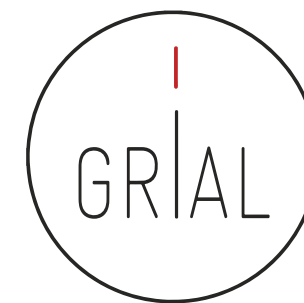




VNiVERSiDAD
D SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL

Inteligencia Artificial

Una perspectiva desde la ficción a la realidad

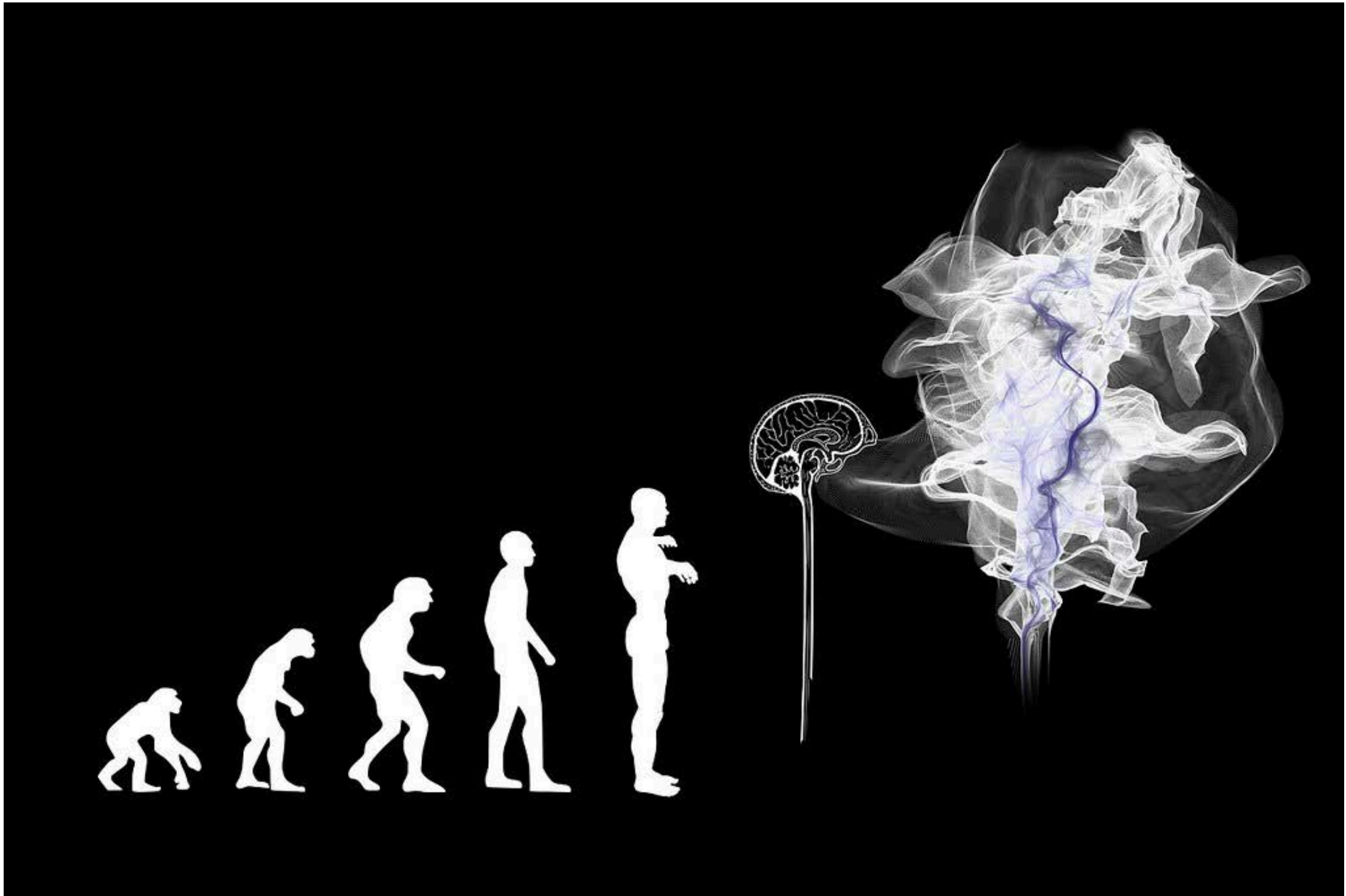
Dr. D. Francisco José García Peñalvo

GRupo de investigación en InterAcción y eLearning (GRIAL)
Instituto de Ciencias de la Educación
Departamento de Informática y Automática
Universidad de Salamanca, España

