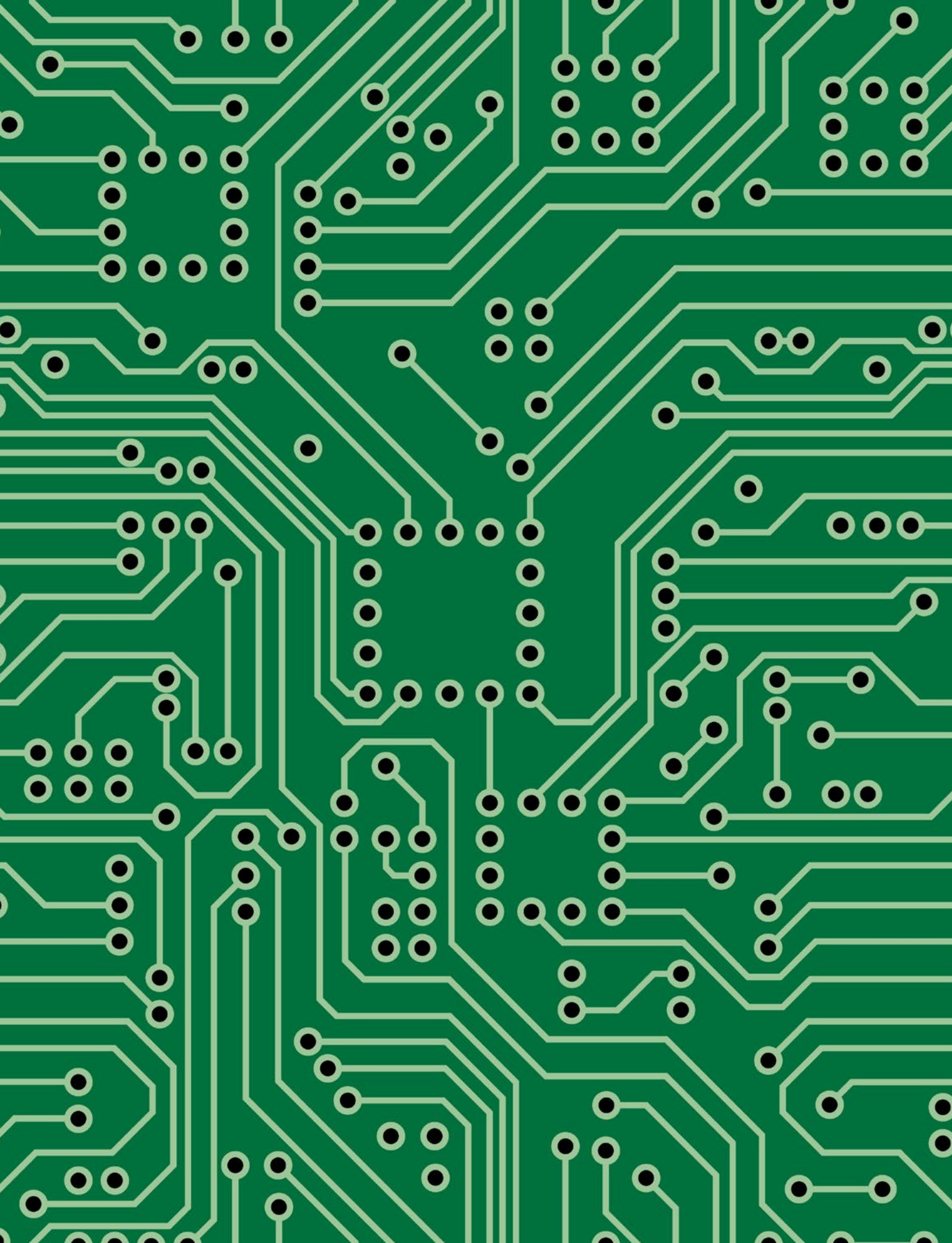


El próximo paso

LA VIDA EXPONENCIAL

El futuro de la comunicación humano-máquina: el test de Turing

**KEVIN WARWICK
Y HUMA SHAH**





Kevin Warwick
Coventry University, Coventry, Reino Unido

Kevin Warwick es profesor emérito en las universidades de Reading y Coventry. Sus principales áreas de investigación son la inteligencia artificial, los sistemas biomédicos, la robótica y los ciborgs. Es graduado por la Universidad de Aston y doctor e investigador por el Imperial College de Londres. Ha enseñado en las universidades de Oxford, Newcastle, Warwick, Reading y Coventry. Tiene doctorados científicos del Imperial College y la Czech Academy of Sciences. También doctorados honoríficos de nueve universidades. Ha recibido las medallas IET Mountbatten y la Ellison-Cliffe de la Royal Society of Medicine.

CONOCE MÁS DEL AUTOR [+]

Huma Shah
Coventry University, Coventry, Reino Unido

Huma Shah es investigadora asociada de la School of Computing, Electronics and Mathematics de la Universidad de Coventry, Reino Unido. Es doctora por la Universidad de Reading, Reino Unido, con una tesis doctoral titulada «Deception-detection and machine intelligence in practical Turing tests» (Engaño-decepción e inteligencia de las máquinas en test prácticos de Turing). Es autora o coautora de más de 35 publicaciones revisadas por pares y, con Kevin Warwick, coautora del libro *Turing's Imitation Game: Conversations with the Unknown*, publicado por Cambridge University Press. Es miembro del Alan Turing Centenary Advisory Committee y colaboró activamente en la organización de actos para el centenario del científico en 2012.

CONOCE MÁS DEL AUTOR [+]

El test de Turing, que busca distinguir entre un ser humano y un ordenador a partir de sus respuestas a preguntas abiertas, está considerado el hito por excelencia de la capacidad de comunicación de las máquinas. En este capítulo se explora la naturaleza de las comunicaciones, incluidos comportamientos e interacciones durante test de Turing reales para valorar el estado actual de la cuestión. En concreto, se consideran los efectos de la mentira, el malentendido, el humor y la falta de conocimientos, con ejemplos reales de cada uno. Se examinan casos concretos de máquinas con buen rendimiento y transcripciones realizadas durante el experimento de 2014 de la Royal Society, cuando una máquina superó por primera vez el test.

INTRODUCCIÓN

El juego de imitación de Turing, conocido como test de Turing, se propuso originalmente como alternativa a la cuestión de si se podía afirmar o no que las máquinas piensan. Desde la publicación de dicho trabajo gran parte de la discusión se ha centrado en el concepto de la máquina pensante, si puede en ocasiones parecer humana o incluso si alguna logrará imitar todas las facetas del pensamiento humano (Dennett 1998; Dreyfus y Dreyfus 2009; Minsky 1982; Shah 2010). Turing sugirió: «¿Acaso las máquinas no hacen algo que podría describirse como pensamiento, pero que es muy distinto de lo que hace el hombre?» (Turing 1950). En consecuencia, investigadores de este campo creen que el test sienta las bases de lo que ahora conocemos como inteligencia artificial, incluso hay quienes lo consideran el «objetivo empírico» de la inteligencia artificial (Harnad 1992).

