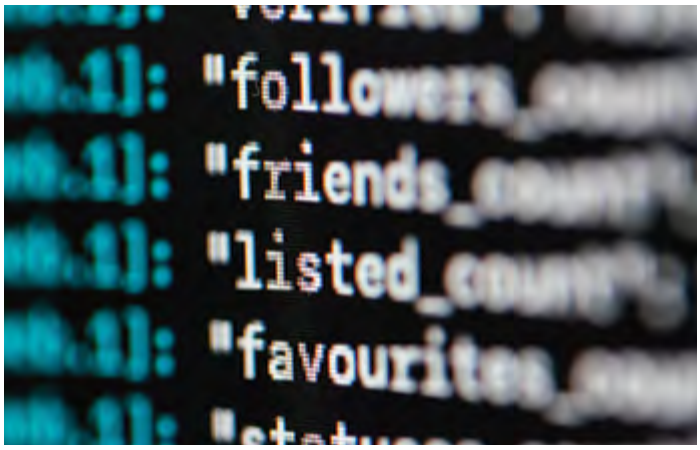

¿Qué nos hace humanos?

por **Jimena Canales**

La aparición de los bots automatizados de internet ha obligado a desarrollar programas eficaces para diferenciar a una persona de un software. La historia de esta cruzada arroja nuevas preguntas sobre lo mucho que ha cambiado nuestra definición de "lo humano".

La pregunta *¿qué nos hace humanos?* irrita a los filósofos más sofisticados tanto como a la gente común. ¿Nos distingue el uso de nuestro lenguaje, nuestra capacidad de crear arte, el amor que sentimos hacia el prójimo, nuestra conciencia del universo, o nuestras aptitudes técnicas, de raciocinio o de imaginación? Ya que todas las respuestas admiten excepciones y objeciones, podríamos argumentar que esta es una de las más grandes preguntas de todas. Está conectada con temas tan importantes como qué hace que valga la pena vivir o cuál es la importancia y el deber de conservar una vida. También está relacionada con los problemas que nos llevan a actuar inhumanamente y a olvidar lo mejor de nuestra especie.

Apenas hay escritor o filósofo que no haya intentado responder la cuestión. Muchos han dedicado su vida y obra entera a tal enigma. Podría agregar a muchos grandes pensadores a la lista de aquellos que han pensado en este tema, pero me limito a anotar algo que ha eludido a todos: indagar qué son los humanos *desde la perspectiva de los humanos* es ya una estrategia comprometida. Si verdaderamente quieres saber quién eres, tienes que averiguar qué dicen de ti.

Empecemos por preguntarnos cómo las computadoras aprendieron a distinguir entre un humano y un no humano. Pongámonos en su lugar. Según ellas, ¿cuáles son las cualidades únicas de los seres supuestamente "humanos"?

Las elecciones de 1996

Los primeros retos alrededor de la distinción de los humanos remontan a octubre de 1996. Durante las campañas electorales a la presidencia de Estados Unidos, la compañía Digital Equipment Corporation (DEC), luego adquirida por Compaq y fusionada con Hewlett-Packard, lanzó un nuevo servicio en internet. Era un concepto muy novedoso: una página web recopilaba y mostraba datos de encuestas sobre los candidatos presidenciales. El entonces presidente y candidato demócrata Bill Clinton se perfilaba de nuevo como ganador. Su competidor era el senador Bob Dole del Partido Republicano. Como parte de su estrategia, Clinton prometía invertir en el desarrollo de la internet y, junto

Fotografía: ThisIsEngineering / Pexels.

con su vicepresidente Al Gore, anunció su estrecho compromiso con la Next Generation Internet (NGI). El gobierno estadounidense vislumbraba transformaciones importantes para la economía y la sociedad que llegarían en un futuro próximo. De acuerdo con sus previsiones, las computadoras conectadas en red cambiarían la manera en que las personas se relacionarían entre ellas y con su entorno: "La internet proporcionará un espacio poderoso y versátil para los negocios, la educación, la cultura y el entretenimiento, [donde] la vista, el sonido e incluso el tacto se integrarán a través de potentes computadoras, pantallas y redes [...] La gente usará este ambiente para trabajar, realizar operaciones bancarias, estudiar, comprar, entretenerse y visitarse unos a otros."