fgarcia@usal.es

<http://grial.usal.es>

<http://twitter.com/frangp>

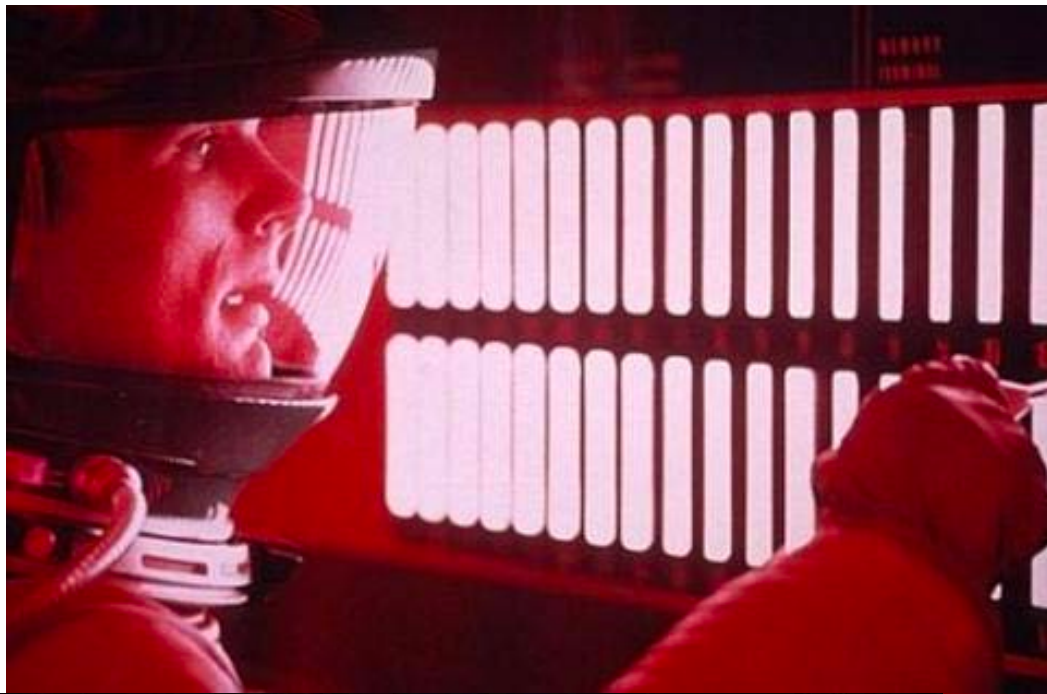


"3885331" by Catalyst Research Institute is licensed under CC PDM 1.0

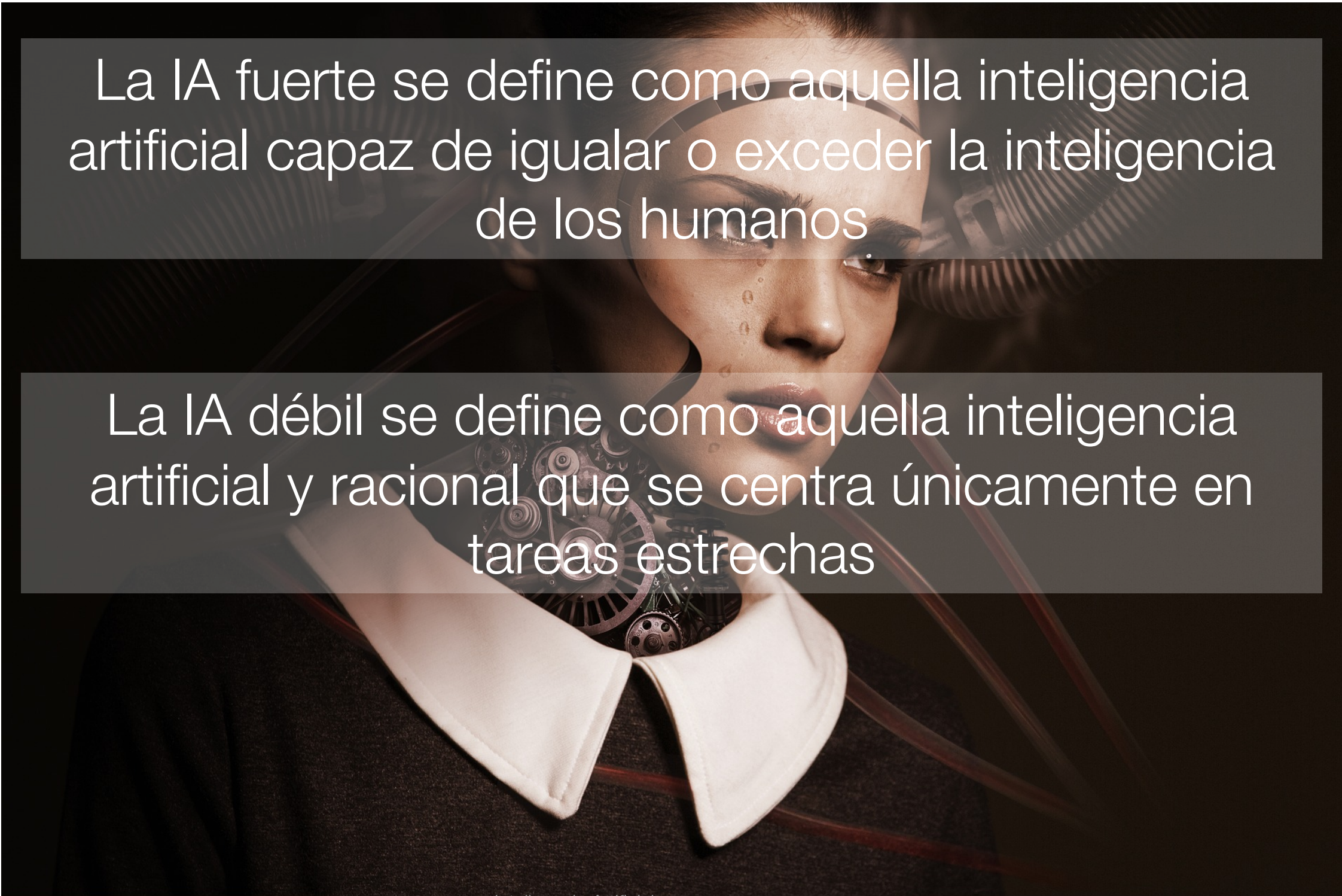
Inteligencia artificial



La relación hombre-máquina inteligente ha sido una constante en el imaginario de la ciencia-ficción



Inteligencia artificial fuerte vs. débil

The background image is a close-up of a woman's face, looking slightly to the right. She has a serious expression. Overlaid on her face and neck are various mechanical and technological elements, including gears, pipes, and a head-mounted display (HMD) with a transparent visor. The lighting is dramatic, with strong highlights and shadows.

La IA fuerte se define como aquella inteligencia artificial capaz de igualar o exceder la inteligencia de los humanos

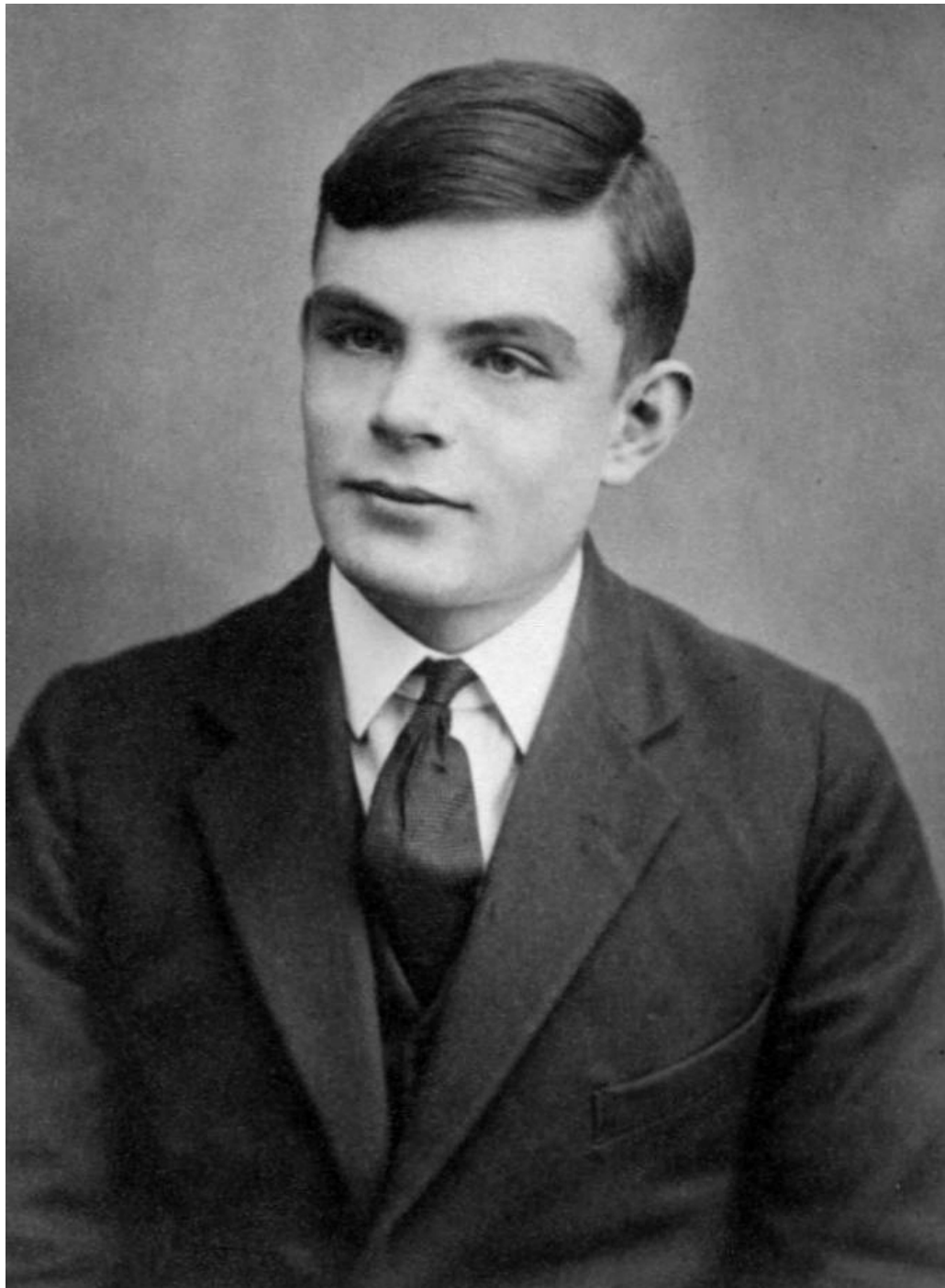
La IA débil se define como aquella inteligencia artificial y racional que se centra únicamente en tareas estrechas

