

Juego Doble: Te Quiero Machine

Emociones de Laboratorio y Electrodomésticas

Axel Arturo Barceló Aspeitia
Instituto de Investigaciones Filosóficas, UNAM

Desde mediados de la década de los noventa, el Laboratorio de Inteligencia Artificial del Instituto de Tecnología de Massachussets (MIT) ha dedicado muchos de sus recursos a desarrollar lo que en los ojos de muchos pertenece mas al ámbito de la ciencia ficción que de la ciencia: la construcción y desarrollo de robots emocionales. Pero no solo el MIT, sino también el la Sociedad de Computo en el Mundo Real del MITI (Ministerio de Industria y Comercio Internacional) y el Instituto Nacional de Ciencia y Tecnología Industrial Avanzada (AIST) del Japón han buscado, por más de una década, innovar el campo de la inteligencia artificial a través del desarrollo de máquinas emocionales. Detrás de estos proyectos descansa una nueva visión de la inteligencia (humana y artificial) donde emociones y razonamiento juegan un papel igualmente sustancial e indisoluble. Es menester, por lo tanto, detenernos un momento a preguntarnos, como sociedad, para qué queremos este tipo de máquinas con emociones.

De entre las muchas razones que se han citado a favor de la construcción de máquinas que piensen de manera inteligente, indudablemente dos han sido las que han ejercido la mayor influencia en el desarrollo de la Inteligencia Artificial (MacKenzie 2001): (i) Como modelos explicativos del pensamiento humano – es decir, para entender mejor cómo es que efectivamente pensamos los humanos – y (ii) para *ayudarnos* a pensar. Esto último se puede lograr, ya sea por que, al reducir los riesgos de error, las máquinas nos pueden ayudar a pensar *mejor* o porque, al ahorrarnos trabajo, nos permiten reducir los

recursos humanos dedicados al pensamiento. Estos dos objetivos nos ofrecen diferentes ideas sobre el tipo de pensamiento que esperamos de las máquinas. Para unos (Herbet Simon, John McCarthy, Marvin Minsky, Herbert Gelernter y Rodney Brooks, entre otros), lo que se busca son máquinas que piensen *tal y como lo hacemos los humanos*, con todo y nuestras limitaciones e imperfecciones. Para otros (Martín Davis, Hao Wang, Ever Berth, Dag Prawitz etc.), lo que se espera que las máquinas inteligentes es que superen nuestras capacidades cognitivas y piensen de manera racionalmente *perfecta*. Para este segundo grupo de investigadores, la inteligencia artificial debe encarnar una visión idealizada del pensamiento humano. De esta manera, el objetivo de la inteligencia artificial es hacer realidad el viejo sueño de Leibniz de automatizar el razonamiento lógico

Ahora que la Inteligencia Artificial ha entrado al campo de las emociones, la pregunta adquiere una nueva relevancia: ¿para qué queremos que las máquinas tengan emociones? El proyecto del MIT se enmarca de lleno en la primera tradición científica de la Inteligencia Artificial. Su objetivo es modelar artificialmente la mente humana, con todos y emociones. Según la Dra. Breazel (1999, 2002), del MIT, la Inteligencia Artificial sin Emociones Artificiales es un proyecto destinado al fracaso, ya que, según ella, el pensamiento inteligente requiere, entre otras cosas, un buen manejo de las emociones.

Existen, por lo tanto, dos proyectos independientes dentro del área de las Emociones Artificiales. El primero, al que pertenece la Dra. Breazel, busca crear lo que yo llamo emociones *de laboratorio*, es decir robots que, en el ambiente controlado del laboratorio muestren conductas que podamos reconocer como emocionales. Su objetivo es ayudarnos a entender mejor la manera en que las emociones afectan nuestro pensamiento. Como tales, las emociones artificiales son modelos explicativos de las emociones humanas. Su

importancia es fundamentalmente científica. Al igual que otros modelos científicos de laboratorio, su objetivo es ayudar al científico a entender mejor aquello que modelan. En este caso, lo que se modela es el papel de las emociones en la inteligencia humana.

El segundo proyecto, desarrollado especialmente en el Japón, busca crear máquinas que, a través de su capacidad para percibir y expresar emociones, faciliten su interacción con los humanos. Este tipo de máquinas, a las que yo llamo *Electrodomésticos Emocionales*, buscan llevar la relación entre máquinas y humanas a otro nivel de interacción. social La hipótesis que guía este segundo proyecto es muy simple: Se supone que si nuestras máquinas pudieran reconocer nuestros estados emocionales y pudieran comunicarse con nosotros a ese nivel emocional, nos sentiríamos mas *a gusto* con ellas y podríamos interactuar con ellas de manera más efectiva. A diferencia del primer proyecto, el interés fundamental no es científico, sino industrial. Según el estudio reciente de Fong, Nourbakhsh y Dautenham (2003), sus primeras aplicaciones han sido en el campo del diseño de juguetes, herramientas educativas e instrumentos terapéuticos.