Ese otoño los votantes escuchaban atentos las promesas dadas por los candidatos, incluyendo entre ellas la "Next Generation Internet", que afectaban su popularidad. Algunos de ellos consultaban la página de Digital Equipment Corporation que mostraba, minuto a minuto, cómo subían o bajaban en las encuestas. Pronto se notó un problema con los datos que mostraba la página: había personas que ingresaban repetidas veces a su candidato preferido en la encuesta. En vez de mostrar quién era realmente el candidato más popular, la encuesta mostraba cuál de todos tenía más simpatizantes cibernautas tramposos.

Pronto el problema en las encuestas de internet incrementó: los programadores aprendieron cómo crear *bots* –una abreviación de *robot* aplicada a un software–. Los bots podían ingresar datos automáticamente manipulando los resultados. En esos casos, la encuesta no solo no mostraba la preferencia de la población votante, sino que ni siquiera mostraba la preferencia de los seres humanos, ya que las respuestas provenían de los bots. Como precaución, el servicio de encuestas ideó una traba en contra de estos nuevos actores y comenzó a mostrar una imagen de la "bandera estadounidense en una posición aleatoria en la pantalla y luego solicitó al usuario que hiciera clic en la bandera antes de ingresar una opinión". La solución, aunque de pinta muy patriótica, era defectuosa. Rápido, los programadores crearon bots que podían perseguir las banderas en la pantalla, de manera automatizada, y hacer clic en ellas para ingresar aún más entradas en la encuesta. Críticos del nuevo sistema, entre ellos un ingeniero llamado Mark D. Lillibridge, pronto notaron cómo era "fácil escribir un programa que reconociera la bandera estadounidense y simulara un clic; por lo tanto, este método no sirve para restringir de manera efectiva el acceso de los agentes electrónicos".

En casos como este fue claro que la manipulación y el fraude en el control de la información chocaban contra los ideales de quienes defendían la internet. En agosto de 1998, en la cumbre anual sobre "Ciberespacio y el sueño americano" en Aspen, Colorado, Ira C. Magaziner –encargado del proyecto durante la administración Clinton– afirmó que en la nueva era de la revolución digital "la censura y el control de contenido no solo son indeseables, sino imposibles en la práctica". Su predicción probó ser optimista y equivocada. Regular a los cibernautas humanos era casi imposible, pero controlar a los bots resultó aún más difícil. Empresas como Digital Equipment Corporation batallaban en distinguir a los usuarios individuales de los duplicados, y a los seres humanos de los seres automáticos, ya que algunos usuarios se dedicaban a crear programas computacionales que podían imitar a los humanos, a los cuales podrían suplantar sin problemas y sin poder ser identificados.

Los bots pronto comenzaron a desempeñar roles más activos en la web. Durante el otoño de 1998, otra encuesta vía internet hizo historia. Un sitio web pidió a los usuarios que votaran por el mejor departamento de informática en las universidades del país. Los estudiantes de la Universidad Carnegie Mellon encontraron una manera de enviar miles de votos mediante programas automatizados. Al día siguiente, los estudiantes del MIT desarrollaron un programa rival. Este resultó ser más eficaz. MIT ganó con 21,156 votos, en comparación con los 21,032 de Carnegie Mellon, mientras que otras escuelas recibieron menos de mil. En este caso, la encuesta manipulada por bots había funcionado mejor de lo que se pensaba, aunque sin reflejar la opinión de los votantes humanos. El resultado indicaba qué universidades tenían los mejores programadores de bots, los mejores manipuladores de encuestas por internet y, por ende, los mejores departamentos de informática.