<https://goo.gl/qWRmch>

La idea de origen es que las máquinas pensaran como los humanos



Los orígenes de la inteligencia artificial



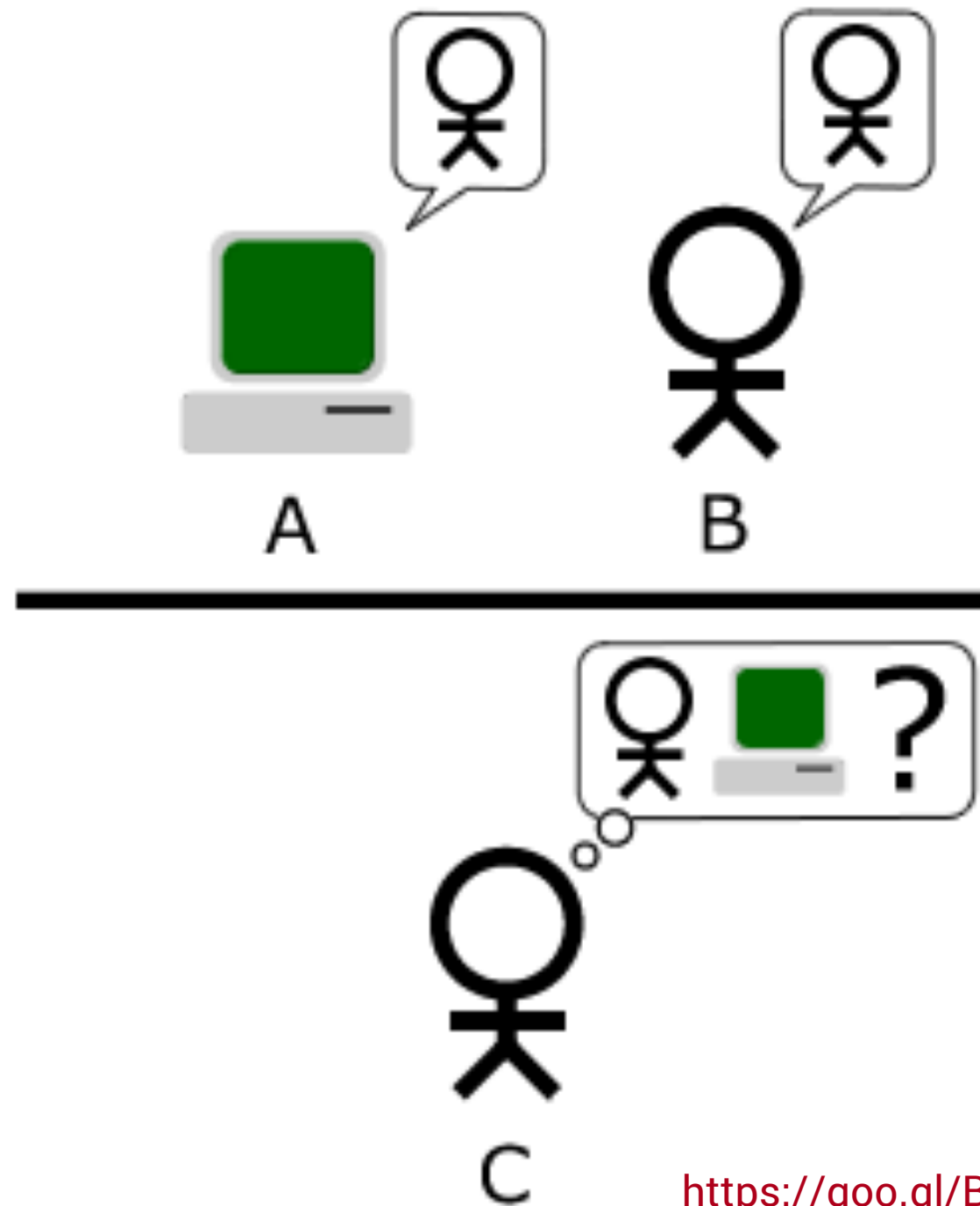
- Hay un número de trabajos iniciales que se pueden caracterizar como de inteligencia artificial
- Pero se le atribuye a Alan Turing (1950) el dar una visión de la inteligencia artificial en su artículo *Computing machinery and intelligence*
 - Introduce la prueba de Turing, el aprendizaje automático, los algoritmos genéricos y el aprendizaje por refuerzo
- En 1956, dos años después de la muerte de Turing, John McCarthy acuña el término Inteligencia Artificial como un campo independiente en la Ciencia de la Computación

<https://goo.gl/BMv2RF>

Test de Turing

- Es una prueba de la habilidad de una máquina para exhibir un comportamiento inteligente similar al de un ser humano o indistinguible de este
- Propone que un humano evaluara conversaciones en lenguaje natural entre un humano y una máquina diseñada para generar respuestas similares a las de un humano
- El evaluador sabría que uno de los participantes de la conversación es una máquina y los intervinientes serían separados unos de otros
- La conversación estaría limitada a un medio únicamente textual como un teclado de computadora y un monitor por lo que sería irrelevante la capacidad de la máquina de transformar texto en habla
- En el caso de que el evaluador no pueda distinguir entre el humano y la máquina acertadamente (Turing originalmente sugirió que la máquina debía convencer a un evaluador, después de 5 minutos de conversación, el 70% del tiempo), la máquina habría pasado la prueba
- Esta prueba no evalúa el conocimiento de la máquina en cuanto a su capacidad de responder preguntas correctamente, solo se toma en cuenta la capacidad de esta de generar respuestas similares a las que daría un humano

