

SUS  
2041



Investigación-Acción:  
métodos para  
entender territorios  
en transformación

URBANISMO CIUDADANO:  
GUIA EINSTEIN - IDEAS CLAVES

[convidalacalle.org](http://convidalacalle.org)

2021

## Problematizando las Máquinas “Morales”

Versión en español de una entrevista con el profesor Robert Sparrow, filósofo, profesor de Monash University, autor de “Why machines cannot be moral” (por qué las máquinas no pueden ser morales). La entrevista es de David Rutledge, The Philosopher’s Zone, Australian Broadcasting Corporation, domingo 28 marzo, <https://www.abc.net.au/radionational/programs/philosopherszone/the-problem-with-moral-machines/13275676>.

Catalina Olivares  
Ayudante, SUS 2041

[convidalacalle.org](http://convidalacalle.org)

### DEFINICIONES CLAVES

**Ética:** Rama de la filosofía que estudia el bien, el mal y su relación con el comportamiento humano (personal)

**Moral:** Conjunto de normas del comportamiento humano que se consideran buenas en una comunidad. (colectivo)

**Síndrome Dunning-Kruger:** sesgo cognitivo. La gente incompetente percibe su habilidad como mayor de lo que es y no valora los conocimientos, incluso mucho mayores, de otras personas. También puede ocurrir que las personas muy competentes subestiman su habilidad, sobreestimando los conocimientos y capacidades de otras personas, con menos competencia.

## Más informaciones:

Sparrow, R. Why machines cannot be moral. *AI & Soc* (2021). <https://doi.org/10.1007/s00146-020-01132-6>

Ahora que la IA (inteligencia artificial) está con nosotros, se presentan varios dilemas éticos, y para hablar de ellos tenemos un invitado poco común ya que es de los pocos filósofos dedicados a IA con estudios formales en el asunto.

## ¿Por qué es importante discutir este tema?

Normalmente los ingenieros pensarán que por estar capacitados para armar un robot, tienen la altura de mira necesaria para afrontar los problemas éticos y miran a la filosofía y la ética con arrogancia, o como si no importasen, pero pongámonos en el siguiente caso: Si un robot bancario tiene que elegir si darte un crédito o no, y para decidir discrimina por tus búsquedas de internet, o tu código postal, es un dilema ético e importante, si un robot en medicina te atiende, cómo balanceas o programas para los falsos positivos y negativos?

## ¿Cuáles son las ideas hasta ahora para cumplir con la ética y en qué fallan?

Una opción es pre programar los posibles escenarios y decisiones, pero en cuanto el robot encuentre una situación inesperada, o algo que no cuadre en su código, no sabrá qué hacer y no podrá hacer nada, ni siquiera reconocerá la situación como un problema. Otra opción es que cuando el robot encuentre estas situaciones inesperadas, que sobrepasen su código, vuelvan con su creador y reciban instrucciones humanas de cómo proceder, pero el problema con esto, es que muchas veces, las decisiones éticas deben tomarse en menos de un segundo, a veces son cosas urgentes e inmediatas y simplemente no habrá tiempo para detenerse consultar, descargar la data y ejecutar, es por eso que el contexto de la IA importa, y mucho.

## El experimento sobre la naturaleza de lo ético

Imaginemos que existe una máquina ética. ¿Dejarías que esa máquina tomase la responsabilidad de tus propias decisiones éticas? Si el dilema fuese desconectar a tu padre del soporte vital en el hospital, ¿dejarías que una máquina tomara esa decisión por ti? La ética no es algo que podamos traer desde fuera o exportar, las decisiones éticas son personales, los dilemas éticos también son personales, las consecuencias de nuestras decisiones son personales. Lo que para mi puede ser un dilema ético no necesariamente lo es para otros, por lo tanto si tenemos un dilema, no podemos escapar de nuestra responsabilidad, podemos pedir consejos, pero últimamente las decisiones son nuestras.