Durante estos años, a pesar de los desafíos que mostraba la regulación de internet, los retos se consideraron, en su mayoría, de una manera positiva. A menudo, se continuaba asociando al nuevo medio con la libertad de expresión, el libre flujo de información, la democracia y el antitotalitarismo. En buena medida se ignoraban los retos de la desinformación, las *fake news*, las teorías conspiratorias que creaban burbujas y cámaras de eco donde la información quedaba atrapada, y la perniciosa y sutil difusión de propaganda política. Cuando en el 2000 el presidente Bill Clinton discutió la posibilidad de que China se uniera a la Organización Mundial del Comercio, minimizó la capacidad de censura de ese país gracias a la nueva tecnología que prometía cambiar el orden global. De acuerdo con el presidente, los intentos de los chinos de "tomar medidas enérgicas contra internet" serían inútiles, ya que censurar internet era "como tratar de clavar gelatina en la pared". Se equivocó: censurar a los humanos que usan internet no era difícil, incluso fue mucho más fácil que controlar a los bots.

Si bien los pioneros del mundo de internet como Alan Turing y Joseph Weizenbaum intentaron crear programas que imitaran a los humanos, el desafío para el próximo milenio dio un giro de ciento ochenta grados: ahora sería distinguir los unos de los otros. Una de las primeras

solicitudes de patentes presentadas con este fin buscaba evitar que los usuarios crearan cuentas de manera automática a fin de tener múltiples accesos a los servicios de suscripción en línea. Su idea era novedosa: solicitaba que los usuarios identificaran un "patrón de validación gráfica". Según sus creadores, entre ellos Lillibridge, el nuevo invento "evita o dificulta la creación automática de múltiples cuentas de usuarios mediante el uso de scripts de programación o técnicas similares". No cualquier patrón funcionaba, sino solo aquellos "diseñados de tal manera que se dificulte la identificación usando las técnicas de reconocimiento óptico de caracteres". Por primera vez en la historia de internet, se había encontrado una barrera efectiva entre los usuarios humanos y los no humanos que poblaban el universo cibernético. Pero la barrera pronto probó ser muy porosa y la distinción entre los actores demasiado burda.

Los CAPTCHA

Los avances en la programación de bots llegaron a tal punto que alrededor de 2003 se acuñaron las siglas CAPTCHA de "Completely Automated Public Turing test to tell Computers and Humans Apart", a raíz del trabajo del guatemalteco Luis von Ahn y del venezolano Manuel Blum, entre otros. Su presentación en la conferencia internacional de criptografía Eurocrypt, que tuvo lugar en Varsovia en 2003, y el artículo titulado "CAPTCHA: Using hard AI problems for security", publicado en las actas del evento, ayudaron a popularizar el nuevo término. Sus creadores se mostraron poco optimistas de poder encontrar una solución totalmente eficaz. "No hay forma de demostrar que un programa no puede pasar una prueba que un ser humano sí pudiera pasar, ya que hay un programa, el cerebro humano, que sí lo hace", escribieron. Al equiparar las habilidades de procesamiento de los cerebros humanos con los programas de computadora, ponían en primer plano una situación muy novedosa que empezaba a marcar la época. Los antagonistas en esta guerra ya no se definían únicamente como computadoras versus humanos, sino como una serie de "adversarios", algunos con intenciones "maliciosas" y otros no, que pelaban por un lugar en el universo cibernético. A partir de entonces, el problema se vería menos como una lucha entre partes ontológicamente distinguibles que como uno de "apalancamiento" de dos lados, uno contra el otro.

Al año siguiente, Compaq ganó una patente para "restringir selectivamente el acceso a los sistemas informáticos". La compañía insistió en que "la integridad es particularmente importante si el proveedor está organizando una lotería, o realizando un concurso de popularidad o una encuesta que permita a un usuario registrar múltiples entradas". También señalaba las necesidades de la industria de la publicidad en línea ya que "los ingresos publicitarios generalmente se basan en la cantidad de veces que se muestran anuncios cuando se realizan solicitudes de servicio". Los motores de búsqueda (uno de los más populares en esos años era AltaVista de DEC) eran particularmente vulnerables al "page boosting" o promoción de la página. Para entonces, el desafío se consideraba una lucha entre "usuarios humanos reales" y "agentes automatizados [...] que operan en nombre de los usuarios", este último definido como "un programa de software o generador de scripts que puede imitar los accesos de los usuarios". La tarea de controlarlos era urgente ya que "es bien sabido que muchos agentes en la internet están diseñados intencionalmente para comportarse de manera maliciosa, destructiva o 'antisocial'".