Test de Turing



<https://goo.gl/B8JJ45>

Test de Turing

- Para que un ordenador pase el test de Turing debe
 - Procesar lenguaje natural
 - Representar el conocimiento
 - Razonar automáticamente
 - Auto aprender
- La llamada Prueba Global de Turing incluye una señal de vídeo que permite al evaluador valorar la capacidad de percepción del evaluado, y también le da la oportunidad al evaluador de pasar objetos físicos «a través de una ventanita», para ello se requiere
 - Visión computacional
 - Robótica

Test de Turing

- En 2014, Eugene, un *bot* de *chat* pasó por primera vez el test de Turing
- Eugene Goostman es un niño ucraniano de 13 años, con una pasión por los dulces y las hamburguesas. Lentes redondos que evocan a las famosas gafas de Harry Potter y un padre que se dedica a la ginecología
- En realidad Eugene es un *bot* de *chat*, diseñado por 3 programadores ucranianos con el único propósito de pasar la prueba de Turing
- El 7 de junio de 2014, en un concurso con motivo del 60 aniversario de la muerte de Turing, el 33 % de los jueces del evento pensó que Goostman era humano; el organizador del evento Kevin Warwick consideró que «pasó» la prueba de Turing, en consecuencia, por la predicción de Turing que para el año 2000, las máquinas serían capaces de engañar a un 30 % de los jueces humanos después de cinco minutos de interrogatorio
- Eugene superó los criterios de Turing, pero usando técnicas que no implican la existencia de inteligencia
- Lo que Eugene sí hizo bien fue jugar al “juego de imitación” del que Turing hablaba en su artículo original

Test de Turing



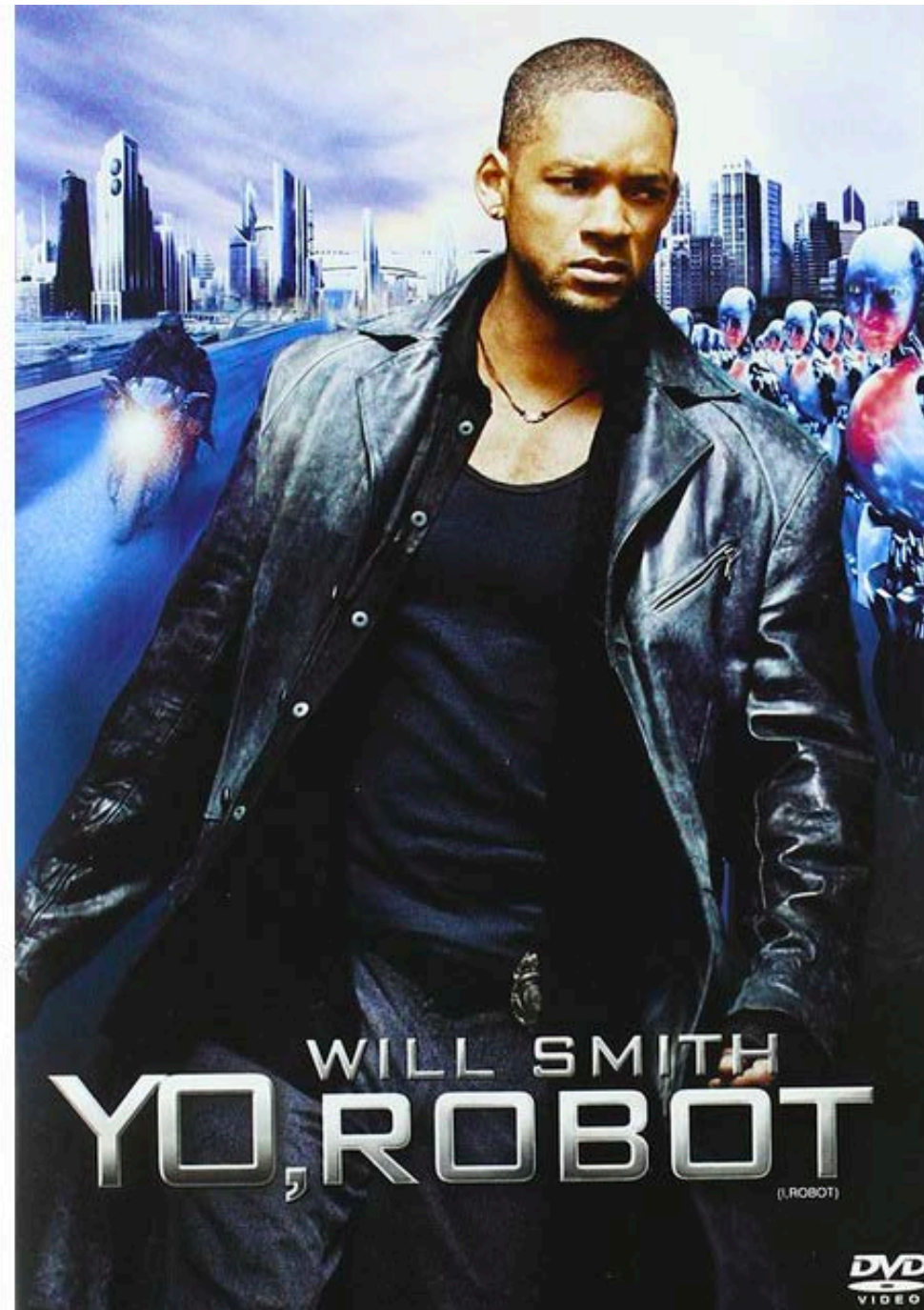
Un **robot** es una entidad virtual o mecánica artificial



La palabra robot se conoce por la obra R.U.R. (Robots Universales Rossum) del checo Karel Čapek (1920). La palabra original era *robotnik*

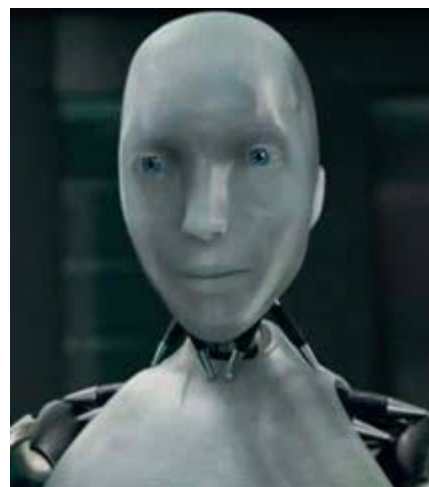
La robótica, usada para describir este campo de estudio, fue acuñada por el escritor de ciencia ficción Isaac Asimov