**En su paper dice que pedir consejos éticos desde fuera (desde otros) es no tomarse la ética en serio, pero uno podría decir que es tomárselo aún más serio, ya que reconozco que necesito ayuda de alguien más capacitado para tomar la decisión, entonces ¿Qué es realmente tomarse la ética en serio?**

Existe una gran diferencia. Si yo tengo un dilema y le pregunto a todo el mundo qué debería hacer, no me lo estoy tomando en serio, porque estoy esperando que otros me digan qué hacer, estoy delegando la responsabilidad de mi decisión y no lo estoy procesando personalmente. Distinto sería, preguntarle a alguien capacitado o buscar sabiduría que me guíe para yo mismo evaluar mi situación y tomar mi propia decisión, después de meditarlo, en ese caso si me lo estoy tomando en serio.

## ¿Cómo se explica que la ética sea a la vez subjetiva y objetiva?

porque la ética vacambando con los años, y nosotros también nos damos cuenta cuando está mal, por ejemplo, hace muchos años nos dimos cuenta de que la esclavitud estaba mal, de que el racismo estaba mal y lo corregimos dentro de lo que entendemos como ético, entonces, la ética no es una vara rígida de valores sino que puede estar correcta o equivocada. Si un modelo para lo humano o ético es científico o matemático, se vuelve más difícil incluir la ética y la cualidad humana, lo cual es otra de las razones porque la IA no es compatible con el carácter moral.

## **Si voy a pedir un consejo necesito a alguien capacitado, con autoridad moral ¿Qué es la autoridad moral? ¿De dónde viene?**

Ya que la ética es algo personal, lo que cada uno aporta se traduce en un consejo único y personalizado. Un buen consejo debería ser bueno por que quien lo da, tiene sabiduría que ha acumulado en base a la experiencia con sus propios dilemas. No podemos enseñarle a la gente a ser ética, adquirir sabiduría requiere de un viaje, un proceso de reflexión, evolución y desarrollo, no viene de clases lectivas sobre la teoría de la ética, y no existe algo tal como "experticia ética", ya que el conocimiento teórico y de libros, no te hace ético.

## **¿Cómo describe el rol de la fisicalidad en el dilema con la IA?**

La relación entre el cuerpo y la mente es en sí una cuestión filosófica. ¿es lo mismo desconectar un robot con IA que desconectar a un humano con soporte vital?. El ser un individuo es también tener un cuerpo y una cara que encarnen tu identidad, lo que no pasa con la IA. Quienes somos, es tanto cuerpo como mente, el cuerpo expresa la personalidad y la encarna, expresando a la vez la ética personal, y al momento de pedir ayuda con dilemas éticos, la corporalidad importa.

## **La máquina nunca puede tener carácter moral**

Sin un cuerpo o una vida, no podemos establecer que la IA tenga compasión o sabiduría, no pueden entender dilemas éticos, ya que estos requieren una historia y una corporalidad. El rol de la persona en la ética significa que el proyecto de dar a una máquina rienda suelta en decisiones éticas está perdido y condenado desde el principio, ya que la ética personal nunca estará ahí, no pueden aprenderla de otros y nunca adquirirán la sabiduría de las experiencias para obtenerla por sí mismos.

## **Imaginemos que se crea una máquina que no fuese preprogramada con códigos éticos, sino que realmente crezca, aprenda y entienda, algo que verdaderamente pase a tener carácter moral o ético**

Mucha gente me hace este escenario parecido al de la película Blade Runner en que ya no se puede distinguir entre humano e IA, pero no me parece apropiado porque perdemos contacto con la idea de que son máquinas. Si creció, tuvo una niñez, aprendió y es mortal, ¿es realmente una máquina? En este escenario perdemos de vista los conceptos básicos y empezamos a cambiar de discusión porque ya no estaríamos hablando de los mismos sujetos.

## **¿Es una preocupación que se estén creando máquinas preprogramadas con códigos morales basados en un mal entendimiento de la ética?**

Lo apropiado es siempre evaluar al humano y sus decisiones humanas, ya que todo lo que haga una máquina es, al fin y al cabo, la instrucción de una persona y los robots se vuelven simples conductos de la voluntad de quienes los programan.