Una vez más, nos encontramos con dos visiones diferentes del tipo de emociones que esperamos de las máquinas. Para los científicos del MIT, el objetivo es modelar las emociones humanas, como todas sus debilidades e imperfecciones. Para los ingenieros y científicos inmersos en el segundo proyecto, las máquinas deben mostrar solo las emociones *correctas* para facilitar su interacción con los humanos. En otras palabras, el comportamiento de los *Electrodomésticos Emocionales* deberá ser complaciente, como ha de esperarse de un aparato al servicio del uso humano. En consecuencia, las emociones artificiales han de ser una versión benigna e idealizada de las emociones humanas.

Estas dos diferentes concepciones de las Emociones Artificiales también definen de manera distinta el *éxito* de su empresa. Para los científicos cognitivos, que ven a las mentes mecánicas como modelos experimentales de la mente humana, la prueba última de que una máquina es efectivamente inteligente ha sido lo que tradicionalmente se llama la *Prueba de Turing* (bautizada así en honor del filósofo y matemático inglés Alan Turing, quien la presentó originalmente en 1950). Dicha prueba está basada en la idea de que una máquina muestra verdadera inteligencia sólo cuando no podemos distinguir su manera de pensar de la que esperaríamos de un humano. Una máquina inteligente debería simular de tal manera el pensamiento humano que, en nuestra interacción con ella, pudiéramos confundir dicho pensamiento con el de un ser humano. Para el proyecto de Emociones Artificiales, esto significaría poder reconocer en la máquina las mismas emociones que reconocemos en otros humanos.

En contraste, el éxito de un electrodoméstico emocional no puede definirse en el laboratorio, sino que ha de ponerse a prueba en el uso cotidiano de la máquina. Para poder interactuar exitosamente con un robot emocional e inteligente no es necesario *confundirlo* con un ser humano. En muchos casos, por el contrario, sería contraproducente que nuestras respuestas emocionales a un robot fueran idénticas a las que tenemos con otro ser humano. Es necesario mantener el mínimo de distancia emocional para poder *usarlo*. Después de todo, como clásicamente lo señaló Immanuel Kant (1785), dotar de humanidad a otro implica éticamente el tratarlo como un fin y nunca meramente como un medio. Hay, por lo tanto, una contradicción ética entre la herramienta y el humano. Hao Wang (1960) bien podía ver a las computadoras como esclavos inteligentes, realizando trabajos intelectuales demasiado engorrosos para el ser humano. Sin embargo, una vez que dichas máquinas

inteligentes muestran también rasgos emocionales, la metáfora del robot-esclavo empieza a dar pie a pesadillas distópicas. ¿Qué tipo de trabajo emocional podemos calificar de demasiado engorroso para ser dejado a las máquinas en vez de los humanos? En vez de *humanizar* a las máquinas, el desarrollo de este tipo de robots podría resultar en la *deshumanización* de las emociones. Si ya no podemos distinguir a los humanos de las máquinas por su manejo de emociones, podemos dejar de usar esas emociones para anclar nuestro compromiso ético con los demás. Reitero, es tiempo de preguntarnos qué tipo de sociedad está dispuesta a delegar a las máquinas el trabajo emocional que comúnmente nos liga con los demás.

BIBLIOGRAFÍA:

- Breazeal, Cynthia, (1999), “Robot in Society: Friend or Appliance”, Reporte del Laboratorio de Inteligencia Artificial del Instituto de Tecnología de Massachussets. *Proceedings of the 1999 Autonomous Agents Workshop on Emotion-Based Agent Architectures*, Seattle, Washington, pp. 18-26. Disponible en <http://www.ai.mit.edu/people/jvelas/ebaa99/breazeal-ebaa99.pdf>
- (2002), *Designing Sociable Robots*, Cambridge, Massachussets, MIT Press.
- Fong, Terrence et. al., (2003), “A Survey of Socially Interactive Robots”, *Robotics and Autonomous Systems*, 42, 143-166. Disponible en <http://www.cs.cmu.edu/~illah/PAPERS/socialroboticssurvey.pdf>.
- Kant, Immanuel, (1785), *Fundamentación de la Metafísica de las Costumbres*, Real Sociedad Económica Matritense de Amigos del País. Madrid.
- MacKenzie, Donald, (2001), *Mechanizing Proof: Computing, Risk, and Trust*. Cambridge, Massachussets, MIT Press. Una versión previa y resumida de los aspectos relevantes a este ensayo se encuentra disponible en línea como “The Automation of Proof: A Historical and Sociological Exploration” (1994) <http://129.215.101.90:1001/publications/PICTworkingpapers/WP56.pdf>
- Simon, Herbert y A. Newell, (1958), “Heuristic problem solving: The next advance in operations research”, *Operations Research*, 6, (1), pp. 1-10.
- Turing, Alan, (1950), “Computing Machinery and Intelligence”, *Mind*, 49, pp. 433-460. Disponible en <http://cogprints.org/499/00/turing.html>.

Wang, Hao, (1960), "Toward Mechanical Mathematics," *IBM Journal of Research and Development*, 4, pp. 2-21. Disponible en

<http://www.research.ibm.com/journal/rd/041/ibmrd0401B.pdf>.