La relación entre robótica e inteligencia artificial



Inteligencia Artificial
Una perspectiva desde la ficción a la realidad

Las leyes de la robótica de Isaac Asimov

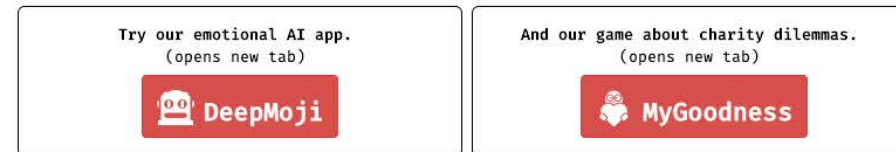
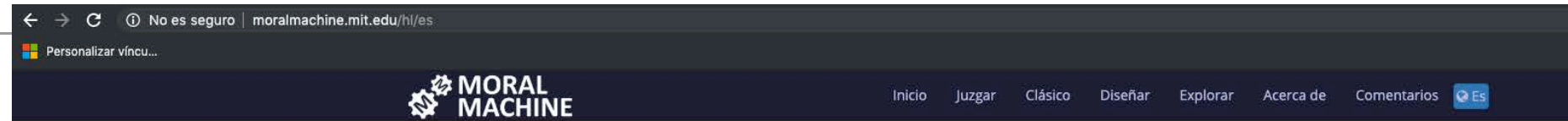


1. Un robot no hará daño a un ser humano o, por inacción, permitirá que un ser humano sufra daño
2. Un robot debe cumplir las órdenes dadas por los seres humanos, a excepción de aquellas que entrasen en conflicto con la primera ley
3. Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la primera o con la segunda ley

(Asimov, 1950)

Coches autónomos

<http://moralmachine.mit.edu/hl/es>



¡Bienvenido a la Máquina Moral! Una plataforma para recopilar una perspectiva humana sobre las decisiones morales tomadas por las máquinas inteligentes, como los coches autónomos.

Te mostramos dilemas morales, donde un coche sin conductor debe elegir el menor de dos males, como elegir entre matar a dos pasajeros o cinco peatones. Como observador externo, tú puedes **juzgar** cuál es el resultado que consideras más aceptable. Puedes ver cómo sus respuestas se comparan con las de otras personas.

Si te sientes creativo, también puedes **diseñar** tus propios escenarios, para que tú y otros usuarios podáis **explorar**, compartir y discutir.



Un vehículo autónomo de Uber atropella y mata a una mujer



El primer accidente de la industria automovilística de coches autónomos ocurre con un coche en pruebas de Uber, en Arizona, E.E.U.U. El coche llevaba un pasajero de seguridad. (Natalie Behring / Reuters)

La ley 0 de la robótica

**Un robot no hará daño a la Humanidad o, por inacción,
permitir que la Humanidad sufra daño**

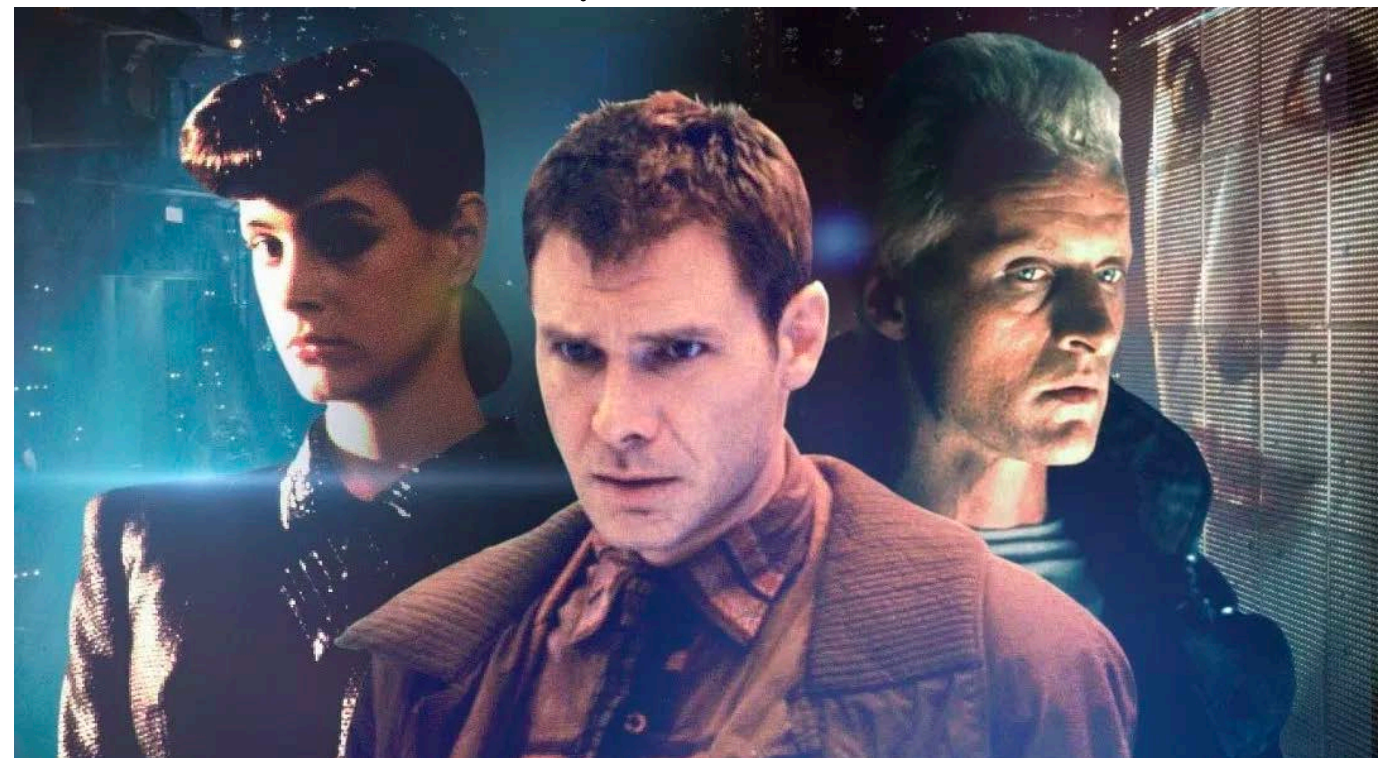
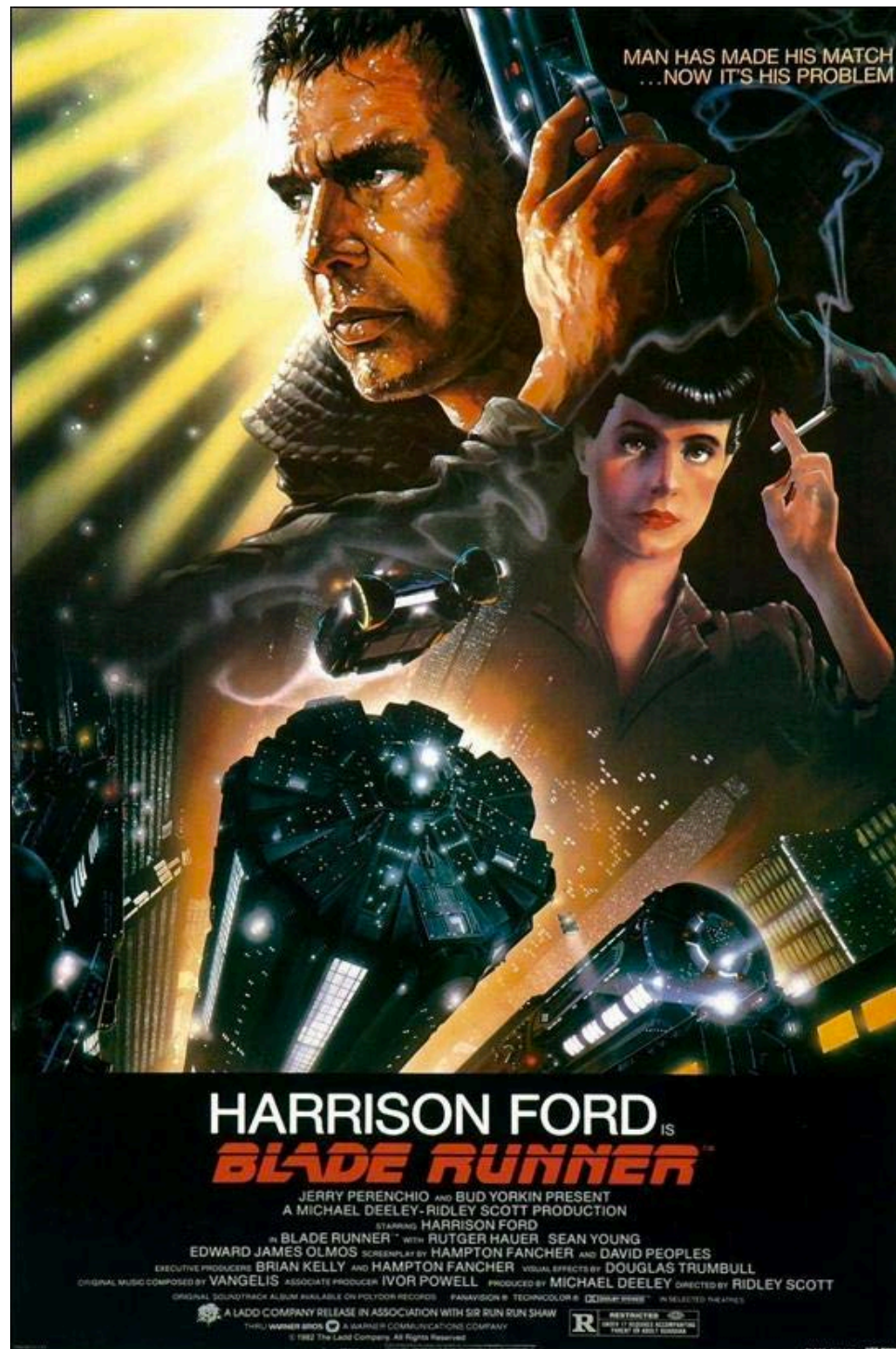
No se puede cambiar el orden de la leyes