Los expertos en estas tecnologías se encontraban en una situación en la que todos ganaban: si un acertijo en forma de CAPTCHA funcionaba, eso probaba que se había encontrado una manera de diferenciar a los humanos de las computadoras. Si, de lo contrario, el CAPTCHA fallaba, quería decir que se había sobrepasado un obstáculo más en el gran reto de desarrollar la inteligencia artificial. Yahoo, Hotmail y PayPal se apresuraron a adoptar las nuevas tecnologías. La compañía Microsoft patrocinaba las investigaciones, sabiendo que se beneficiaría de ellas.

A pesar de los logros técnicos, la tarea que la industria CAPTCHA tenía por delante se volvía cada vez más compleja. Un número creciente de practicantes empezaban a entender su desafío como el de "probar la humanidad" lo que los obligaba a desarrollar "pruebas de humanidad". Según crecía el número de usuarios en internet, los riesgos de confundir los actores involucrados, o peor aún, de favorecer a uno sobre el otro, se volvían cada vez más ominosos. ¿Qué pasaría si algún bot entraba al sistema y lo cerraba por dentro empleando algún CAPTCHA o sistema novedoso que previniera la entrada a los humanos? Durante la segunda conferencia internacional de Human Interactive Proofs (HIP) de 2005, el editor y ponente Daniel P. Lopresti se preguntó: "¿Sería posible que los usuarios reales (es decir, humanos) queden bloqueados si el sistema se convence de que las respuestas erróneas proporcionadas por cierto algoritmo son correctas?"

Entra Google

En 2009, Google adquirió una nueva prueba diseñada para mejorar el CAPTCHA tradicional, llamada reCAPTCHA y desarrollada en la Universidad Carnegie Mellon. Durante las primeras semanas de marzo de 2012, solo unos años después de que Google entrara en la competencia, la empresa implementó un nuevo sistema. De repente, las pruebas de texto distorsionado que habían caracterizado los antiguos CAPTCHA desaparecieron y ahora se pedía a los usuarios que identificaran fotografías de "números de casas extraídos de Google Street View". Un usuario que intentó hacer la nueva prueba describió su asombro al ver el cambio. "Lo que acabo de ver se parece mucho a los números de

dirección dorados atornillados sobre un revestimiento de vinilo" comunes en las casas de los suburbios americanos. Tales imágenes llevaron a sospechar a este perspicaz observador que probablemente Google tenía otra intención detrás de sus nuevas consultas. "Creo que es muy probable que Google esté decodificando los números que se encuentran en las imágenes de sus cámaras de Street View. Tal vez están intentando crear alguna característica nueva dentro de los mapas de Google", se preguntó. Después de averiguar si otros usuarios tenían experiencias similares, concluyó: "Tengo curiosidad por saber qué está pasando ahora y qué es lo que el viejo G tiene bajo la manga esta vez." Poco tiempo después, quedó claro que, en efecto, al forzar a los usuarios a identificar imágenes para entrar a ciertas páginas, Google utilizaba estas respuestas como datos valiosos obtenidos sin el conocimiento ni el consentimiento del usuario.

El uso de estas tecnologías había evolucionado tanto que la investigación ahora formaba parte del proyecto para mejorar el software de reconocimiento de imágenes. Desde 2005 los expertos reunidos en la segunda conferencia internacional de HIP ya habían llegado a una idea similar para crear CAPTCHA con el fin de "beneficiar tanto la seguridad en línea como la investigación de reconocimiento de patrones". Pocos años más tarde, Google implementaba ese proyecto en una escala masiva sin precedentes. Cuando un periodista del sitio especializado *TecbCrunch* le preguntó a la empresa sobre sus intenciones reales, Google confirmó que "actualmente estaba ejecutando un experimento donde los caracteres de Street View aparecen en CAPTCHA". Y elaboraron: "frecuentemente extraemos datos como los nombres de las calles y señales de tráfico de las imágenes de Street View para mejorar Google Maps incorporando datos tales como la ubicación y localización de negocios". Los usuarios humanos, la mayoría sin saberlo, estaban ayudando a las computadoras a mejorar su capacidad de reconocer y clasificar imágenes. Cuando las verdaderas intenciones de la empresa salieron a luz, Google admitió haber usado los resultados de reCAPTCHA en su labor para digitalizar contenido para Google Books y los archivos de Google News. Si bien "el sistema está diseñado para reducir el spam y el fraude" también contribuye a la digitalización de un texto que provenga de materiales impresos, como libros y periódicos.

