

Mesa 1: Bioética, gobernanza y mejora

La inteligencia artificial como tecnología de mejora humana y el debate sobre sus aplicaciones de doble uso

Miguel Moreno Muñoz

Universidad de Granada

mm3@ugr.es

Contenido

- **Desarrollo de la Inteligencia Artificial (IA)**
 - Carácter especulativo de las primeras propuestas y parón en la innovación
 - Combinación de desarrollos parciales: computación, robótica, sensores, software
 - Desarrollos importantes en los programas de Inteligencia Artificial aplicada
- **El potencial de la IA actual**
 - Desarrollos espectaculares para tareas complejas → especialización
 - Integración de tecnologías con resultados que superan las capacidades humanas
 - Del volcado de cerebros en la nube (Nicholas Agar vs Kurzweil) a múltiples aplicaciones de doble uso: robótica industrial, transporte, traducción, lenguaje natural... → Implicaciones socio-laborales
- **Las armas robóticas autónomas como amenaza global**
 - Un proceso irreversible y continuamente incentivado
 - Debate filosófico acerca de la IA general como riesgo existencial (Bostrom)
 - Consideraciones sobre el trato que merecen los sistemas de IAG avanzada

Contextualización

- **Carácter especulativo de las aportaciones iniciales:**
 - Entusiasmo inicial ante aplicaciones futuristas de la robótica dotada de IA avanzada
 - Moderación de expectativas más optimistas de desarrollo acelerado y validación contra posibles riesgos
 - Un nuevo escenario de desarrollos y progresos parciales susceptibles de integración
- **Desarrollos acumulativos, singularidad remota**
 - Vehículos capaces de conducción autónoma (aeronaves, carreteras, tráfico marítimo)
 - Sistemas de control automático de procesos / cooperación hum.-máq. entorno industrial
 - La perspectiva de máquinas superinteligentes, dotadas de inteligencia artificial general
- **Del optimismo tecnológico inicial al análisis de riesgos existenciales**
 - Ray Kurzweil (nanotech, IA y robótica): *La era de las máquinas inteligentes*
 - Biomejora mediante integración de biomecánica, electrónica y robótica
 - Dificultades para un establecer un marco que regule el uso de armas robóticas autónomas
 - Nicholas Agar y dificultades para hacer aplicar el principio de precaución en el desarrollo y aplicación de tecnologías de doble uso

Importa anticipar escenarios

- Agar discrepara de quienes sugieren **posponer el debate** sobre opciones morales en escenarios tecnológicos aún distantes:
 - Contra **Onora O'Neill**, para quien **el estado primitivo actual de las tecnologías de mejora reduce la urgencia del debate moral** sobre sus aplicaciones. Deberíamos concentrarnos en **evaluar tecnologías que ya están aquí**.
- **La mejora humana ha dejado de ser un futuro lejano:**
 - **El PGD o la clonación reproductiva permiten elegir** algunas características.
 - El desarrollo científico ha proporcionado múltiples **sorpresas morales**.
 - En la década de **1930** no se consideraba verosímil que la **energía atómica** pudiera ser utilizada con fines destructivos, algo que estaba muy próximo.
 - En la década de **1980**, los biólogos Davor Solter y James McGrath realizaron una serie exhaustiva de experimentos con manipulación de núcleos de células de ratón diferenciadas en óvulos fertilizados y concluyeron que “la **clonación** de mamíferos, mediante una simple transferencia nuclear [de células somáticas], es **biológicamente imposible**”. (Solter, McGrath, *Science* 226, 1984: 1317–19).

Ray Kurzweil y los programas de IA fuerte

- **La inteligencia artificial (AI) y las tecnologías de la computación han acelerado el rendimientos humano:**
 - Los desarrollos en el diseño de sistemas computacionales avanzados han potenciado el **programa fuerte de AI**
 - Sus logros no podrán distinguirse de los de seres humanos en tareas complejas
- **Su impacto puede suponer una mejora exponencial (el promedio de rendimiento actual)**
 - La **Singularidad** –período con un ritmo de cambio tecnológico acelerado- transformará la vida humana irreversiblemente.
 - Bostrom (*Superintelligence*, 2014) riesgos de los programas de AI fuerte (IAG).
 - Resulta fácil emular funciones especializadas del cerebro humano
 - Aunque diseñar máquinas capaces de sustentar procesos conscientes complejos sea un **desafío tecnológico descomunal** todavía.