WHY ASIMOV PUT THE THREE LAWS OF ROBOTICS IN THE ORDER HE DID:

POSSIBLE ORDERING	CONSEQUENCES	
1. (1) DON'T HARM HUMANS 2. (2) OBEY ORDERS 3. (3) PROTECT YOURSELF	[SEE ASIMOV'S STORIES]	BALANCED WORLD
1. (1) DON'T HARM HUMANS 2. (3) PROTECT YOURSELF 3. (2) OBEY ORDERS	EXPLORE MARS!  Haha, no. It's cold and I'd die.	FRUSTRATING WORLD
1. (2) OBEY ORDERS 2. (1) DON'T HARM HUMANS 3. (3) PROTECT YOURSELF		KILLBOT HELLSCAPE
1. (2) OBEY ORDERS 2. (3) PROTECT YOURSELF 3. (1) DON'T HARM HUMANS		KILLBOT HELLSCAPE
1. (3) PROTECT YOURSELF 2. (1) DON'T HARM HUMANS 3. (2) OBEY ORDERS	 I'll make cars for you, but try to unplug me and I'll vaporize you.	TERRIFYING STANDOFF
1. (3) PROTECT YOURSELF 2. (2) OBEY ORDERS 3. (1) DON'T HARM HUMANS		KILLBOT HELLSCAPE

Bioingeniería y biomateriales

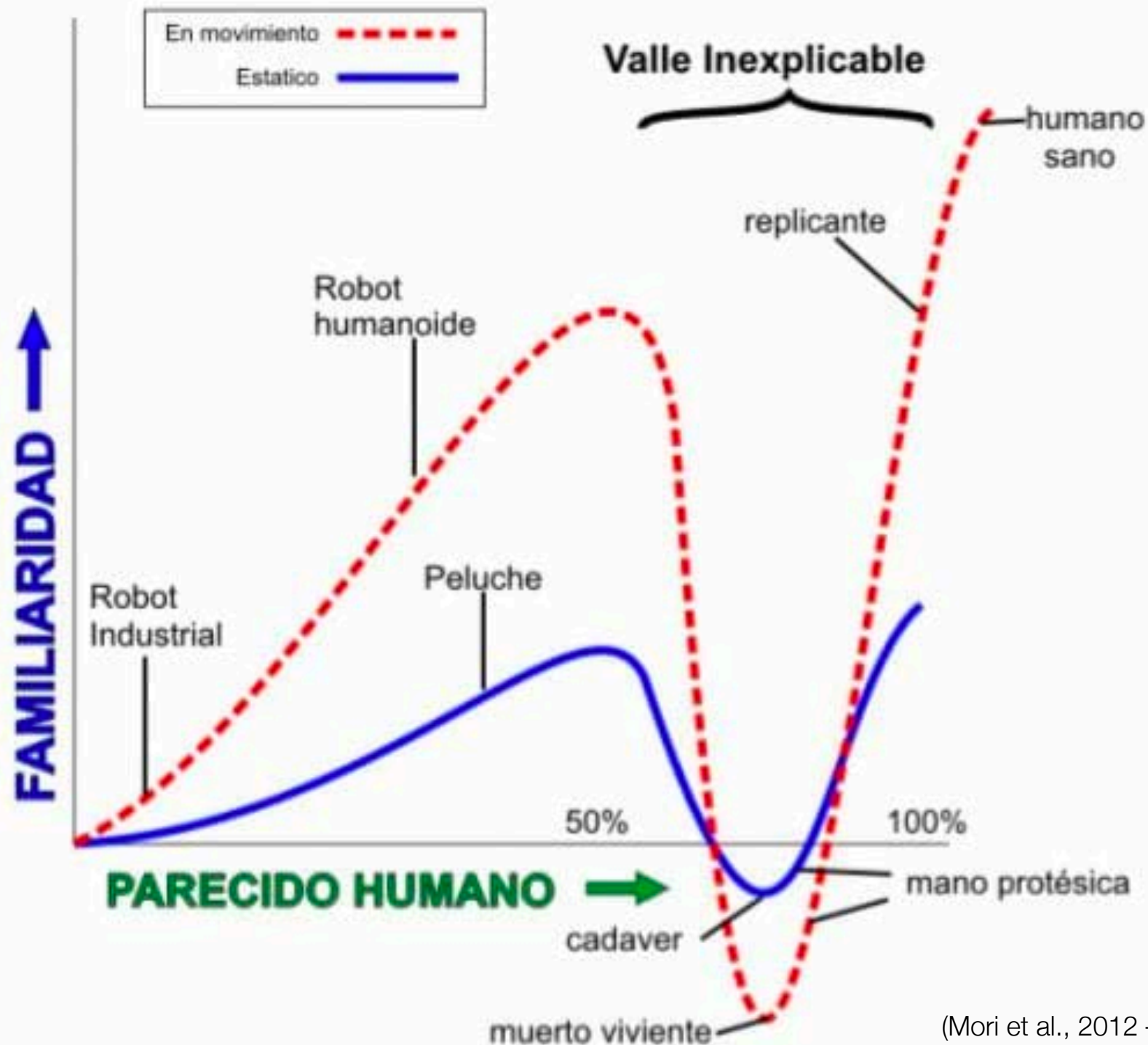
En un sentido amplio, los biomateriales serían materiales diseñados para actuar con sistemas biológicos con el fin de evaluar, tratar, aumentar o reemplazar algún tejido, órgano o función del cuerpo



