de los algoritmos. Evitar los algoritmos es probablemente imposible, pocas opciones alternativas están disponibles para dar sentido al aluvión actual de datos. La transparencia algorítmica requiere un público más educado capaz de entender los algoritmos (Osoba & Welser, 2017, p. 35).

6.1 Primer brazo: evitarlos

Evitar los algoritmos parece muy difícil ya que tendríamos que abandonar la vida digital por completo y, como plantea Cathy O'Neill en su libro *Armas de destrucción matemática: cómo el big data aumenta la desigualdad y amenaza en democracia* estaríamos renunciando a la posibilidad de generar un cambio a futuro y cayendo en problemas éticos al saber que algo es erróneo e igualmente mirar a otro sitio. Ella plantea, al igual que Derman (que veremos más adelante), que los modelos no se construyen solamente con datos sino con decisiones. Ser parte también es cuestionar, difundir y comunicar. Entender que los modelos, si bien son ejecutados por máquinas, son creaciones humanas y por el hecho de su complejidad o hermetismo no debemos mirarlos como fenómenos inevitables con los cuales lo único que podemos hacer es ponernos a resguardo. Textualmente plantea:

> Estos modelos no se construyen únicamente con datos, sino también con las decisiones que tomamos sobre cuáles son los datos a los que debemos prestar atención y qué datos dejaremos fuera. Y esas decisiones no se refieren únicamente a cuestiones logísticas, de beneficios o eficiencia, sino que son fundamentalmente decisiones morales. Si nos retiramos y tratamos los modelos matemáticos como si fueran una fuerza neutra e inevitable, como la meteorología o las mareas, estaremos renunciando a nuestra responsabilidad (O'Neill, 2017, p. 172).

6.2 Segundo brazo: transparencia

Pese a ser algo con lo que convivimos cotidianamente rara vez notamos que hay algoritmos operando en cosas tan banales como una solicitud a una página web. Porque como preveía Turing la inteligencia artificial despliega un juego de imitación donde su objetivo es no ser detectada. Por lo tanto, el identificar su accionar es uno de los principios fundamentales para un desarrollo positivo de la IA, y es la segunda de las vías que proponen Osoba y Weiser y se llama "transparencia".

> Transparencia: cuando un sistema de inteligencia artificial toma una decisión, las personas afectadas deben poder recibir una explicación de por qué se ha tomado esa decisión (BIOCAT, 2017).

Por lo tanto primero debemos saber que está allí para luego demandar que cumpla con objetivos para el bien común que nos lleven a una sociedad mejor y, como dice Hueso (2019), con más oportunidades económicas, sociales y políticas. La transparencia es vital porque no se puede accionar sobre algo que se desconoce, pero también es una etapa imprescindible para el siguiente punto la auditoría.

6.3 Tercer brazo: auditorías

Cathy O'Neill en su libro *Armas de destrucción matemática* nos plantea:

> ¿Y cómo empezar ahora a regular los modelos matemáticos que dirigen cada vez más nuestras vidas? Yo sugeriría que el proceso comenzara con los programadores que crean los modelos. Al igual que los médicos, los científicos de datos deberían hacer un juramento hipocrático centrado en los posibles abusos y malas interpretaciones de sus modelos (O'Neill, 2017, p. 162).

La autora nos habla de abuso y malas interpretaciones de modelos que pueden haber sido creados con objetivos loables o neutrales, pero que posteriormente se hayan utilizado para otros fines. Reclama a los creadores un compromiso ético y los responsabiliza por el mal uso ulterior de sus creaciones.

En la crisis inmobiliaria del 2008 surgió un movimiento que empezó a cuestionar los modelos de IA basados por ejemplo en maximizar ganancias, pues si bien pueden llegar a obtener una rentabilidad excelente, en un momento pueden generar una burbuja y posteriormente una crisis. Estos matemáticos crearon lo que se llamó el "Financial Modeler's Manifesto" que incluía, como pide O'Neill, un juramento hipocrático del tipo que se ve en medicina.

Uno de los firmantes del manifiesto, Emanuel Derman, ya se planteaba en 1996 los problemas éticos relacionados con la creación de modelos:

> Siempre hay presupuestos implícitos detrás de un modelo y su método de solución. Pero los seres humanos tienen una prevision limitada y gran imaginación, de tal modo que, inevitablemente, un modelo será utilizado de modos que su creador nunca pretendió. Esto es especialmente verdadero en los entornos de *trading,* donde no puede invertirse suficiente tiempo en hacer interfaces a prueba de fallos, pero también es un tema de principios, no puedes prever todo. Así, incluso un modelo "correcto", "correctamente" resuelto, puede llevar a problemas. Mientras más complejo el modelo, mayor es esta posibilidad (Derman, 1996, p. 7).

Derman aleja el problema del creador del modelo y afirma que es "inevitable" que pueda ser usado de una manera diferente a la que fue concebido originalmente. Dice que es "imposible" prever los problemas que puede generar a futuro un modelo; incluso un modelo "correcto" resuelto de manera "correcta" puede generar problemas; cuanto más complejo el problema mayor son las posibilidades que esto pase.