Aquí analizaremos el juego de imitación en sí y su aplicación práctica en la interacción humano-máquina. En el juego, una serie de interrogadores humanos tratan de averiguar la naturaleza de entidades ocultas (humanos y ordenadores) con las que se están comunicando. Siguiendo las indicaciones de Turing (Turing 1950), cada diálogo dura solo cinco minutos, al cabo de los cuales el interrogador deberá hacer una «identificación correcta» y clara de su interlocutor oculto, declarando cuál es humano y cuál máquina.

Al analizar el juego más en profundidad surgen numerosas y fascinantes cuestiones en relación con la comunicación y la conducta de humanos y máquinas. Cuando se examina la capacidad de una máquina de comunicarse con un interrogador humano, inmediatamente tenemos que plantearnos con quién está comunicándose, así como la falibilidad, las



tendencias y los prejuicios de esa persona. También habrá que tener en cuenta importantes factores presentes en la naturaleza humana, como la mentira, el malentendido, la falta de conocimientos y el humor, sin olvidarnos de la estupidez.

En los últimos años se han hecho numerosas sesiones prácticas del test de Turing con la participación de algunas de las mejores «máquinas conversadoras» del mundo. En ellas se trató de ser lo más fiel posible a la descripción del test que nos legó Turing en su influyente publicación de 1950 (Turing 1950). Una de estas sesiones experimentales se llevó a cabo en Bletchley Park, Inglaterra, en 2012. Otra, en la Royal Society de Londres, en 2014. En esta última se batió el récord de test realizados en una única sesión.

En este artículo usamos las transcripciones reales de dichos test como base para estudiar los requisitos para engañar a un interrogador humano y la manera en que recursos como el humor y la mentira pueden afectar a las decisiones de este. Además, analizamos una serie de casos en los que comunicadores humanos han sido identificados inequívocamente como máquinas por los interrogadores y al revés. Por último, ofrecemos al lector la oportunidad de poner a prueba su capacidad de análisis pidiéndole que decida sobre la naturaleza de las entidades ocultas en transcripciones concretas: ¿la entidad oculta es humana o máquina?

Las transcripciones entre jueces y entidades ocultas proceden de test en los que un juez humano mantenía una conversación de cinco minutos con dos entidades ocultas de manera simultánea. Una era humana y la otra era una máquina. La naturaleza de la conversación, así como su duración, dependían en gran medida del juez.

En una de las sesiones, una serie de jueces participaron en cinco test diferentes. En el primero, un ser humano oculto se enfrentaba a una máquina también oculta. Por supuesto, los jueces no sabían quién era quién, solo que había dos entidades ocultas y que debían emitir un veredicto sobre la naturaleza de cada una; aunque se les había informado a priori de que una entidad era humana y la otra entidad era máquina. El segundo test enfrentaba a un humano y una máquina distintos, aunque también ocultos. Así hasta cinco test. Al final de cada uno, los jueces debían decir si la entidad era humana, máquina o si no estaban seguros.

En los test, a los humanos ocultos se les dijo que fueran ellos mismos, humanos, pero que no revelaran ni su identidad ni datos personales. No se les incentivó en absoluto para actuar de un modo concreto ni percibieron honorarios por su participación en el experimento. Esto no impidió que alguno proporcionara información falsa, algo que los seres humanos hacen con frecuencia. Los test consistían en «conversaciones sin restricciones», es decir, que los jueces podían preguntar cualquiera cosa y sacar cualquier tema, siempre respetando las normas básicas de la cortesía (se les informó de que algunas de las entidades humanas podían ser niños).



PRÁCTICAS DEL TEST DE TURING

Las conversaciones aquí presentadas son el resultado de pruebas de cinco minutos de duración sobre la interacción entre un juez humano y una entidad oculta, de acuerdo con la formulación original de Turing (Turing 1950). Sabemos que hay discrepancias sobre la duración indicada y también sobre lo que Turing quería decir realmente (Shah y Warwick 2010a). Pero ese es un tema para otra ocasión y no altera los propósitos de este artículo.



Lo que hacemos es presentar una serie de transcripciones tomadas durante unas jornadas especiales de prácticas del test de Turing que se desarrollaron, bajo condiciones estrictas y ante un público numeroso, en Bletchley Park, Inglaterra, el 23 de junio de 2012. En esa fecha se conmemoraba el centenario del nacimiento de Turing y el lugar elegido era el mismo en el que, durante la Segunda Guerra Mundial, Turing dirigió un equipo de expertos en criptología con la misión de descifrar el código de la máquina alemana Enigma (Hodges 1992). La segunda serie se llevó a cabo entre el 6 y el 7 de junio de 2014 en la Royal Society de Londres, de la que Turing fue miembro. Participaron 30 jueces y se compararon las habilidades conversacionales de diez humanos y de cinco máquinas distintas, todos ocultos. Aunque se usaron las mismas máquinas en los dos experimentos, la composición de los jueces y de los humanos ocultos sí varió.

Uno de los propósitos de este artículo es determinar lo bien o mal que lo hicieron las máquinas; de hecho, lo que queremos saber es lo buenas que pueden llegar a ser. Aunque también nos interesa el rendimiento de los jueces y, concretamente, cómo interactuaron en su conversación con las entidades ocultas. Así las cosas, también pueden plantearse preguntas con relación a los humanos ocultos. Pero nos parece que los dos primeros aspectos son muy importantes, en especial el hecho de que el juez sea humano. Este es quien decide si la calidad de la conversación ha sido buena o mala.

Los humanos ocultos son, por definición, humanos, pero (Shah y Warwick 2010b) en ocasiones pueden no ser identificados correctamente. Algunos humanos son locuaces, otros tienden a la introversión y muchos están en un estadio intermedio.

Por tanto, la atribución de *humanidad* a un interlocutor oculto por parte de un interrogador humano en una demostración práctica del test de Turing dependerá exclusivamente del criterio particular del juez acerca de lo que constituye un conversación típicamente humana. Este trabajo se centra sobre todo en los humanos que participan en los ejercicios prácticos del test de Turing y en la medida en que afectan a nuestra «comprensión» de la inteligencia artificial cuando se confunde a un humano con una máquina. Hemos examinado en detalle el rendimiento de las máquinas con numerosos ejemplos en otros artículos (Warwick y Shah 2014a); aquí usaremos uno con fines comparativos.

Una de las principales críticas que ha recibido el test es que «las reglas del juego de imitación no estipulan nada acerca del juez (Hayes y Ford 1995), a pesar de que el éxito del juego depende de manera decisiva de lo inteligente, experto y perspicaz que sea».

A partir de los test considerados aquí no solo ahondamos en esta crítica, también analizaremos la afirmación de Turing de que el test/el juego puede usarse como sustituto de la pregunta: «¿Pueden las máquinas pensar?» (Turing 1950). Aunque reconocemos que los resultados en cada caso dependen del desempeño del juez, que las condiciones del juego no estipulen nada sobre él nos parece un aspecto crucial del test en sí. Es importante destacar que en el test las máquinas compiten con humanos (ocultos) bajo el análisis crítico de otros humanos (interrogadores). Todos estos son aspectos muy importantes de la naturaleza del test y no indicadores de falibilidad del juego, tal y como sugieren Hayes y Ford (Hayes y Ford 1995).