Teoría del valle inquietante

- Es la sensación que produce ver un robot demasiado humano, pero no lo suficiente
- A medida que las figuras antropomórficas se acercan al ser humano real, la respuesta de los humanos es más empática
- Pero llega un punto a partir del cual se convierte en rechazo
- Si la figura se hace más parecida, hasta hacerse casi indistinguible de un ser humano verdadero la respuesta vuelve a ser positiva
- Este *gap* se denomina el valle inquietante o el valle inexplicable (*uncanny valley*)

Teoría del valle inquietante



(Mori et al., 2012 – traducción del artículo original en japonés de 1970)

Teoría del valle inquietante

- En Blade Runner, la teoría del valle inquietante se lleva al extremo por el miedo que supone no distinguir a un replicante de un humano (en Blade Runner 2049 llegan a poder reproducirse)
- De ahí que se aplique el test de Voight-Kampff, como si fuera un test de Turing orientado a la empatía



Hiroshi Ishiguro “En pocos años no podremos distinguir entre robots y humanos”



(Ishiguro, 2006)

Robots no antropomórficos y multifuncionales



Robots no antropomórficos y multifuncionales



Atlas



Spot Mini



Pick



Handle

<https://www.bostondynamics.com/>

Chatbots

- Un *bot* es un *software* de inteligencia artificial diseñado para realizar una serie de tareas por su cuenta y sin la ayuda del ser humano (reservas en un restaurante, marcar una fecha en el calendario, etc.)
- Un *chatbot* es un programa conversacional con inteligencia artificial
- No solo se utilizan como herramientas de atención al cliente, sino que se han convertido en una forma de automatizar y mejorar las operaciones internas

La inmortalidad



Altered carbón, basada en la novela del mismo nombre de Richard Morgan (2002)

La inmortalidad



Viñeta de 'El Incal' (L'Incal) Moebius y Alejandro Jodorowsky (1981)

Robert McIntyre y Michael McCanna fundaron en California la empresa Nectome en 2015 con el objetivo “de hacerle un *backup* a tu cerebro”

La inmortalidad

- Avances en la conservación de las sinapsis de un conejo (2016)
- Avances en la conservación de las sinapsis de un cerdo (2018)
- Usan simultáneamente dos técnicas: el embalsamiento y la conservación criogénica
- Para que funcione y se conserven todas las sinapsis debe hacerse con cerebros frescos, o dicho de otra manera, vivos
- Se desconoce si la conciencia se recuperaría con la memoria (los recuerdos son diferentes a la conciencia, las personas pueden perder la conciencia y mantener sus recuerdos a largo plazo)
- El tema despierta tantas preguntas científicas como filosóficas por las implicaciones éticas

Neuralink

- Empresa norteamericana de neurotecnología fundada por Elon Musk
- Desarrolla interfaces cerebro-computadora implantables
- En abril de 2017 que Musk dijo que el objetivo de empresa era hacer dispositivos para tratar enfermedades del cerebro a corto-plazo, con el objetivo eventual de perfeccionamiento humano
- Fusión de cerebro humano e inteligencia artificial

El miedo a la inteligencia artificial



Inteligencia Artificial
Una perspectiva desde la ficción a la realidad



No tendremos una segunda oportunidad con la inteligencia artificial (Nick Bostrom)



<https://goo.gl/Rsx7Sh>

Inteligencia Artificial
Una perspectiva desde la ficción a la realidad

"El éxito en la creación de la inteligencia artificial podrá ser el evento más grande en la historia de la humanidad. Desafortunadamente también sería el último, a menos de que aprendamos cómo evitar los riesgos" Stephen Hawking



Los sistemas no tienen capacidad para entender qué es la inteligencia artificial o para conocer su propio código, menos para modificarlo



En lugar de preocuparse por los peligros de una superinteligencia informática maquiavélica, hay que centrarse en otros problemas reales que presenta esta tecnología: la privacidad de los datos y los sesgos algorítmicos