Kurzweil acertó en la idea básica y enfoque:

Simular la distribución de neuronas del neocórtex en una red neural

<http://www.technologyreview.com/featuredstory/513696/deep-learning/>

- **Incremento de la capacidad de computación y desarrollo de fórmulas matemáticas /algoritmos avanzados permiten modelar muchas capas de neuronas virtuales.**
 - **Avances importantes en recon./procesamiento del lenguaje natural:**
 - **June 2015:** a **Google deep-learning system** that had been shown 10 million images from YouTube videos proved almost twice as good as any previous image recognition effort at identifying objects such as cats.
 - **October 2015:** Microsoft chief research officer Rick Rashid, in China, with a demonstration of speech **software that transcribed his spoken words into English text with an error rate of 7 percent, translated them into Chinese-language text, and then simulated his own voice uttering them in Mandarin.**
 - **October 2015:** a team of three graduate students and two professors won a contest held by Merck to **identify molecules that could lead to new drugs.**
 - Deep learning to zero in on the molecules most likely to bind to their targets.

¿Se sobrevalora el potencial de la IA?

<http://feeds.weblogssl.com/~r/xataka2/~3/uN3ojZ6FI0k/los-academicos-nos-dicen-la-verdad-sobre-el-estado-del-hype-de-la-inteligencia-artificial>

- Senén Barro: hubo optimismo excesivo, pero **ahora parece justificado**, aunque en esta ocasión "hay un enfoque más pragmático".
- Juan Pavón: "muchas de las aplicaciones de IA están embebidas en servicios y sistemas que utilizamos comúnmente"
 - "La cuestión es **dónde está la frontera de lo que es IA**, ya que **muchos problemas**, una vez resueltos, dejan de ser considerados del ámbito de la IA y **pasan a la ingeniería informática más convencional**".
- José Manuel Molina: la IA se puede aplicar **cada vez más a entornos reales**, "pero **queda mucho para poder desarrollar una IA que sea capaz de entender la realidad mediante conceptos abstractos** y poder interpretar, razonar, tomar decisiones y aprender en múltiples entornos, **del mismo modo que lo hacen los seres humanos**".
- **La creación de una "máquina pensante" continúa siendo un reto:**
 - "La investigación en IA no ha cumplido sus objetivos" (→ M. Minsky), aunque ha aportado "soluciones inteligentes" muy útiles para problemas concretos
→ **Un punto de vista ya obsoleto!**

¿Se está exagerando el potencial de la IA?