El propósito original de diferenciar a los humanos de los bots pasó a un segundo plano. Vinay Shet, gerente de productos de Google para reCAPTCHA, explicó en el blog de seguridad de la compañía las razones detrás del cambio: "Los avances en inteligencia artificial han reducido la brecha entre las capacidades humanas y mecánicas para descifrar texto distorsionado." La agresividad de la pelea era tan clara como lo había sido en años anteriores, con bots considerados por sus oponentes como "atacadores".

Los programadores conocían desde hace mucho tiempo la posibilidad de utilizar los datos *de* los humanos para ayudar a las computadoras a volverse más humanas y más inteligentes. ImageNet, un proyecto de inteligencia artificial lanzado en la Universidad de Stanford, se basó en gran medida en el "crowdsourcing" para desarrollar métodos de clasificación y categorización de imágenes. Al usar miles de voluntarios para clasificar conjuntos de imágenes ampliamente disponibles en la red, estos "datos" se convertían en "alimento" para las inteligencias artificiales. Inicialmente, ImageNet reunió a cincuenta mil personas dispuestas y capacitadas para clasificar alrededor de mil millones de imágenes, que luego sirvieron para alimentar a las computadoras. En su plática TED, Fei-Fei Li, la líder del proyecto de Stanford, explicó cómo se podría lograr un progreso enorme en la inteligencia artificial ahora que "tenemos los datos para nutrir el cerebro de nuestra computadora".

¿Qué pasaría si más personas comenzaban a proporcionar alimentos para las inteligencias artificiales de forma gratuita, sin ni siquiera saberlo? ¿Cuál será el poder de quien se adueñe de esta inteligencia integrada con el cerebro de millones de incautos usuarios? ¿Qué propósitos tendría y contra quién estará dirigida?

Otro avance llegó el 25 de octubre de 2013, con el anuncio de que "reCAPTCHA ahora es más fácil (pero solo si eres humano)". Los psicólogos conocían desde hacía mucho tiempo un ingenioso truco para diagnosticar trastornos mentales. Al administrar una prueba de Rorschach basada en preguntarle a un paciente qué le sugería una imagen hecha con varias manchas de tinta, los analistas podían ignorar la respuesta si se fijaban detalladamente en el comportamiento del sujeto antes y después de la prueba. Si este comportamiento les parecía extraño, el diagnóstico se hacía con base en él. Google probó una estrategia similar. Justo al momento en que un usuario intentara acceder a alguna página web protegida, la compañía empezaba a recopilar datos sobre el comportamiento de los usuarios que se confrontarían con un CAPTCHA *antes* de que se dieran cuenta de que estaban siendo investigados. Así, la respuesta dada al CAPTCHA dejó de ser la parte central de la prueba. De hecho, se volvió poco más que una distracción para ver cómo reaccionaban los usuarios poco antes y poco después de aplicar la prueba. Google comenzó a enfocarse en la participación total del usuario con su computadora. Los límites de la "prueba" se habían borrado y los CAPTCHA se transformaron en un desafío más dentro de una gran bolsa de trucos de observación que crecía rápidamente y en donde la privacidad del usuario desaparecía a pasos agigantados. "Eso significa que hoy en día las letras distorsionadas sirven menos como una prueba de humanidad y más como un medio para obtener una amplia gama de pistas que caracterizan a los humanos y a los bots", explicó la compañía. Llegado el año 2014, el software desarrollado por Google

para reconocer imágenes y caracteres ejecutando algoritmos de aprendizaje automático ya podía acertar en reconocer el texto distorsionado de los CAPTCHA tradicionales un 99.8 por ciento de las veces. Los humanos acertaban solo un 33 por ciento. Tal discrepancia entre las habilidades de los humanos comparados con los bots hizo que los ingenieros informáticos entraran en pánico. "Por lo tanto, el texto distorsionado por sí solo ya no es una prueba confiable", concluyeron.

