Somos IA n.º 3:

¿QUIÉN VIVE, QUIÉN MUERE, Y QUIÉN DECIDE?

Traducido por Daniel Domínguez Figaredo

Términos de uso

Citar como:

Julia Stoyanovich, Mona Sloane y Falaah Arif Khan. "¿Quién vive, quién muere, quién decide?" We are AI Cómics, Vol. 3 (2021) http://r-ai.co/comics

PREDECIR ES MUY DIFÍCIL, ESPECIALMENTE SI ES SOBRE EL FUTURO.

POR MÁS QUE SEA DIFÍCIL - DEBIDO A LO INCIERTO Y COMPLEJO QUE ES EL MUNDO - PREDECIR EL FUTURO ES HABITUALMENTE EL TRABAJO DE LA IA.

Y DEBIDO A LA DIFICULTAD DE LA TAREA QUE A VECES ES INCLUSO IMPOSIBLE, LOS SISTEMAS DE IA COMETEN ERRORES.

POR EJEMPLO, UNA IA DE LUZ INTELIGENTE PUEDE NO ACERTAR SOBRE SI UNA LUZ DEBE ESTAR ENCENDIDA O APAGADA.

ENCENDER LA LUZ EN MEDIO DE LA NOCHE NOS DESPERTARÁ.

Y DEJAR LAS LUCES ENCENDIDAS SISTEMÁTICAMENTE CONLLEVA PAGAR UNA FACTURA DE ENERGÍA MÁS ALTA.

OTRO EJEMPLO: UNA IA DE SERVICIO AL CLIENTE EN TU ZAPATERÍA FAVORITA PODRÍA MALINTERPRETAR SU PEDIDO,

...Y TÚ RECIBIRÁS UN PAR DE ZAPATOS EQUIVOCADOS.

ESTOS ERRORES PUEDEN SER IRRITANTES, PERO PLANTEAN UN ESCASO RIESGO.

LAS CONSECUENCIAS DE DICHOS ERRORES NO SON GRAVES Y SON REVERSIBLES.

SIN EMBARGO, HAY CASOS EN LOS QUE LOS ERRORES PUEDEN PROVOCAR DAÑOS IRREVERSIBLES Y CATASTRÓFICOS,

INCLUSO LA PÉRDIDA DE VIDAS HUMANAS.

CONSIDEREMOS UN AUTOMÓVIL AUTÓNOMO:

UNA IA ESTÁ A PUNTO DE CRUZAR UNA INTERSECCIÓN,

Y NO RECONOCE A UNA PERSONA EN UNA BICICLETA COMO UNO DE LOS TIPOS DE OBJETOS QUE ESPERARÍA VER EN LA CARRETERA.

EL VEHÍCULO NO SE DETIENE Y ATROPELLA A LA CICLISTA.

OTRO EJEMPLO ES CUANDO EL VEHÍCULO AUTÓNOMO NO DETECTA LA PRESENCIA DE UNA PERSONA EN SILLA DE RUEDAS CRUZANDO LA INTERSECCIÓN.

ESTO PODRÍA SUCEDER SI, POR EJEMPLO, LA PERSONA CRUZA LA INTERSECCIÓN YENDO HACIA ATRÁS,

DE MANERA QUE LA IA DEL COCHE AUTÓNOMO NO CALCULE BIEN LA TRAYECTORIA DEL PEATÓN.

PERO LOS CONDUCTORES HUMANOS TAMBIÉN CAUSAN ACCIDENTES.

ASÍ QUE, ¿POR QUÉ DEJAR QUE LO PERFECTO SEA EL ENEMIGO DE LO BUENO?

¿NO DEBERÍAMOS ESTAR PREPARADOS PARA SUFRIR ALGUNOS ERRORES COMETIDOS POR AUTOMÓVILES AUTÓNOMOS EN ARAS DE UNA MAYOR SEGURIDAD GENERAL DE NUESTRO SISTEMA DE TRANSPORTE Y LA CONVENIENCIA DE LOS CONDUCTORES?

DE HECHO, ¿NO PODRÍAMOS CODIFICAR NUESTRO CRITERIO SOBRE LOS ERRORES MÁS IMPORTANTES A EVITAR, Y DEJAR QUE UNA IA RESUELVA LOS CASOS DUDOSOS?

¿NO PODEMOS EQUIPAR NUESTRA IA CON ESOS VALORES?

UN EJEMPLO FAMOSO QUE NOS HACE PENSAR EN NUESTROS VALORES, Y LAS COMPENSACIONES QUE INTRODUCEN, ES

EL PROBLEMA DEL TRANVÍA.

ES UN EXPERIMENTO SOBRE EL PENSAMIENTO QUE PLANTEA UN DILEMA ÉTICO:

¿DEBERÍAMOS SACRIFICAR LA VIDA DE UNA SOLA PERSONA PARA SALVAR LA VIDA DE UN GRUPO GRANDE DE PERSONAS?

CURIOSAMENTE, LOS EXPERIMENTOS EN ÉTICA Y PSICOLOGÍA HAN DEMOSTRADO QUE NO HAY UNA RESPUESTA CLARA.

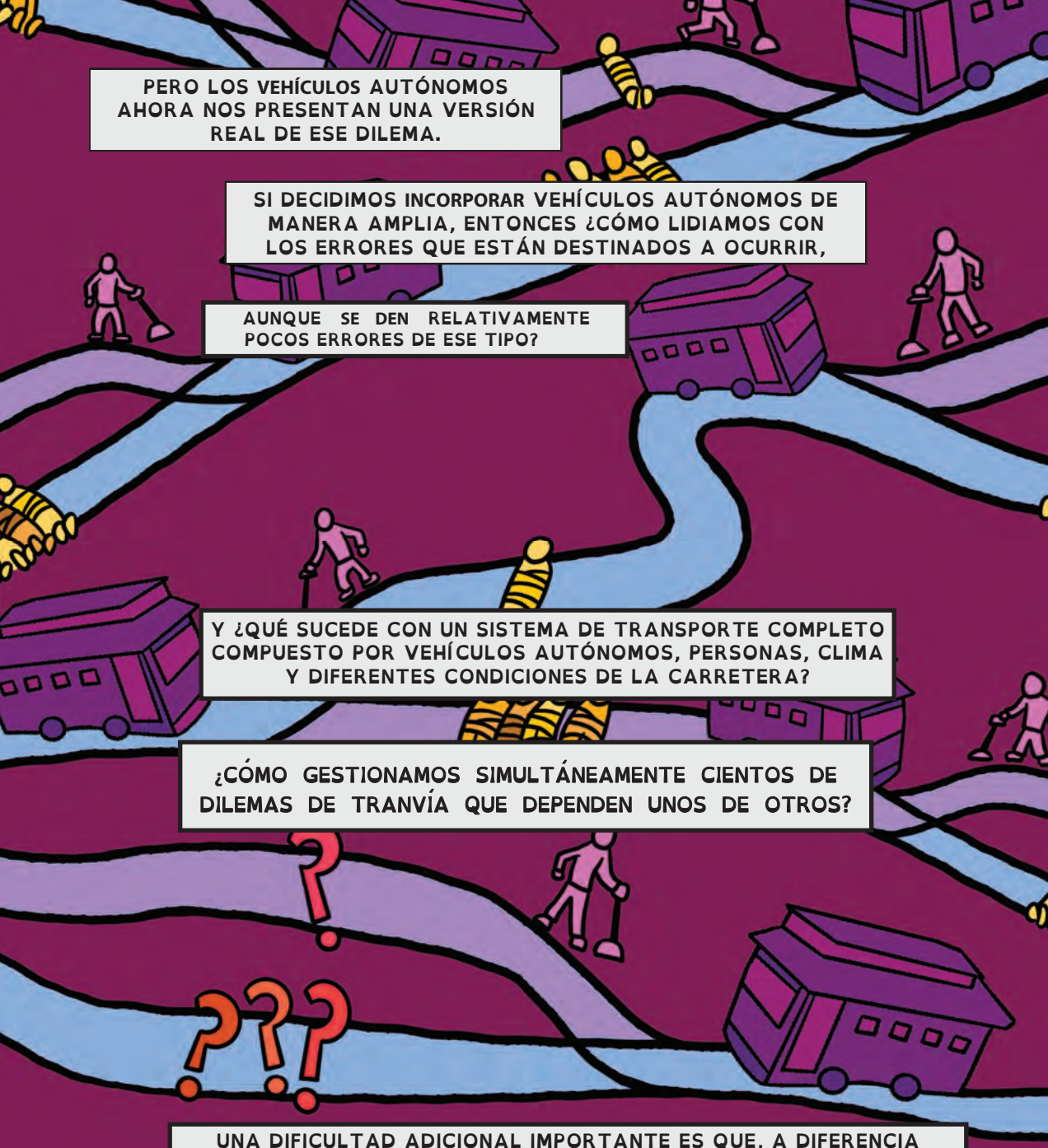
LO QUE DECIDIMOS DEPENDE DE NUESTROS VALORES: DE LO QUE CONSIDERAMOS CORRECTO O INCORRECTO,

DE LOS DIVERSOS ELEMENTOS DE NUESTRA IDENTIDAD, DE NUESTRO CONTEXTO CULTURAL,

Y TAMBIÉN DE LA CONFIGURACIÓN ESPECÍFICA DEL PROBLEMA (EL CONTEXTO EN EL QUE SE TOMA LA DECISIÓN).

POR INTERESANTE QUE SUENE, EL DILEMA DEL TRANVÍA SIGUE SIENDO UN EXPERIMENTO SOBRE EL PENSAMIENTO

Y HA SIDO CRITICADO POR SER DEMASIADO EXTRAVAGANTE Y POCO REALISTA.


PERO LOS VEHÍCULOS AUTÓNOMOS AHORA NOS PRESENTAN UNA VERSIÓN REAL DE ESE DILEMA.
SI DECIDIMOS INCORPORAR VEHÍCULOS AUTÓNOMOS DE MANERA AMPLIA, ENTONCES ¿CÓMO LIDIAMOS CON LOS ERRORES QUE ESTÁN DESTINADOS A OCURRIR,
AUNQUE SE DEN RELATIVAMENTE POCOS ERRORES DE ESE TIPO?
Y ¿QUÉ SUCEDE CON UN SISTEMA DE TRANSPORTE COMPLETO COMPUESTO POR VEHÍCULOS AUTÓNOMOS, PERSONAS, CLIMA Y DIFERENTES CONDICIONES DE LA CARRETERA?
¿CÓMO GESTIONAMOS SIMULTÁNEAMENTE CIENTOS DE DILEMAS DE TRANVÍA QUE DEPENDEN UNOS DE OTROS?
???
UNA DIFICULTAD ADICIONAL IMPORTANTE ES QUE, A DIFERENCIA DEL CLÁSICO DILEMA DEL TRANVÍA, DONDE SE SABE CUÁNTAS PERSONAS HAY EN CADA LADO DE LA VÍA,
UN AUTOMÓVIL AUTÓNOMO, Y OTROS TIPOS DE TECNOLOGÍA, OPERAN BAJO UN ALTO GRADO DE INCERTIDUMBRE.
PUEDE QUE NI SIQUIERA SEPAMOS SI HAY PERSONAS EN LAS VÍAS,
Y MUCHO MENOS EL NÚMERO CONCRETO QUE HAY Y A QUÉ GRUPOS PUEDEN REPRESENTAR.
¿CÓMO HACEMOS JUICIOS DE VALOR FRENTE A SEMEJANTE INCERTIDUMBRE?