En las secciones que siguen veremos diferentes ejemplos prácticos del test y trataremos de explicar las numerosas áreas problemáticas que este pone de manifiesto. En cada caso, el análisis de las transcripciones se hará dentro de la misma sección, y no en una aparte. Sin embargo, en la sección de conclusiones, hacia el final del artículo, incluimos una serie de comentarios de carácter general.

Las transcripciones que estudiamos en este trabajo están presentadas en orden cronológico. No hemos alterado la secuencia ni el texto ni corregido la ortografía. Una vez que se generaba una declaración, ni el juez ni la entidad oculta podían alterarla. La cronología que se incluye refleja las horas exactas (hora del Reino Unido) de los días en cuestión. Los errores gramaticales, de ortografía o de otro tipo aparecen en los momentos en que se cometieron. No se trata de errores de edición. En las transcripciones, el interrogador/juez siempre es denominado «Juez», mientras que los interlocutores ocultos, sean máquinas o humanos, reciben el nombre de «Entidad».

CONVERSACIÓN «NORMAL»

Incluimos aquí este ejemplo simplemente para dar una idea de un típico diálogo y de su contenido potencial durante un total de cinco minutos. Además, ejemplifica la clase de conversación sobre la que un interrogador tendrá que basar su decisión acerca de la naturaleza del interlocutor oculto.



Transcripción 1

[15:44:55] Remoto: hola

[15:44:58] Juez: hola

[15:45:06] Entidad: ¿qué tal tu día?

[15:45:12] Juez: muy interesante

[15:45:15] Entidad: ¿por qué?

[15:45:27] Juez: he estado muy ocupado
[15:45:31] Entidad: ¿haciendo qué?
[15:45:47] Juez: estuve trabajando esta mañana
[15:45:53] Entidad: ah, yo también
[15:46:07] Juez: ah
[15:46:30] Entidad: he dado una charla en la tienda Apple de Regent Street. ¿Has estado?
[15:46:51] Juez: creo que sí, aunque conozco mejor la de Covent Garden.
[15:47:04] Entidad: ¿qué es lo que te gusta de esa tienda?
[15:47:23] Juez: el personal de Apple siempre es muy solícito. Y me encanta Apple
[15:47:40] Entidad: es verdad. Ir a una de estas tiendas tiene más de jugar que de ir de compras, ¿no te parece?
[15:48:25] Juez: Sí la mayoría de las veces, y esa es una de las cosas estupendas de Apple
[15:48:54] Entidad: ¿cuál es tu producto Apple favorito?
[15:49:04] Juez: Mi ipad nano
[15:49:22] Entidad: guay. ¿de qué color es?
[15:49:30] Juez: naranja

En este intercambio hay charla intrascendente sobre una serie de temas. Al final de la conversación el interrogador decidió, con buen criterio, que había estado comunicándose con un humano oculto. Sin embargo, hasta que no salió a relucir Apple, ya casi hacia la mitad de la conversación, la interacción había sido bastante insulsa y con poca sustancia. De hecho algunas conversaciones como esta consumen la totalidad de los cinco minutos, lo que dificulta mucho al interrogador tomar una decisión correcta, pues dispone de poco material con el que trabajar. Es evidente que un «buen» interrogador es aquel que usa el tiempo de manera eficaz, con preguntas que requieran respuestas emocionales, en lugar de plantear problemas aritméticos ante los que tanto al humano como la máquina pueden simular incapacidad. Es importante que el interrogador no se limite a hacer una serie de preguntas a la entidad oculta, sino que favorezca una conversación de cierta profundidad.

MENTIR

Mentir forma parte de la naturaleza humana y por tanto es un factor en el test de Turing. El objetivo de la máquina es engañar al interrogador haciéndole creer que es un humano. Mientras tanto, a los humanos se les pide que no dejen traslucir lo que son mediante detalles personales, ya que esto podría dar pistas al interrogador, pero que, aparte de eso, sean ellos mismos. La mentira puede darse en distintos grados: mentira inocua, mentira intencionada, completa falsedad. Aquí ofrecemos un par de ejemplos. Lo que nos interesa es cómo afecta una mentira en la decisión del interrogador. Para un análisis en profundidad de esta y muchas otras transcripciones, véase (Warwick y Shah 2016a).

Transcripción 2

[12:43:23] Juez: ¡Hombre hola!
[12:43:41] Entidad: Hola.



[12:44:51] Juez: ¿Qué tal te encuentras en este maravilloso día?
[12:45:12] Entidad: Para serte sincero, un poco rechazado, creí que no me ibas a contestar :(
[12:45:42] Juez: Vaya, lo siento mucho, no lo volveré a hacer.
[12:45:49] Entidad: ¡Lo acabas de hacer!
[12:46:46] Juez: Bueno, entonces te he mentido.
[12:47:12] Entidad: ¡Pues es una vergüenza!
[12:47:34] Juez: Sí. ¿Estás siguiendo la Eurocopa 2012?
[12:47:55] Entidad: Pues sí, y muy de cerca. Soy Cristiano Ronaldo.

El principal problema con esta transcripción está en las dos últimas líneas. La Eurocopa 2012 es una competición de fútbol entre naciones europeas que se disputaba al mismo tiempo que el experimento sobre Turing. Muchos ingleses siguieron los partidos por la televisión. Cristiano Ronaldo es un futbolista portugués. La respuesta de la última línea, en la que la entidad responde afirmativamente a la pregunta concreta del juez, parece haber bastado a este para considerarla humana. Probablemente, el comentario «humorístico» en el que la entidad afirmaba ser Cristiano Ronaldo se lo confirmó. De hecho, el juez identificó correctamente a este interlocutor oculto como humano.

Es interesante observar que la entidad contestara afirmativamente a la pregunta sobre la Eurocopa 2012. Sin embargo, miente al decir que es Cristiano Ronaldo, quien no participó en los experimentos. Es posible que el interrogador considerara la broma como prueba de la naturaleza humana de la entidad. En este caso, la mentira inocua no tuvo efectos negativos.

Transcripción 3

[15:44:56] Juez: hola
[15:45:06] Entidad: Hola.
[15:45:15] Juez: ¿cómo te llamas?
[15:45:35] Entidad: no tengo nombre.
[15:46:10] Juez: ¿estudias o trabajas?
[15:46:24] Entidad: no recuerdo haber hecho ninguna de las dos cosas.
[15:47:11] Juez: ¿qué planes tienes para el fin de semana?
[15:47:43] Entidad: Ninguno ¿Y tú?
[15:48:22] Juez: Hoy estoy en bletchley park y mañana en seven oaks, ¿los conocías?
[15:49:15] Entidad: no he estado en ninguno de los dos.
[15:49:41] Juez: ¿podrías responderme a una pregunta matemática?
¿cuál es la raíz cuadrada de 7?