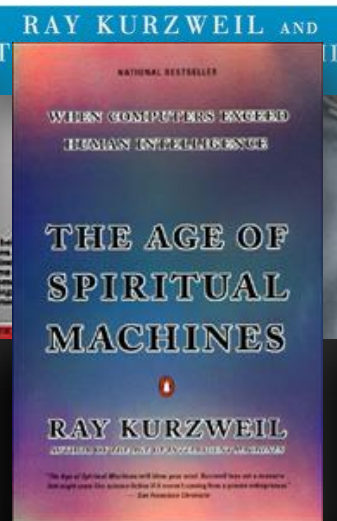
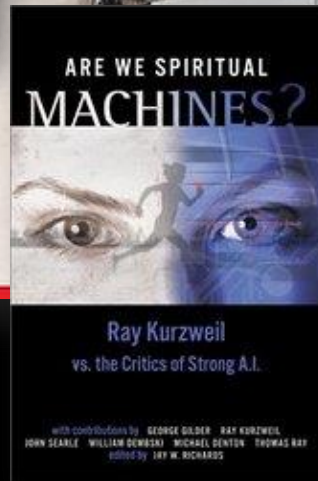
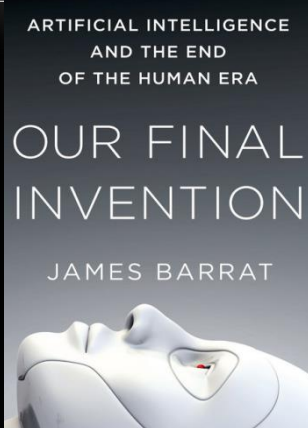
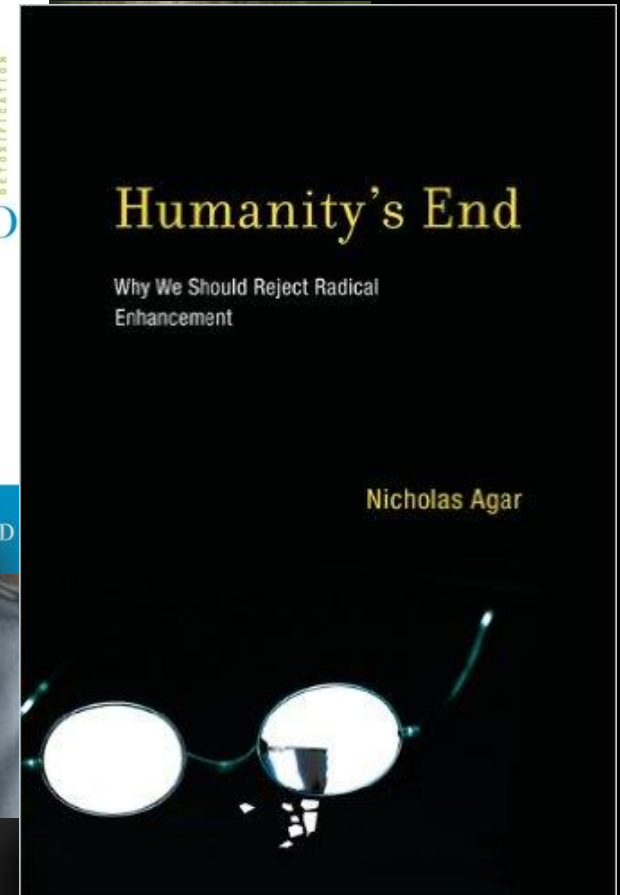
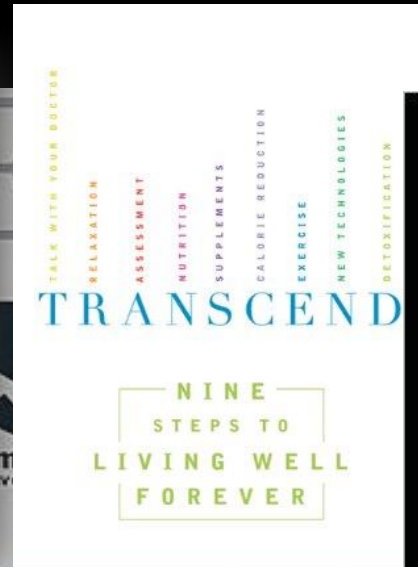
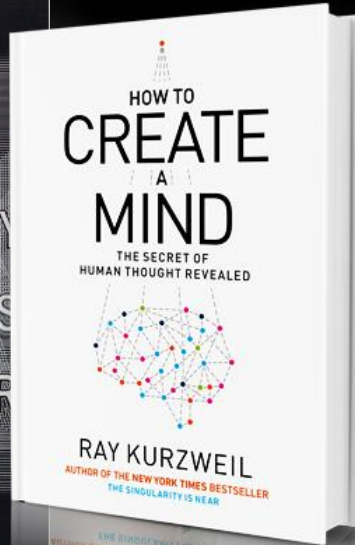
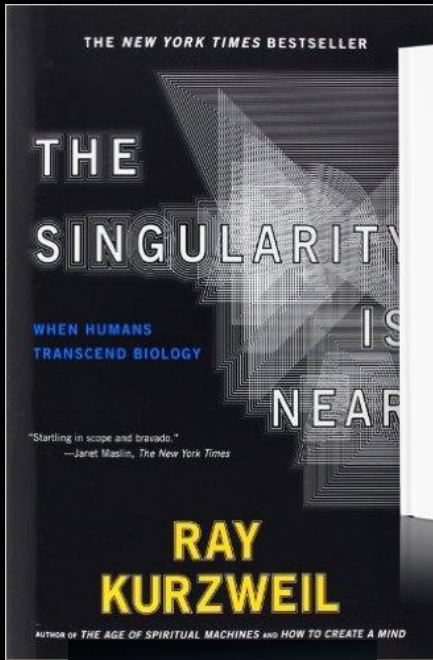
- **Ramón López de Mantaras: La IA será cada vez mayor en ámbitos como el de los coches autónomos, biología molecular y farmacología**
 - aplicaciones relacionadas con la web y las redes sociales
 - "optimistas sólo si estos progresos sirven para llevar a cabo aplicaciones socialmente responsables que mejoren la sociedad y nuestras vidas".
- **Enrique Carmona: las tecnologías en IA empiezan a tener valor económico**
 - [...] han cruzado la frontera que separa la investigación de laboratorio de la aplicación práctica en el mundo real, y atrae inversión cuantiosa de empresas.
- **Barro: el principal reto consiste en el aprendizaje máquina ("clave de bóveda de la IA").**
- **Molina: retos relacionados con la robustez de los modelos teóricos**
 - Algoritmos y sistemas de aprendizaje aplicables a la realidad ("sistemas capaces de tomar las decisiones 'mejor' que un humano")
 - El reto filosófico por excelencia de la IA: Construir máquinas como los replicantes de Blade Runner o robots como en ExMachina, indistinguibles de humanos.

¿Se está exagerando el potencial de la IA?

- **Ramón López de Mantaras: el mayor reto es "dotar de conocimientos de sentido común a las máquinas".**
 - La lección más importante en 60 años con líneas de trabajo en inteligencia artificial es que **lo que parecía más difícil** (diagnosticar enfermedades, jugar al ajedrez a nivel de gran maestro, etc.) **ha resultado ser relativamente fácil y lo que parecía más fácil ha resultado ser lo más difícil.**
- **Enrique Carmona: un reto inmediato es ayudar a hacer progresar todas las ramas de la IA**
 - Procesamiento del lenguaje natural, visión artificial, aprendizaje automático, razonamiento automático y robótica
 - Aportarían las mejoras necesarias para desarrollar el futuro "robot humanoide" avanzado, capaz de "mostrar emociones"
 - ