

Somos IA #3:

¿QUIÉN VIVE, QUIÉN MUERE, QUIÉN DECIDE?



TERMINOS DE USO

Todos los paneles de este libro de historietas tienen licencia CC BY-NC-ND 4.0. Consulte la página de licencias para obtener detalles sobre cómo puede usar este material gráfico.

TL;DR: Se puede usar paneles/grupos de paneles en presentaciones/artículos, siempre y cuando

1. Se proporcione la cita adecuada
2. No se realicen modificaciones a los paneles individuales

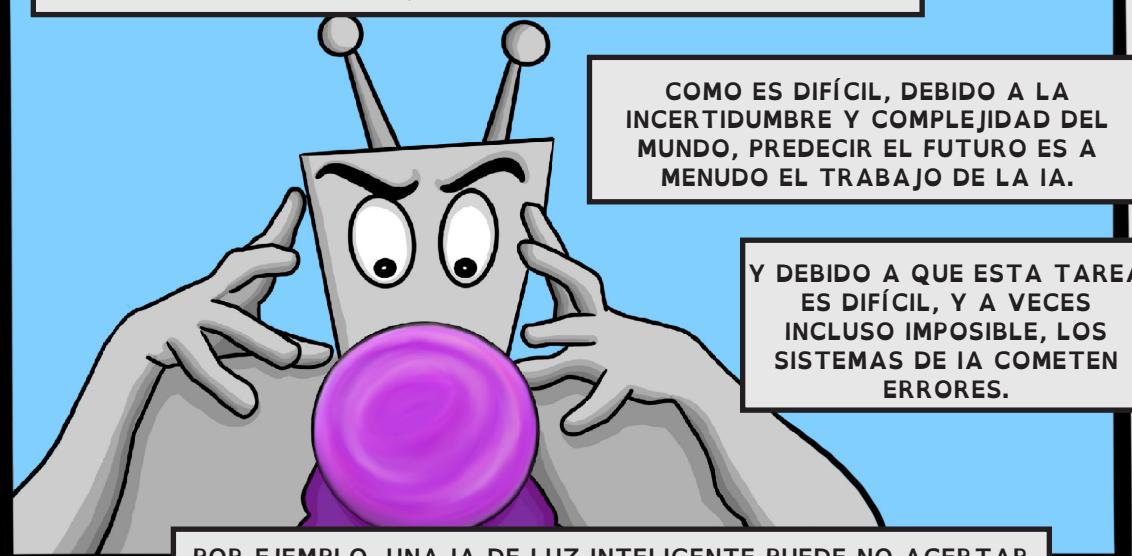
Citar como:

Julia Stoyanovich, Mona Sloane y Falaah Arif Khan. ¿Quién vive, quién muere, quién decide? Somos IA, Vol. 3 (2022) https://dataresponsibly.github.io/we-are-ai/comics/vol3_es.pdf

Contacto:

Dirija cualquier consulta sobre el uso de elementos de este cómic a themachinelearnist@gmail.com, con copia a stoyanovich@nyu.edu

LA PREDICCIÓN ES DIFÍCIL, ESPECIALMENTE SI ES DEL FUTURO.



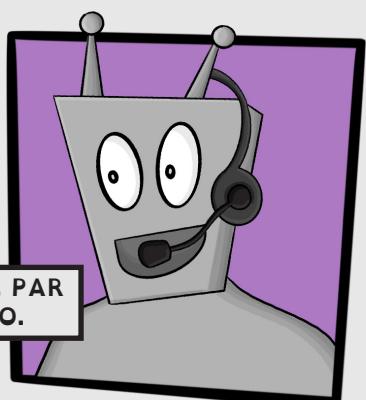
COMO ES DIFÍCIL, DEBIDO A LA INCERTIDUMBRE Y COMPLEJIDAD DEL MUNDO, PREDECIR EL FUTURO ES A MENUDO EL TRABAJO DE LA IA.

Y DEBIDO A QUE ESTA TAREA ES DIFÍCIL, Y A VECES 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.



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



ESTOS ERRORES PUEDEN SER IRRITANTES, PERO PRESENTAN POCO 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 CATASTRÓFICOS,

INCLUSO PÉRDIDA DE VIDAS HUMANAS.

CONSIDEREMOS UN AUTOMÓVIL AUTÓNOMO:

UNA IA QUE 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 AUTO NO SE DETIENE Y ATROPELLA A LA CICLISTA.



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

ESTO PODRÍA SUceder si, por ejemplo, la persona cruzara la intersección yendo hacia atrás,

POR LO QUE LA IA DEL AUTO AUTÓNOMO NO CALCULARÍA BIEN LA TRAYECTORIA DEL PEATÓN.



ENTONCES, ¿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 PODEMOS CODIFICAR NUESTRO CRITERIO SOBRE QUÉ ERRORES SON MÁS IMPORTANTES DE EVITAR Y DEJAR QUE UNA IA RESUELVA LAS COMPENSACIONES?

¿NO PODEMOS EQUIPAR NUESTRA IA CON VALORES?

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

EL PROBLEMA DEL TRANVÍA.

ES UN EXPERIMENTO DE
PENSAMIENTO QUE PLANTEA UN
DILEMA ÉTICO:

¿DEBERÍAMOS SACRIFICAR LA VIDA DE
UNA 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 PROBLEMA
DEL TRANVÍA SIGUE SIENDO UN EXPERIMENTO
DE PENSAMIENTO

Y HA SIDO CRITICADO POR SER TAN
EXTRAVAGANTE COMO POCO REALISTA.

PERO LOS AUTOS AUTÓNOMOS AHORA
NOS PRESENTAN UNA VERSIÓN REAL DE
ESTE DILEMA.

SI DECIDIMOS DESPLEGAR VEHÍCULOS AUTÓNOMOS DE
MANERA AMPLIA, ENTONCES ¿CÓMO LIDIAMOS CON LOS
ERRORES QUE ESTÁN DESTINADOS A OCURRIR,

AUNQUE HAYA RELATIVAMENTE
POCOS ERRORES DE ESTE TIPO?

Y ¿QUÉ SUCEDA CON UN SISTEMA DE TRANSPORTE COMPLETO
COMPUESTO POR VEHÍCULOS AUTÓNOMOS, PERSONAS, CLIMA
Y DIFERENTES CONDICIONES DE LA CARRETERA?

Y ¿QUÉ SUCEDA CON UN SISTEMA DE TRANSPORTE COMPLETO
COMPUESTO POR VEHÍCULOS AUTÓNOMOS, PERSONAS, CLIMA
Y DIFERENTES CONDICIONES DE LA CARRETERA?

UNA DIFICULTAD ADICIONAL IMPORTANTE ES QUE, A DIFERENCIA DEL
PROBLEMA CLÁSICO DEL TRANVÍA, DONDE SE SABE CUÁNTAS
PERSONAS HAY EN CADA LADO DE LA PISTA,

UN AUTOMÓVIL AUTÓNOMO, Y OTROS TIPOS DE TECNOLOGÍA, OPERAN
BAJO UN ALTO GRADO DE INCERTIDUMBRE.

PUEDE QUE NI SIQUIERA SE SEPA SI HAY
PERSONAS EN LAS VÍAS,

Y MUCHO MENOS CUÁNTAS HAY Y QUÉ
GRUPOS PUEDEN REPRESENTAR.

¿CÓMO HACEMOS JUICIOS DE VALOR FRENTE A LA INCERTIDUMBRE?

EL PROBLEMA DEL TRANVÍA ILUSTRA UNA DOCTRINA ESPECÍFICA DE FILOSOFÍA MORAL:

¿PODRÍA ESTA DOCTRINA BRINDARNOS ALGUNAS PAUTAS?

UTILITARISMO

EL UTILITARISMO ES UN PRINCIPIO MORAL QUE SOSTIENE QUE EL CURSO DE ACCIÓN CORRECTO, EN CUALQUIER SITUACIÓN,



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

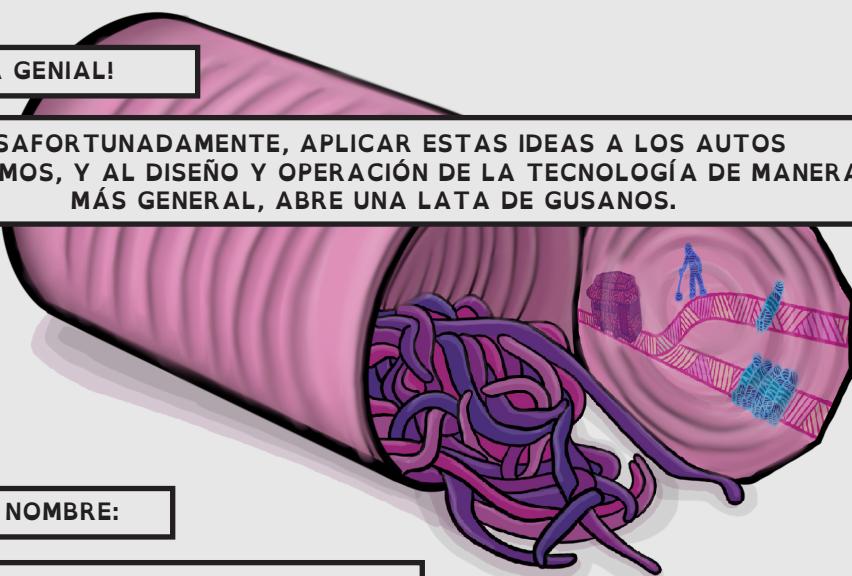


EL UTILITARISMO PROVIENE DE LOS FILÓSOFOS Y ECONOMISTAS INGLESES DE FINALES DEL SIGLO XVIII Y PRINCIPIOS DEL XIX, JEREMY BENTHAM Y JOHN STUART MILL.

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

¡SUENA GENIAL!

DESAFORTUNADAMENTE, APLICAR ESTAS IDEAS A LOS AUTOS AUTÓNOMOS, Y AL DISEÑO Y OPERACIÓN 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 EL RAZONAMIENTO MORAL A LOS ALGORITMOS.

Y HACERLO ES PROBLEMÁTICO. ESTA ES LA RAZÓN:



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



¿Y CÓMO CODIFICAMOS ESTAS MEDIDAS EN UN CONJUNTO DE OBJETIVOS QUE COMPRENDERÁ UN ALGORITMO?

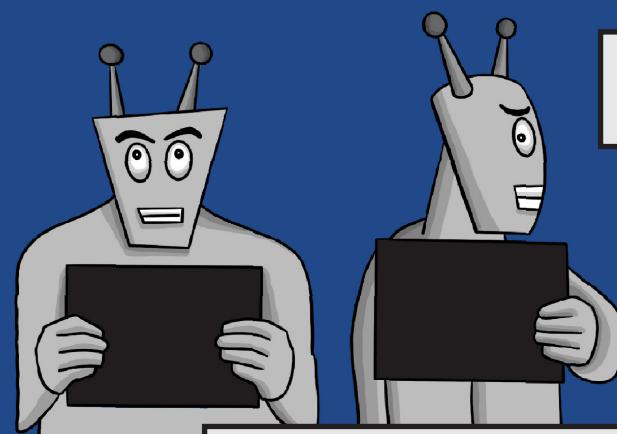
RARA VEZ EXISTE UNA FÓRMULA MATEMÁTICA O UNA DECLARACIÓN LÓGICA QUE PUEDA CAPTURAR EL EQUILIBRIO ENTRE LOS BENEFICIOS Y LOS DAÑOS.

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



Y NO EXISTE UNA FÓRMULA PARA LOS VALORES, Y PARA CÓMO SURGEN Y CAMBIAN LOS 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 CARRO NOS MOSTRÓ QUE NO
PODEMOS DELEGAR LA ÉTICA A 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 EL TRABAJO
DE SER HUMANO A UNA MÁQUINA.

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

DEBEMOS PENSAR EN QUÉ VALORES SE CUECEN
EN ESTOS SISTEMAS,

QUIÉN SE BENEFICIA CUANDO LOS
SISTEMAS FUNCIONAN BIEN,

Y QUIÉN SE PERJUDICA
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 ARRAIGA EL DISEÑO DE LA TECNOLOGÍA EN LAS PERSONAS.