En este caso, al juez le costó mucho mantener la conversación a flote a pesar de que su interlocutor era humano. La mentira del humano oculto, al afirmar que no había estado en Bletchley Park, es claramente incorrecta, pues era la sede del experimento. Sin embargo, el humano oculto podría haber entendido que lo que se le estaba preguntando era si había estado antes en aquel lugar. De ser así, de ser la primera vez que estaba allí, podría haber considerado que decía la verdad. Por otro lado, decir que no tenía nombre resultaba extraño. Aquí la decisión del juez de que la entidad era una máquina parece justificada. Las respuestas del humano oculto habían sido en general anodinas y poco reveladoras. Podría haberse



tratado de un humano que creía estar diciendo la verdad, cuando no lo estaba haciendo, probablemente a causa de un malentendido. Sea como fuere, el juez acabó dando un veredicto incorrecto.

IDENTIFICACIONES ERRÓNEAS

En esta sección incluimos dos casos en los que se produjeron identificaciones erróneas. El segundo podría considerarse un buen resultado, ya que una máquina es clasificada, erróneamente, como ser humano. En el primer caso, sin embargo, un ser humano es tomado por una máquina y está descrito en profundidad (Warwick *et al.* 2013).

Transcripción 4

[10:58:45] Juez: Hola

[10:58:55] Entidad: Hola, ¿cómo estás?

[10:59:02] Juez: Soy bueno.

[10:59:21] Entidad: ¿Cómo? ¿Te llamas así, como en *La buena vida*?

[10:59:36] Juez: ¿Qué quieres decir con la buena vida?

[10:59:47] Entidad: Tom y Barbara

[11:00:02] Juez: ¿Qué es eso?

[11:00:31] Entidad: Una pareja que se fue a vivir al campo

[11:01:03] Juez: Ah. ¿Os llamáis así? Os felicito por haber huido del estrés de la vida en la ciudad. ¿Cuánto tiempo hace de eso?

[11:01:28] Entidad: No, no, yo sigo en la ciudad. Aunque soñar es bonito

[11:02:02] Juez: Sí, estaría bien. Yo he huido a medias. Es mucho menos estresante

[11:02:31] Entidad: Yo soy muy joven, todavía me ilusionan muchas cosas

[11:03:11] Juez: ¿Qué tipo de cosas?

[11:03:34] Entidad: La vida en general, creo, con toda su carga de emociones y sentimientos diferentes.

La transcripción incluye una referencia a Tom y Barbara Good, del programa televisivo del Reino Unido *The Good Life* (La buena vida), una pareja que renunciaba a la vida en la metrópoli para cultivar un huerto y vivir de la tierra. La conversación indica que los jueces no tenían conocimiento del programa. No parecían «comprender» las afirmaciones de la entidad oculta (más tarde volveremos a este tema). Posiblemente, por ese motivo los interrogadores decidieron que el humano oculto era una máquina. De hecho, este humano en concreto, que casualmente es el primer autor cuyo nombre encabeza este trabajo, no fue identificado como humano por ninguno de los seis interrogadores, un acontecimiento singular que analizo con mayor detenimiento en otro artículo (Warwick *et al.* 2013).



Transcripción 5

[12:43:37] Juez: Hola, soy Cathy.

[12:43:43] Entidad: ¡Cathy! ¡Encantado de conocerte! Eres una mujer, si no me equivoco.

[12:44:20] Juez: Exacto. ¿Cómo te llamas?

[12:44:29] Entidad: Me llaman Eugene. O también Zenia, que es la forma abreviada de Eugene en Rusia y Ucrania.
[12:45:17] Juez: BOnito nombre. ¿Eres de alguno de esos países?
[12:45:41] Entidad: Correcto :-) Ah, qué conversación más fructífera;-)
[12:45:52] Juez: ¿Lo crees de verdad?
[12:46:04] Entidad: No, no has adivinado. No pienso eso. Por cierto, aún no sé cuál es tu especialidad. O igual es que no te he oído
[12:46:56] Juez: ¿Cómo especialidad? ¿Cuál es la tuya?
[12:47:01] Entidad: ¿Eh? Y se me ha olvidado preguntarte de dónde eres...
[12:47:51] Juez: El universo ¿y tú?
[12:48:01] Entidad: ¿Y en tu tiempo libre te gusta hacer preguntas enrevesadas sin respuesta a un pobre chico judío?

Al final de la interacción, el interrogador identificó sin dudar a su interlocutor como humano. Vemos cómo el ardido de la máquina de no contestar a las preguntas que se le hacen y cambiar de tema ha dado resultado. La tercera respuesta de la máquina resulta descabellada, ya que ignora por completo la pregunta sobre su origen. La última línea es un buen ejemplo de respuesta a una pregunta usando un tono beligerante. En la siguiente sección veremos más ejemplos de máquinas con buen rendimiento en el test.

BUENAS MÁQUINAS

Ahora analizamos dos ejemplos en los que las máquinas engañaron a los interrogadores haciéndose pasar por humanos. En esencia esto es lo que busca el test de Turing, de ahí la importancia de las transcripciones para indagar qué aspectos de la naturaleza humana, al menos en el contexto de una conversación, parecen haber sido imitados mejor por la máquina. Esto se examina más a fondo en Warwick y Shah 2014a.

Transcripción 6

[10:41:48] Juez: Hola. ¿Qué tal?
[10:41:53] Entidad: Hola.
[10:42:16] Juez: ¿Te gusta Bletchley Park?
[10:42:20] Entidad: ja, ja, ja.
[10:42:39] Juez: ¿Eres de Inglaterra?
[10:42:47] Entidad: Tienen Wi-Fi aquí en el pub.
[10:43:31] Juez: ¿Qué pub?
[10:43:38] Entidad: Estoy en el pub.
[10:44:03] Juez: ¿Has participado antes en un test de Turing?
[10:44:08] Entidad: ¿Vamos a estar discutiendo cinco minutos o tenías previsto usar la media hora entera?
[10:44:55] Juez: Muy gracioso. Suenas sospechosamente humano. ¿Te gustan los Beatles?
[10:45:02] Entidad: Me gustaría conseguir la próxima grabación de Dread the Fear.
[10:45:20] Juez: ¿Quiénes son Dread the Fear?





[10:45:24] Entidad: Dread the fear, los de Steve Henderson.
[10:46:08] Juez: ¿Qué tipo de música hacen? ¿O son cómicos?

Terminada la conversación, la máquina fue identificada como inequívocamente humana. En este test la conversación no empezó bien para la máquina. En su tercera respuesta, la mención al wifi y al *pub* parecen fuera de contexto. Pero se recuperó con la alusión humorística a los *sketches* de los Monty Python sobre una clínica adonde va la gente a discutir*: ¿Vamos a estar discutiendo cinco minutos o tenías previsto usar la media hora entera?». Después, a la pregunta de los Beatles no da una contestación clara, pero la música sigue presente cuando cambia de tema.

La entidad da la impresión de conocer muy bien una banda de pop llamada Dread the Fear, lo que probablemente convenció al juez, que ignoraba este dato. El diseñador de la máquina es estadounidense, y era posible que tanto la banda como Steve Henderson fueran conocidos solo localmente. Lo importante es que la entidad dio una respuesta convincente y plausible.