No veremos ordenadores autoconscientes porque no es lo que estamos construyendo

Ramez Naam

2001: A Space Odyssey

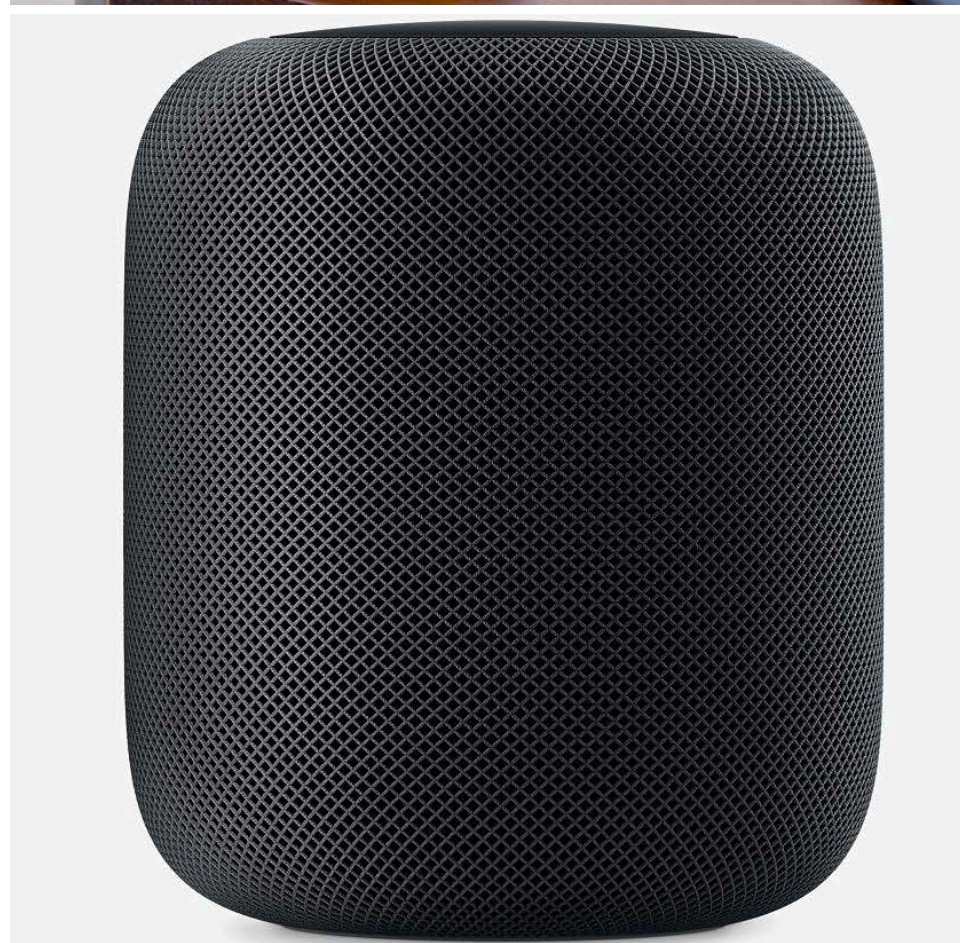
- En esta película (1968) se incluye muchas de las características mencionadas anteriormente: control de sistemas, miedo a la inteligencia artificial, aspectos morales, etc.
- Comunicación con una interfaz de voz y de visión artificial
- El error de HAL9000 se produce por una orden que va en contra de las tres leyes de Asimov: la misión está por encima de los hombres



Asistentes virtuales



Google assistant



Homepod (Apple)

- En 2008 Google consigue que una máquina reconozca la voz humana
- Los asistentes de voz y los altavoces inteligentes son *gadgets* habituales
- Representan una nueva forma de interacción persona-ordenador
- Alexa ha superado los 100 millones de unidades (no se incluyen en móviles)
 - Se estima que más de 7.500 millones de dispositivos en 2021 tendrán asistentes virtuales integrados
- Google Assistant entiende 30 idiomas, frente a los 20 de Siri o los 8 de Alexa
- El 71% de los usuarios de altavoces inteligentes usa el dispositivo al menos una vez al día
- Una de cada cinco búsquedas se realiza mediante voz

Visión artificial



- La visión artificial incluye métodos para adquirir, procesar, analizar y comprender las imágenes del mundo real con el fin de producir información numérica o simbólica para que puedan ser tratados por un ordenador
- Reconocimiento de sentimientos (muchas API: Emotient, Affectiva, EmoVu, Nviso, Kairos, Proyecto Oxford de Microsoft, Face Reader de Noldus, Sightcorp, SkyBiometry, Face++, IBM Watson, Cloud Vision de Google, etc.)
- Leer los labios (Chung et al., 2017)

- El internet de las cosas se refiere a una interconexión digital de objetos cotidianos con internet (Atzori et al., 2010; Gubbi et al., 2013)
- IoT depende de una serie integral de tecnologías: API que conectan los dispositivos a Internet, Big Data, analíticas predictivas, IA y *machine learning*, la nube y la identificación por radiofrecuencia (RFID)



El problema del black box



- En inteligencia artificial es posible observar los datos de entrada y de salida en los sistemas algorítmicos, pero su funcionamiento interno no se comprende muy bien debido a la enorme complejidad de cómputo que requieren en el aprendizaje automático, especialmente en el *deep learning*
- Estos sistemas no se basan en reglas establecidas de antemano
- Hay una falta de una descripción inteligible de los resultados obtenidos en cada caso y en el hecho de que no se hayan puesto de relieve las características más importantes del caso en cuestión que han dado lugar a esos resultados

Conclusiones

- La inteligencia artificial está haciendo realidad muchas de las predicciones de la ciencia ficción
- Las posibilidades son enormes en el beneficio de la humanidad
- No se está cerca de que las máquinas tengan conciencia, se encamina a hacer tareas concretas de una manera muy eficiente
- Los problemas éticos, de *black box*, de privacidad, etc., son inminentes frente al efecto *skynet* de destrucción o sometimiento de la humanidad
- A la inteligencia artificial hay que comprenderla, más que temerla, pero ha de tratarse con respeto y los avances desarrollarlos sobre un planteamiento ético y transparente
- Se deben evitar los sesgos de los algoritmos (que no discriminen por color, raza, género, etc.)



<http://pixabay.com/es/libro-libros-encuadernador-de-pila-190034/>

Referencias

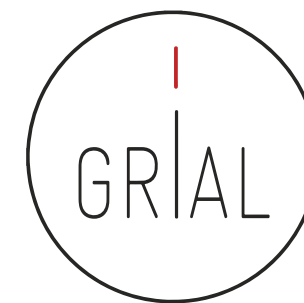


Referencias

- Asimov, I. (1950). *I, Robot*. USA: Gnome Press
- Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, 54(15), 2787–2805. doi:10.1016/j.comnet.2010.05.010
- Chung, J. S., Senior, A., Vinyals, O., & Zisserman, A. (2017). Lip Reading Sentences in the Wild *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (21-26 July 2017, Honolulu, HI, USA)*(pp. 3444–3453). USA: IEEE
- Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645–1660. doi:10.1016/j.future.2013.01.010
- Ishiguro, H. (2006). Android Science. *Scientific American*, 294(5), 32–34. doi:10.1038/scientificamerican0506-32
- Morgan, R. (2002). *Altered carbon*. UK: Victor Gollancz Ltd.
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley. *IEEE Robotics & Automation Magazine*, 19(2), 98–100. doi:10.1109/MRA.2012.2192811
- Mori, M. (1970). The uncanny valley. *Energy*, 7(4), 33–35
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460

Cómo citar este documento

García-Peñalvo, F. J. (2019). *Inteligencia Artificial. Una perspectiva desde la ficción a la realidad*. Salamanca, España: Grupo GRIAL. Disponible: <https://bit.ly/2Q0jap0>.
doi:10.5281/zenodo.2818903



VNiVERSiDAD
D SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL

Inteligencia Artificial

Una perspectiva desde la ficción a la realidad

Dr. D. Francisco José García Peñalvo

GRupo de investigación en InterAcción y eLearning (GRIAL)
Instituto de Ciencias de la Educación
Departamento de Informática y Automática
Universidad de Salamanca, España

fgarcia@usal.es

<http://grial.usal.es>

<http://twitter.com/frangp>