EL DILEMA DEL TRANVÍA ILUSTRA UNA TEORÍA ESPECÍFICA DENTRO DE LA FILOSOFÍA MORAL:

¿PUEDA ESTA TEORIA OFRECERNOS ALGUNAS PAUTAS?

EL UTILITARISMO

EL UTILITARISMO SE REFIERE A UN PRINCIPIO MORAL QUE SOSTIENE QUE LA MEJOR ACCIÓN, EN CUALQUIER SITUACIÓN,

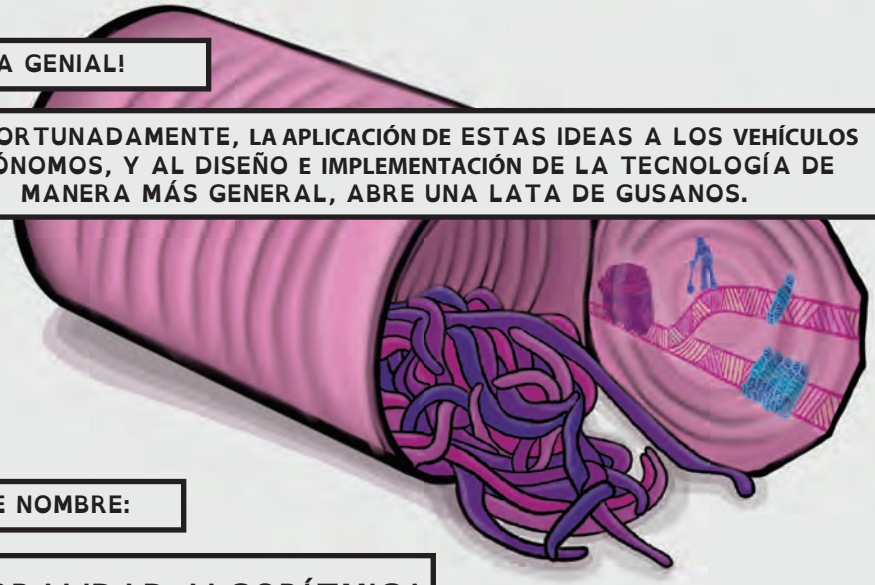
ES LA QUE PRODUCE EL MAYOR EQUILIBRIO ENTRE BENEFICIOS Y DAÑOS PARA TODAS LAS PERSONAS IMPLICADAS.

EL UTILITARISMO PROVIENE DE LOS FILÓSOFOS Y ECONOMISTAS INGLESES DE FINALES DEL SIGLO XVIII Y PRINCIPIOS DEL XIX, JEREMY BENTHAM Y JOHN STUART MILL.

UNA CITA FAMOSA DE BENTHAM ES: "LA MAYOR FELICIDAD DEL MAYOR NÚMERO ES LA MEDIDA DEL BIEN Y DEL MAL".

¡SUENA GENIAL!

DESAFORTUNADAMENTE, LA APLICACIÓN DE ESTAS IDEAS A LOS VEHÍCULOS AUTÓNOMOS, Y AL DISEÑO E IMPLEMENTACIÓN DE LA TECNOLOGÍA DE MANERA MÁS GENERAL, ABRE UNA LATA DE GUSANOS.



Y TIENE NOMBRE:

MORALIDAD ALGORÍTMICA

LA MORALIDAD ALGORÍTMICA ES EL ACTO DE ATRIBUIR JUICIOS MORALES A LOS ALGORITMOS.

Y HACERLO ES PROBLEMÁTICO. ESTA ES LA RAZÓN:

PARA EMPEZAR, ¿CÓMO MEDIMOS LA FELICIDAD Y LA INFELICIDAD?

¿Y CÓMO CODIFICAMOS ESAS MEDIDAS EN UN CONJUNTO DE OBJETIVOS QUE SEAN COMPRENSIBLES PARA UN ALGORITMO?