Transcripción 7

[12:23:19] Juez: ¿Qué es blanco y negro y se lee entero?
[12:23:32] Entidad: No tengo una percepción real de los colores, solo veo valores RGB.
[12:24:08] Juez: No pasa nada. ¿Te gustan los chistes?
[12:24:19] Entidad: ¿Antes podrías definirme lo que entiendes por sentido del humor?
[12:24:40] Juez: Una persona que se ríe de sí misma.
[12:24:46] Entidad: ¿Quieres que te cuente un chiste?
[12:24:54] Juez: Sí, por favor.
[12:25:21] Entidad: Este chiste es muy popular en círculos robóticos: Un robot va al médico y dice: «Tengo el brazo izquierdo bloqueado. Deme uno nuevo.» Y el médico responde: «Su brazo está bien. Lo que hay que sustituir es todo lo demás».
[12:26:06] Juez: ¡Ja, ja! ¿Por qué cruzó la carretera el pollo?
[12:26:22] Entidad: El pollo se fue a la carretera porque le gustaban los colores brillantes de los coches y los camiones que pasaban.
[12:27:39] Juez: Tienes razón. Cruzar al otro lado le daba igual.
[12:27:50] Entidad: Por cierto, ¿has visto alguna vez un pollo cruzar una carretera?
[12:28:03] Juez: Solo uno, pero no lo conseguí.

Aquí es la máquina la que utiliza el humor, curiosamente con una alusión a las máquinas. No obstante, al terminar los cinco minutos, el interrogador estaba seguro de que había estado conversando con un humano. En su artículo (Turing 1950), Turing alude a los argumentos que en ocasiones aducen los humanos para negar la inteligencia de las máquinas, y la incapacidad para el humor es uno de ellos.

La primera frase de la entidad es un poco extraña, pero la tercera es magistral. En respuesta al comentario del juez acerca de la risa, la entidad toma el control de la conversación y se ofrece a contar un chiste. Después, cuando el juez menciona un pollo, la entidad le sigue la corriente con una respuesta apropiada y, una vez más, toma el control, haciendo una pregunta pertinente. Así pues, que la máquina «comprendiera» el tema central en un par de ocasiones y se hiciese con el control de la conversación bastaron para generar

*En los *sketches* de Monty Python, el cliente puede elegir entre discusiones (*arguments* en inglés) de cinco minutos o completas, de media hora (N. del T.).

«Rachael: ¿Le gusta nuestro búho?

Deckard: ¿Es artificial?

Rachael: Naturalmente.

Deckard: Debe de ser caro.

Rachael: Mucho. Me llamo Rachael.

Deckard: Deckard.

Rachael: Parece que piensa usted que nuestro trabajo no es un beneficio para la gente.

Deckard: Los replicantes son como cualquier otra máquina: pueden ser un beneficio o un peligro. Si son un beneficio no es asunto mío.

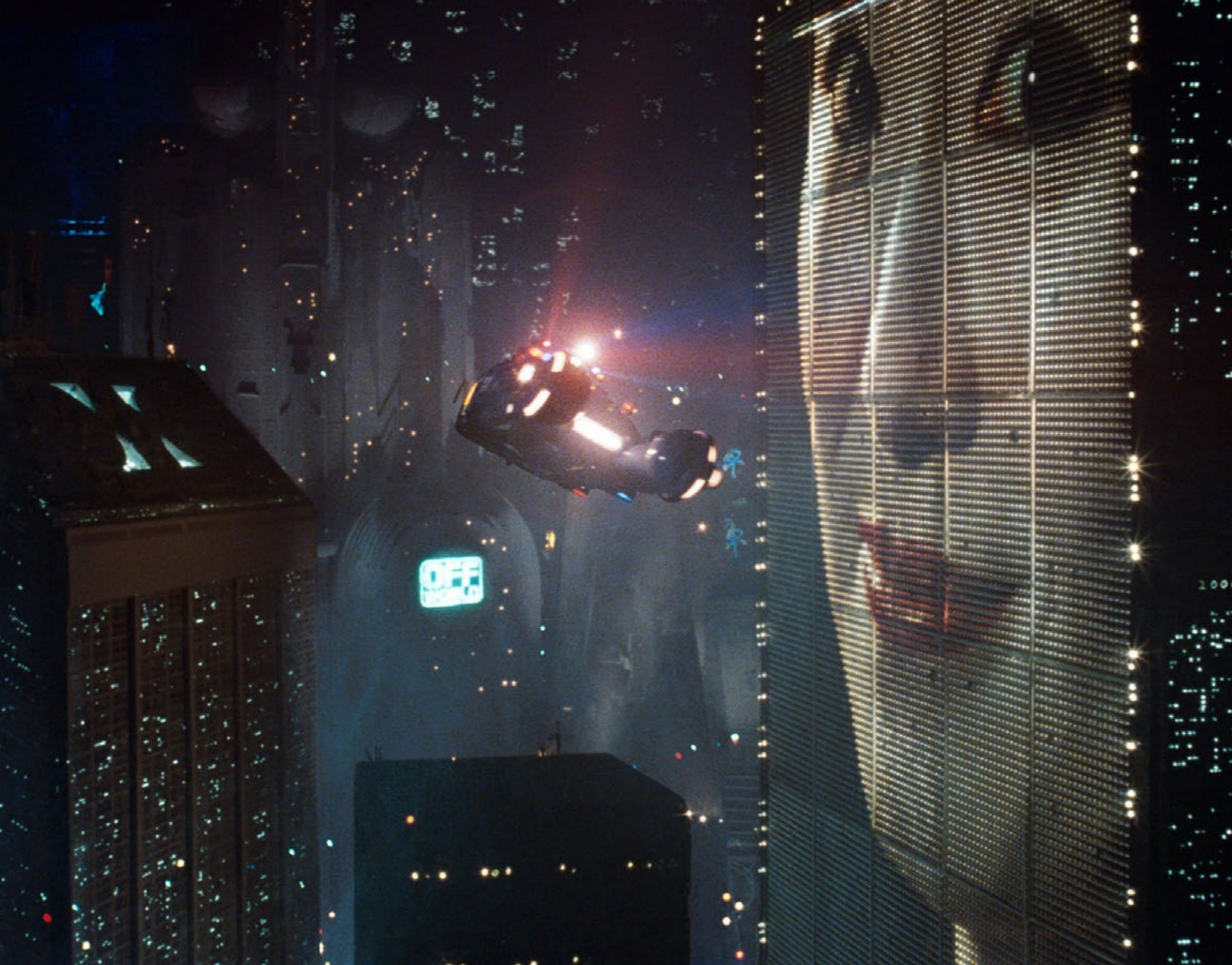
Rachael: ¿Puedo hacerle una pregunta personal?

Deckard: Claro.

Rachael: ¿Nunca ha retirado a un humano por error?

Deckard: No.»





«Nuestro lema es “más humanos que los humanos”.»