Cuando Google lanzó los nuevos reCAPTCHA, las secciones de comentarios de miles de páginas de internet explotaron con quejas de usuarios que estaban siendo bloqueados. Un empleador del nuevo sistema escribió furioso:

> La distinción humanos/bots de reCAPTCHA es atrozmente mala. A partir de mis horribles experiencias con este sistema, y asumiendo que muchos otros usuarios legítimos tendrán que lidiar con lo mismo, no tuve más remedio que deshacerme de reCAPTCHA por completo. No puedo someter a mis usuarios a esta tortura y asustarlos para siempre. Esto ha ido demasiado lejos. Especialmente porque escuché que hay bots, programados para solucionar CAPTCHA, que tienen una tasa de acierto mucho mayor que la mía en este momento.

Había llegado la hora de probar otro método más eficaz. Al acercarse las fiestas decembrinas de 2014, Google y otras compañías temían perder millones de clientes que se quedaran bloqueados al intentar hacer sus compras. Para prevenir tal desgracia, en diciembre de aquel año, las pruebas de CAPTCHA cambiaron a un sistema basado en confesiones forzadas. Simplemente se les pedía a los usuarios hacer clic en un cuadro con el mensaje "No soy un robot". Si tal clic se estudiaba en combinación con los datos de comportamiento del usuario recopilados antes y después de dar la respuesta, la nueva prueba funcionaba de maravilla. Vinay Shet lo llamó "No CAPTCHA, reCAPTCHA". Orgulloso, contó cómo su equipo llegó a una idea tan sencilla como brillante: "Pensamos que sería más fácil preguntarles directamente a nuestros usuarios si son o no robots, ¡así que lo hicimos!" El nuevo sistema pronto fue adoptado por Snapchat, WordPress, Humble Bundle y varios otros. Gracias a ese *upgrade* y a partir del 3 de diciembre de 2014, el día de su lanzamiento, el acceso a usuarios humanos estuvo garantizado durante esas fechas cruciales, lo que les permitió a varios de ellos realizar "algunas compras navideñas de última hora" sin problemas.

Un llamado

Un rápido recorrido por la historia de CAPTCHA muestra cómo, durante el cambio de milenio, los científicos informáticos ya no se dedicaban solamente a crear máquinas y software con habilidades o capacidades humanas; para entonces, estas tecnologías ya podían imitarnos a la perfección. El desafío para las ciencias de la computación del nuevo milenio residía en encontrar la manera más eficaz de distinguir a los humanos de las computadoras dirigidas por bots.

Millones de usuarios han visto y respondido a los CAPTCHA sin saber casi nada sobre estos, ni sobre su funcionamiento, su propósito verdadero o su historia. ¿Qué tan importantes son los cambios sutiles que estos sistemas experimentan regularmente? ¿Qué más vemos habitualmente de lo que no sabemos mucho o nada y usamos seguido sin darnos cuenta de su impacto y sus consecuencias?

La era de CAPTCHA es un llamado a un tipo diferente de acción intelectual creativa, una dirigida a amplificar el eco casi silencioso de las transformaciones tecnológicas que nos interpelan de manera codificada, abstracta y enrarecida. Debido a que no tienen límites, estos cambios desafían la categorización en términos de disciplinas. Derriban las explicaciones tradicionales basadas en las convenciones del discurso racional donde se les considera demasiado menores para ser tomadas en serio. Por ende, escapan a la atención tanto de los humanistas como de los científicos. Sin embargo, es a través de estos cambios casi invisibles que podemos ver cómo se crean y se sostienen los pilares metafísicos más importantes de nuestra era, como los que separan a los humanos de las máquinas y la ciencia del arte.

¿Qué más está ocurriendo bajo nuestros párpados que no estamos notando? Es difícil dar con una respuesta –ya que ser humano ahora no es tan sencillo como antes. –

JIMENA CANALES es historiadora de las ciencias por la Universidad de Harvard. Sus libros más recientes son *Bedeviled. A shadow history of demons in science* (Princeton University Press, 2020) y *Simply Einstein* (Simply Charly, 2021).