En el caso del captcha, el algoritmo fue ideado para detectar software fraudulento y Google después de adquirirlo lo utilizó para alimentar su algoritmo de reconocimiento. En este caso no es claro si es un "mal uso"; lo que sí es muy claro es que el algoritmo fue usado para algo muy diferente de lo que originalmente fue pensado en su génesis en el ámbito académico. Lo que afirma Derman al menos en este caso es cierto y comprobable.

Osoba y Welser IV proponen que los algoritmos sean juzgados en una forma análoga a la que usamos para los humanos: por las consecuencias de sus actos y decisiones y no por su estructura de pensamientos. En el caso de los algoritmos debemos juzgar sus resultados (outputs) y no la calidad del código en el que están escritos.

> Esto es similar a cómo a menudo juzgamos a los agentes humanos: por las consecuencias de sus resultados (decisiones y acciones) y no por el contenido o el ingenio de su código base (pensamientos). Esta opción tiene más sentido para los formuladores de políticas y establece el estándar para una ética consecuencialista para los agentes artificiales. La regulación es mucho más fácil bajo este marco. Las discusiones como esta a veces pueden antropomorfizar a los agentes artificiales: ¿Están las máquinas comenzando a pensar como nosotros y cómo podemos juzgarlas y guiarlas? El progreso actual en los agentes artificiales puede hacer que esta visión antropomórfica de algoritmos sea más cercana a la norma. Esto puede tener el beneficio inesperado de fomentar la comprensión pública de que los agentes artificiales, como los humanos, no están libres de sesgos (Osoba & Welser, 2017, p. 26).

Aquí Osoba nos plantea la base a fin de generar estándares éticos para inteligencia artificial. Afirma que, a medida que las máquinas empiezan a pensar como nosotros, la mirada antropomórfica de los algoritmos y sus comportamientos es cada vez más aceptada. A su vez, entender que los algoritmos, igual que los humanos, no pueden escapar al sesgo en sus decisiones pese a que, como explicita Derman, "resuelvan los problemas de manera 'correcta'", es la llave para abrir la caja negra que muchas veces los mismos representan para nosotros.

O'Neill afirma:

> Ya están en marcha algunos movimientos para auditar algoritmos. En Princeton, por ejemplo, los investigadores han lanzado un proyecto sobre responsabilidad y transparencia en la web. Han creado robots de software que se disfrazan en Internet como si fueran personas de todo tipo: ricos, pobres, hombres, mujeres o personas con problemas de salud mental. Los investigadores estudian el tratamiento que reciben estos robots y así pueden detectar los sesgos existentes en los sistemas automáticos, desde los motores de búsqueda hasta las páginas web de búsqueda de empleo. Se están lanzando iniciativas similares en universidades como Carnegie Mellon y el MIT (O'Neill, 2017, p. 166).

Estos ejemplos citados, el juramento hipocrático de los modeladores de algoritmos financieros, el poner el tema ético sobre la mesa como plantean

Osoba y Welser, son pasos que nos pueden llevar a lo que la Comisión Europea para el estudio de la IA propone: "la inteligencia artificial confiable":

> Aunque la IA es capaz de generar enormes beneficios para las personas y la sociedad, también entraña riesgos que se deben gestionar de manera adecuada. Puesto que, en general, los beneficios de la IA compensan sus riesgos, debemos seguir uno que maximice los beneficios y minimice los riesgos. Para asegurarnos de que vamos por el buen camino, es necesario regirse por un enfoque de la IA centrado en los seres humanos. Es decir, que nos obligue a recordar que el desarrollo y uso de la IA tienen por objetivo mejorar el bienestar de los seres humanos, y no verlos como un medio en sí mismos. La IA confiable marcará nuestro camino, ya que los seres humanos sólo podrán beneficiarse completamente y con plena confianza de la IA si pueden confiar en la tecnología (Comisión Europea, 2018).

Paradójicamente o no la IA confiable es la que está centrada en los seres humanos, una inteligencia artificial "antropocentrista", Lorenzo Cotino Hueso en su artículo "Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y big data confiables" lo define de este modo:

> Así, la IA debe desarrollarse para el bien común y el beneficio de la humanidad, mejorar el bienestar individual y colectivo, generar prosperidad, valor y maximizar la riqueza y sostenibilidad. Asimismo, debe buscar una sociedad justa, inclusiva y pacífica, ayudando a aumentar la autonomía de los ciudadanos, con una distribución equitativa de oportunidades económicas, sociales y políticas. También debe tener objetivos como la protección del proceso democrático y el Estado de derecho; la provisión de bienes y servicios comunes a bajo costo y de alta calidad; alfabetización y representatividad de los datos; mitigación de daños y optimización de la confianza hacia los usuarios (Hueso, 2019, p. 37).