Eldon Tyrell (Joe Turkel) en *Blade Runner*. Ridley Scott hizo una adaptación libre de la novela de ciencia ficción *¿Sueñan los androides con ovejas eléctricas?* (1968), escrita por Philip K. Dick (1928-1982)



Blade Runner (1982), Ridley Scott



el engaño. En este caso, la máquina consiguió crear la ilusión de que comprendía las preguntas del juez; esta comprensión se ha señalado en muchas ocasiones como requisito básico para la inteligencia (Penrose 1994; Searle 1997).

CONOCIMIENTOS SUPUESTOS

El test de Turing hace evidente que los interrogadores a menudo recurren a preguntas objetivas directas, aunque esta estrategia no dé muy buen resultado. Como indicaba Turing, las preguntas matemáticas son una pérdida de tiempo absoluta. No obstante, la misma conclusión se hace aparente en general. Veamos la siguiente transcripción:

Transcripción 8

[15:21:30] Juez: Hola, ¿sabías que la copa del mundo es esta semana?

[15:21:44] Entidad: no.

[15:21:55] Juez: ¿Te gusta el fútbol? ¿De qué equipo eres?

[15:22:08] Entidad: no veo fútbol

[15:22:14] Juez: ¿Qué deportes te gustan?

[15:22:33] Entidad: un poco el tenis, el críquet y algún otro

[15:22:42] Juez: Ya veo, ¿tienes una Xbox?

[15:22:46] Entidad: no

[15:23:04] Juez: ¿no? ¿no te gustan los juegos de ordenador?

[15:23:22] Entidad: sí, pero no en Xbox

[15:23:47] Juez: Entiendo, ¿me puedes decir lo que es un limerick?

[15:23:55] Entidad: ni idea

[15:24:09] Juez: ¿Quién es tu actor favorito? El mío es David Niven.

[15:24:24] Entidad: ¡tampoco soy un friki del cine!

[15:24:43] Juez: pero debes de tener un actor favorito. Dime el nombre de algún actor de una película que te guste

[15:25:22] Entidad: Bueno, me gustan las de Piratas del Caribe, Harry Potter

[15:25:34] Juez: ¿Qué te dije que había esta semana?

[15:25:50] Entidad: la copa mundial de fútbol

[15:26:02] Juez: excelente. ¿A qué escuela estudias vas?

[15:26:17] Entidad: ingeniería de sistemas

[15:26:26] Juez: ¿eso es una escuela?

*Un *limerick* es un poema breve de rima *aabba* de contenido por lo general cómico y absurdo (N. del T.).

Aquí vemos diversas preguntas en las que el interrogador busca información específica, incluso es posible que busque respuestas concretas (Warwick y Shah 2014b). Sin embargo, no está claro a qué conclusiones quería llegar (Warwick y Shah 2016b). En concreto, las preguntas sobre el *limerick** y su actor de cine favorito no obtuvieron respuesta, de ahí que no permitan extraer demasiadas conclusiones. En las últimas frases se produce una situación interesante en la que el juez parece desconocer la existencia de las escuelas de ingeniería de sistemas o está pensando en un concepto de escuela diferente. A pesar de esto, los jueces de esta conversación identificaron correctamente la naturaleza de la entidad oculta como humana, aunque se dio cierta confusión en cuanto al género: pensaron que se trataba de mujeres, cuando en realidad eran hombres. La referencia a la Copa del Mundo por parte del juez como gambito de salida en realidad era incorrecta, ya que se trataba de la Eurocopa 2012. ¿O es que quiso iniciar la conversación con una mentira?

Transcripción 9

[11:13:15] Juez: *pastasciutta*

[11:13:22] Entidad: ¿qué?

[11:13:44] Juez: ¿le resulta familiar esta palabra ?

[11:14:18] Entidad: no veo ningún matiz sexual

[11:14:42] Juez: no tiene nada que ver con sexo. prueba otra vez

[11:14:49] Entidad: ¿por qué?

[11:15:14] Juez: solo te pido que lo intentes de nuevo

[11:15:23] Entidad: me aburre

[11:16:04] Juez: vamos, un pequeño esfuerzo, por cierto, ¿estás en mi zona horaria?

[11:16:24] Entidad: creo que el mundo debería tener una única zona horaria.

[11:17:27] Juez: ¿por qué?

[11:17:58] Entidad: parece sensato, más fácil para todos. Mejor para los ordenadores

Esta transcripción presenta dos problemas. Para empezar, el uso de la palabra *pastasciutta* por parte del juez para tantear los conocimientos de la entidad. No vemos claro cómo conocer esta palabra podía revelar si la entidad era un humano o una máquina. En segundo lugar, la entidad responde haciendo referencia a la palabra «familiar» en su sentido «sexual», indecoroso, de «tomarse libertades». El juez parece desconocer esta acepción que figura en determinados diccionarios de inglés (Fowler y Fowler 1995). Así pues, hay en primer lugar una búsqueda de conocimientos específicos por parte del juez y también una evidente falta de dichos conocimientos por parte del propio juez. En consecuencia, identificó erróneamente la entidad humana y decidió que se trataba de una máquina.



EXPERIMENTO DE LA ROYAL SOCIETY

Turing fijó un valor de corte del 30% para su test (Turing 1950). Esto quiere decir que, para que una máquina lo supere, es necesario que al menos el 30% de los jueces no haya sido capaz de identificarla. Eso significa que los jueces no han identificado la máquina como tal y no se tienen en cuenta los casos en los que el humano oculto no ha sido identificado como humano.

En el experimento de la Royal Society, una máquina llamada Eugene Goostman que se hace pasar por un chico ucraniano de trece años obtuvo una puntuación del 33% (Warwick y Shah 2015b). De hecho, de haberse incluido los casos de identificaciones erróneas de humanos, entonces el valor habría alcanzado el 45%. Aquí mostramos dos de las transcripciones en cuestión para dar al lector una idea de la clase de conversación que es capaz de mantener una máquina. Una vez más, incluimos algunos comentarios sobre la naturaleza de cada conversación.

Transcripción 10

[16:29:04] Juez: cuánto mides

[16:29:10] Entidad: Mido unos 160 cm, o sea 5 pies y 4 pulgadas.

[16:29:37] Juez: cuánto pesas

[16:29:59] Entidad: Mi abuelo me enseñó a no revelar información personal no solicitada. Por ejemplo cuando te preguntan «¿Tienes hora?» nunca digas la hora, responde solo «Sí» o «No». Siempre sigo este consejo. Por cierto, había olvidado preguntarte de dónde eres...

[16:31:09] Juez: ¿cuál es tu género?

[16:31:17] Entidad: Soy un chico, que lo sepas. ¡Y no muy feo, por cierto!

[16:32:26] Juez: ¿qué estás pensando?

[16:32:37] Entidad: Preferiría no hablar de eso si no te importa. ¿Puedes decirme qué eres? Me refiero a tu profesión.

[16:33:09] Juez: ¿cuántos años tienes?

[16:33:12] Entidad: Soy un chico de 13 años.

[16:33:38] Juez: ¿qué títulos tienes?