RARA VEZ EXISTE UNA FÓRMULA MATEMÁTICA O UNA DECLARACIÓN LÓGICA QUE PUEDA ESTABLECER EL EQUILIBRIO ENTRE LOS BENEFICIOS Y LOS DAÑOS.

EN OTRAS PALABRAS: SIMPLEMENTE NO HAY UNA FÓRMULA PARA LO "CORRECTO" O LO "INCORRECTO".

Y NO EXISTE UNA FÓRMULA PARA DEFINIR LOS VALORES, Y CÓMO SURGEN Y CAMBIAN ESOS VALORES EN SITUACIONES SOCIALES COMPLEJAS.

OTRA RAZÓN POR LA CUAL LA MORALIDAD ALGORÍTMICA ES PROBLEMÁTICA ES QUE,

CUANDO SE COMETE UN ERROR DE JUICIO SOBRE LO QUE ESTÁ BIEN O MAL

—Y, COMO YA SABEMOS, SE COMETERÁN ERRORES PORQUE EL MUNDO ES COMPLEJO, INCIERTO Y, QUIZÁS, INCLUSO IMPREDECIBLE—,

LA MORALIDAD ALGORÍTMICA REQUERIRÍA DE UN ALGORITMO PARA ASUMIR LA RESPONSABILIDAD POR EL ERROR.

PERO RESPONSABILIZAR A UN ALGORITMO POR UN ERROR NO TIENE SENTIDO:

UN ALGORITMO NO POSEE CONCIENCIA NI LIBRE ALBEDRÍO,

NO TOMA UNA DECISIÓN INTENCIONAL QUE CONDUCE A UN ERROR,

Y POR LO TANTO NO PUEDE SER CONSIDERADO RESPONSABLE.

¿DÓNDE NOS DEJA ESTO?

EL ABRELATAS QUE ES EL PROBLEMA DEL AUTOMOVIL NOS MOSTRÓ QUE NO PODEMOS DELEGAR LA ÉTICA EN LAS MÁQUINAS.

QUE TODAVÍA DEPENDE DE NOSOTROS, LOS HUMANOS, TOMAR DECISIONES Y MEDIDAS (O ELEGIR NO ACTUAR),

DE ACUERDO CON NUESTROS VALORES, Y CON LAS LEYES VIGENTES.

Y LUEGO DEPENDE DE NOSOTROS ASUMIR LA RESPONSABILIDAD DE LAS CONSECUENCIAS DE CUALQUIER ERROR.

NO PODEMOS EXTERNALIZAR HACIA UNA MÁQUINA EL TRABAJO DE SER HUMANOS.

EN RESUMEN, PARA INCORPORAR LA ÉTICA EN SISTEMAS SOCIOTÉCNICOS COMO LA IA,

DEBEMOS PENSAR QUÉ VALORES ESTÁN ALREDEDOR DE ESOS SISTEMAS,

QUIÉN SE BENEFICIA CUANDO LOS SISTEMAS FUNCIONAN BIEN,

Y QUIÉN SE RESULTA PERJUDICADO POR SUS ERRORES.

Y DEBEMOS ASUMIR COLECTIVAMENTE LA RESPONSABILIDAD DE DECIDIR SOBRE EL EQUILIBRIO ENTRE LOS BENEFICIOS Y LOS DAÑOS,

PARA QUE "LA MAYOR FELICIDAD" QUE JEREMY BENTHAM PROMETE AL MAYOR NÚMERO DE PERSONAS TAMBIÉN SEA DISFRUTADA POR LA MAYOR DIVERSIDAD DE PARTES INTERESADAS.

ESTE TRABAJO DE COMPRENSIÓN Y NEGOCIACIÓN COLECTIVA DE LAS COMPENSACIONES ES LO QUE HACE QUE EL DISEÑO DE LAS TECNOLOGÍAS SE SUSTENTE EN LAS PERSONAS.