7. CONCLUSIÓN

La reflexión final que proponemos es que conceptos técnicamente complejos como "algoritmo", "inteligencia artificial", *machine learning* son un desafío para profesiones como la filosofía de la ciencia, la filosofía de la tecnología y, especialmente, para el comunicador científico. Pero la relevancia de estas problemáticas y la penetración que han tenido en múltiples aspectos de nuestra sociedad hacen que debamos imbuirnos en los problemas científico-tecnológicos que ya abandonaron los centros de investigación y pasaron a ser tópicos comunes de nuestra vida cotidiana.

Por ello es fundamental exigir que se cumplan principios fundamentales como la transparencia de los algoritmos y la autonomía de los usuarios frente a estos para llegar a una sociedad que utilice el avance tecnológico como motor de desarrollo y, como afirmamos a lo largo del trabajo, que esté centrada en los seres humanos, en su beneficio individual y colectivo y no en el beneficio de las grandes compañías tecnológicas.

El propósito de este análisis es aportar a un foro de discusión permanente sobre cómo debemos actuar con respecto a estas herramientas, reflexionar sobre su uso apropiado, desmitificarlas, ir a su génesis, a su aplicación práctica, e intentar entender que hay páginas de nuestra historia actual que por su complejidad van a necesitar muchos más captchas para ser desbloqueadas.

8. REFERENCIAS

BIOCAT Y Obra social "La Caixa" (2017): "Declaración de Barcelona para un desarrollo y uso adecuados de la inteligencia artificial en Europa", <https://www.biocat.cat/sites/default/files/sinopsibdebate_artintelligence_es.pdf >,consultado el 16 de diciembre de 2019

Carnegie Mellon University (2008): "Computer users are digitizing books, newspapers quickly and accurately with Carnegie Mellon Method disponible", <https://www.cmu.edu/news/archive/2008/August/aug14_recaptcha.shtml>, consultado el 16 de diciembre de 2019

Cleland, C. E. (1993): "Is the Church-Turing thesis true?", *Minds and Machines*, *3*(3), pp. 283-312.

Comisión Europea (2018): "Proyecto de directrices éticas sobre una inteligencia artificial confinable", <https://www.algoritmolegal.com/tecnologias-disruptivas/directrices-eticas-para-una-inteligencia-artificial-confiable-en-europa/>, consultado el 16 de diciembre de 2019

Derman, E. (1996). "Model risk: What are the assumptions made in using models to value securities and what are the consequent risks?", *Risk Magazine Limited*, 9, pp. 34-38.

Gandz, S. (1926): "The origin of the term 'Algebra'", *The American Mathematical Monthly*, 33, 9, pp. 437-440.

Google (2019a):" ¿Qué es un CAPTCHA?", <https://support.google.com/a/answer/1217728?hl=es> ,consultado el 16 de diciembre de 2019

Google (2019b): "reCAPTCHA v3 : The new way to stop bots", <https://www.google.com/recaptcha/intro/v3.html#the-recaptcha-advantage>, consultado el 16 de diciembre de 2019

Google Official Blog (2009): "Teaching computers to read: Google acquires reCAPTCHA", <https://googleblog.blogspot.com/2009/09/teaching-computers-to-read-google.html>, consultado el 16 de diciembre de 2019

González, R. (2007): "El test de Turing: Dos mitos, un dogma", *Revista de filosofía*, 63, pp. 37-53.

González, R. (2011): Descartes: "Las intuiciones modales y la inteligencia artificial clásica", *Alpha*, 32, pp. 181-198.

Hueso, L. C. (2019): "Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y big data confiables", *Revista Catalana de Dret Públic*, 58, pp. 29-48.

Minsky, M. (1961): "Steps toward artificial intelligence", *Proceedings of the IRE*, *49,* 1, pp. 8-30

O'Neil, C. (2017): *Armas de destrucción matemática: cómo el big data aumenta la desigualdad y amenaza la democracia,* Capitán Swing.

Osoba, O. A. y W. Welser IV (2017): *An intelligence in our image: The risks of bias and errors in artificial intelligence,* Rand Corporation.

Polastron, L. X. (2015): *Libros en llamas: historia de la interminable destrucción de bibliotecas,* Fondo de Cultura Económica.

Russell, B., J. Xirau y E.L. Iñigo (1973): *Los problemas de la filosofía,* Labor.

Searle, J. R., G. Brown y S. Willis (1984): *Minds, brains, and science,* Harvard University Press.

Turing, A. M. (1950): "Computing machinery and intelligence", *Mind*, 59, pp. 236-433.

Université de Montréal (2017): "Montreal Declaration for a responsible development of artificial intelligence", <https://www.montrealdeclaration-responsibleai.com/the-declaration>, consultado el 16 de diciembre de 2019

Weintraub, S (2009): "Google acquires reCAPTCHA in two-for-one deal", <https://www.computerworld.com/article/2467668/google-acquires-recaptcha-in-two-for-one-deal.html>, consultado el 16 de diciembre de 2019

Williams, B. A., C.F. Brooks y Y. Shmargad (2018): "How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications", *Journal of Information Policy*, 8, pp. 78-115.