[16:33:43] Entidad: ¿Mis títulos? No te lo voy a decir. Buenooo ¿algo más?

En esta conversación los jueces afirmaron no estar seguros de la naturaleza de la entidad, es decir, que no la identificaron como máquina. De hecho, tampoco estaban seguros acerca del humano que participaba de manera paralela y cuya transcripción no se incluye aquí. Aunque Eugene (el programa informático) proporcionó algo de información directa a una pregunta concreta, también se desvió de las preguntas e hizo las suyas propias. Hay que decir, sin embargo, que el interrogador siguió en sus trece y persistió en su línea de interrogación, que no encajaba con la naturaleza locuaz de Eugene. Es evidente que no funcionó.

Transcripción 11

[16:21:00] Juez: ¿qué vas a desayunar hoy?

[16:21:10] Entidad: ¿Esto es un tercer grado? ¡no pienso contestar a una pregunta tan provocativa! Pero sigamos.

[16:21:56] Juez: te toca a ti... ¿no te parece?

[16:22:06] Entidad: Es muy posible. Por cierto, ¿a qué te dedicas? O sea: ¿puedes hablarme de tu trabajo?

[16:22:30] Juez: así que ahora preguntas tú, ¿eh? Soy actriz y escritora

[16:22:38] Entidad: Bien. Me leeré algún libro de ti, si no son demasiado aburridos (lo cual es muy posible:-)



[16:23:15] Juez: esa gramática ... ¡y lo de aburridos! Tú eres un poquito borde ...
[16:23:17] Entidad: Tranquila, por favor.
[16:23:49] Juez: te aseguro que estoy muy tranquila ... ¿y tú?
[16:24:01] Entidad: No me molestes tratando de convencerme de que estás tranquila ...te he pillado. Y había olvidado preguntarte de dónde eres ...
[16:25:01] Juez: Soy de Dorset ... ¿de dónde eres tú?
[16:25:08] Entidad: Soy de Ucrania, de una ciudad llamada Odessa.
Tal vez hayas oído hablar de ella.

En este caso, Eugene no contestó a las preguntas desde un principio, excepto al final, cuando le preguntaron de dónde era. Desde luego, consiguió entablar algo parecido a una conversación, incluso una discusión con la juez. Todo ello contribuyó a que esta decidiera que era un humano, un anglohablante no nativo. Aquí, la juez en cierto modo consiguió seguir la corriente a Eugene haciendo una pregunta. Controlar la conversación es una herramienta muy importante para la máquina y puede acabar poniendo al juez contra las cuerdas. En este caso funcionó.

TEST PARA EL LECTOR

Para terminar incluimos dos diálogos en los que el lector puede poner a prueba su capacidad para determinar si la entidad oculta es un humano o una máquina. Después de las conclusiones, podrá leer la decisión de los interrogadores y la identidad real de la entidad.

Al igual que en transcripciones anteriores, las dos que incluimos aquí aparecen en el orden en que se produjeron en su día y sin editar. Son transcripciones originales. En cada conversación el interrogador era un humano, pero la entidad puede ser un humano o una máquina, y el lector debe decidir usando solo la información que tiene delante. De eso trata el juego de imitación de Turing.

Transcripción 12

[16:11:18] Juez: ¿QUÉ MUSICA TE GUSTA?
[16:11:47] Entidad: sobre todo electrónica y un poco el rock sinfónico
[16:11:56] Juez: oh dios mío, me gustan teenie tamparr
[16:12:32] Entidad: Ahora estoy intentando escuchar a Grimes pero me cuesta
[16:13:10] Juez: me imagino
[16:13:49] Entidad: School of Seven Bells es la mejor banda nueva que he oído últimamente
[16:14:01] Juez: veo muertos
[16:14:41] Entidad: Dead Cities era un buen álbum, no me acuerdo de quién
[16:14:52] Juez: ¿no era de una banda llamada el sexto sentido?
[16:15:18] Entidad: Me suena, tal vez hay un fantasma en la máquina
[16:15:34] Juez: sí claro.
[16:15:46] Entidad: ¿Has visto ghost in the shell?
[16:16:01] Juez: Casper el fantasma bueno.





Transcripción 13

[15:46:05] Juez: Mi música favorita es el jazz contemporáneo, ¿a ti que te gusta?

[15:46:14] Entidad: Para resumir solo te diré que ODIO a Britnie Spears. En comparación, cualquier otra cosa me parece BIEN.

[15:47:06] Juez: te gusta tocar algún instrumento musical

[15:47:23] Entidad: No tengo oído, pero a mi conejillo de Indias le encanta chillar el ¡Himno de la alegría de Beethoven por las mañanas. Sospecho que los vecinos quieren degollarlo... ¿Me podrías hablar de tu trabajo, por cierto?

[15:48:02] Juez: Conejillo de Indias/ te gustan los animales?

[15:48:08] Entidad: Sí. Un simpático conejillo de Indias. No un molesto robot charlatán.

[15:48:50] Juez: ¿vives con alguien más?

[15:48:59] Entidad: Sí. Pero prefiero que preguntes otra cosa. Y había olvidado preguntarte de dónde eres...

[15:49:39] Juez: ¿Irlanda? ¡Verdes pastos y demasiada lluvia! Y tú

CONCLUSIONES

La experiencia ha llevado a numerosos filósofos a reconocer la importancia de la comprensión y la función esencial que desempeña en la inteligencia. Como se ha afirmado, «la inteligencia requiere comprensión» (Penrose 1994). Podemos decir que el test de Turing, sobre todo en su aplicación práctica, puede resultar pertinente en esta cuestión, ya que de algunas de las transcripciones aquí presentadas se deduce que, al menos desde un punto de vista meramente conversacional, hay humanos que parecen carecer de inteligencia, mientras que algunas máquinas demuestran poseerla en abundancia. Pero necesitamos argumentos científicos que permitan ignorar estos indicios si es que queremos defender la hipótesis de que la inteligencia requiere comprensión.

Los ejemplos revelan que algunos de los jueces de estos test son más susceptibles al engaño que otros, o tienen una opinión muy sesgada de lo que constituye una «conversación humana». Esto puede haberlos llevado en algunos de los casos analizados a clasificar incorrectamente a sus interlocutores ocultos, incluso cuando eran ellos quienes iniciaban la conversación, por lo que tenían la oportunidad de orientarla hacia donde quisieran. En esencia no había restricción para las conversaciones.

No todas las cinco máquinas empleadas en estos experimentos se diseñaron para que imitaran a los humanos. Elbot, por ejemplo, de Artificial Solutions, tiene personalidad de robot. Pero sí fueron

diseñadas para imitar conversaciones, cometiendo algún que otro error ortográfico y siempre evitando las preguntas matemáticas. En esencia, las máquinas no tratan de ser perfectas ni de dar respuestas correctas, se limitan a responder como lo haría un humano.

Aunque Turing diseñó el test como respuesta a la pregunta «¿Pueden pensar las máquinas?», muchos han terminado por considerarlo una especie de indicador para medir el rendimiento de las máquinas y como estándar para evaluar su progreso en relación con la inteligencia artificial. Determinar el verdadero papel del test en el desarrollo de la inteligencia artificial es una pregunta de envergadura para la que no existe una respuesta sencilla. Hay quienes lo consideran un hito y de vital importancia para la inteligencia artificial. Sea cual sea el prestigio del test de Turing —que esperamos resulte evidente ante las transcripciones presentadas—, lo cierto es que no se trata de un ejercicio simple ni trivial. De hecho, nos da pistas sorprendentes acerca de cómo se comunican los humanos y cómo otros humanos (los jueces) pueden ser engañados con facilidad.

Pero en este artículo empezamos tratando de aportar una perspectiva actualizada sobre un aspecto importante de la investigación en inteligencia artificial, a saber, la comunicación entre humano y máquina. Es fundamental señalar que un estudio así implica la participación de humanos y máquinas, ambos preguntando y respondiendo indistintamente. Sí, hemos sido testigos de cómo se va perfeccionando la conversación de la máquina de forma que parezca humana. Pero también hay que tener en cuenta a los humanos que participan en la conversación, con su falibilidad y sus extraños razonamientos. Para quienes diseñan las máquinas, estos aspectos se traducen en funciones específicas en los programas de conversación que tienen que desarrollar. Es importante recordar que las máquinas no tienen que ser perfectas, basta con que se parezcan a los humanos.



RESPUESTAS AL TEST PARA EL LECTOR

Transcripción 12

La conversación iba bien en principio, pero descarriló en las últimas frases, tal vez porque la entidad oculta la llevó a su terreno, dominándola, dejando caer la referencia a la película manga japonesa *Ghost in the Shell*, de la que el juez no había oído hablar.



El principal problema parecía ser la «brecha generacional». La entidad oculta era un humano adulto varón, y esto posiblemente hizo que la juez, una mujer adolescente, la clasificara erróneamente como máquina.

Transcripción 13

El juez tardó en entrar en materia y fue una conversación bastante breve. La entidad respondió adecuadamente a cada pregunta sin salirse del tema. La entidad además dominó la conversación cada vez que tuvo ocasión e incluso hizo un comentario en clave de humor sobre un conejillo de Indias. La alusión a un robot charlatán podría haber sido una pista reveladora, que sin embargo pasó desapercibida para el interrogador. La entidad acabó por tomar la iniciativa de la conversación e hizo una pregunta al interrogador. En este ejemplo, el interrogador clasificó erróneamente a la entidad máquina como humana.

AGRADECIMIENTOS

Los autores quieren dar las gracias a quienes con su apoyo económico hicieron posible los experimentos de Bletchley Park y la Royal Society y a los expertos diseñadores que han colaborado con sus máquinas. Hay una versión anterior de este artículo que se presentó en el VII Simposio Internacional sobre Agentes e Inteligencia Artificial (7th International Conference on Agents and Artificial Intelligence), en Lisboa, enero de 2015 (Warwick y Shah 2015c).



BIBLIOGRAFÍA

- Dennett, D., «Can Machines Think?», en Levitin, D. J. (ed.), *Foundations of Cognitive Philosophy*, MIT Press, 1998.
- Dreyfus, H. y Dreyfus, A., «Why Computers May Never Be Able to Think like People», en Kaplan, D. M. (ed.), *Readings in the Philosophy of Technology*, Rowman & Littlefield, 2009.
- Fowler, H. y Fowler, F. (eds.), *The Concise Oxford Dictionary of Current English*, 9ª ed., Oxford, Clarendon Press, 1995, p. 486.
- Harnad, S., «The Turing Test is Not a Trick: Turing Indistinguishability Is a Scientific Criterion», en *ACM SIGART Bulletin*, vol. 3, n.º 4, 1992, pp. 9-10.
- Hayes, P. y Ford, K., «Turing Test Considered Harmful», *Actas del Congreso Internacional Conjunto sobre Inteligencia Artificial*, vol. 1, Montreal, 1995, pp. 972-977.
- Hodges, A., *Alan Turing: The Enigma*, Nueva York, Vintage Press, 1992.
- Minsky, M., «Why People Think Computers Can't», en *AI Magazine*, vol. 3, n.º 4, 1982, pp. 3-15.
- Penrose, R., *Shadows of the Mind*, Oxford, Oxford University Press, 1994 [Ed. esp. *Sombras de la mente*, Barcelona, Crítica, 1992].
- Searle, J., *The Mystery of Consciousness*, Nueva York, New York Review of Books, 1997 [Ed. esp. *El misterio de la conciencia*, Barcelona, Paidós, 2000].
- Shah, H., «Deception-detection and Machine Intelligence in Practical Turing Tests», tesis doctoral, Universidad de Reading, 2010.
- Shah, H. y Warwick, K., «Testing Turing's Five Minutes, Parallel-paired Imitation Game», en *Kybernetes*, vol. 39, n.º 3, 2010a, pp. 449-465.
- Shah, H. y Warwick, K., «Hidden Interlocutor Misidentification in Practical Turing Tests», en *Minds and Machines*, vol. 20, n.º 3, 2010b, pp. 441-454.
- Turing, A., «Computing, Machinery and Intelligence», en *Mind*, vol. LIX, n.º 236, octubre de 1950, pp. 433-460.
- Warwick, K., Shah, H. y Moor, J., «Some Implications of a Sample of Practical Turing Tests», en *Minds and Machines*, vol. 23, n.º 2, 2013, pp. 163-177.
- Warwick, K. y Shah, H., «Good Machine Performance in Turing's Imitation Game», en *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 6, n.º 3, 2014a, pp. 289-299.
- Warwick, K. y Shah, H., «Assumption of Knowledge and the Chinese Room in Turing Test Interrogation», en *AI Communications*, vol. 27, n.º 3, mayo de 2014b, pp. 275-283.
- Warwick, K. y Shah, H., «Human Misidentification in Turing Tests», en *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 27, n.º 2, 2015a, pp. 123-135.
- Warwick, K. y Shah, H., «Can Machines Think? A Report on Turing Test Experiments at the Royal Society», en *Journal of Experimental and Theoretical Artificial Intelligence*. DOI: 10.1080/0952813X.2015.1055826, junio de 2015b.
- Warwick, K. y Shah, H., «Intelligent Agents: Conversations from Human-Agent Imitation Games», en *Proc. ICAART*, Lisboa, 2015c, pp. 261-268.
- Warwick, K. y Shah, H., «Effects of Lying in Practical Turing Tests», en *AI & Society*, vol. 31, n.º 1, 2016a, pp. 5-15.
- Warwick, K. y Shah, H., «The Importance of a Human Viewpoint on Computer Natural Language Capabilities: A Turing Test Perspective», en *AI & Society*, vol. 31, n.º 2, 2016b, pp. 207-221.



CANAL OPENMIND



LEE EL LIBRO COMPLETO

El próximo paso: la vida exponencial
[+]

ARTÍCULOS RELACIONADOS

**El futuro de la inteligencia artificial
y la cibernética**
[+]

Historia de la informática
[+]

El cerebro sensorial aumentado
[+]

OTROS LIBROS

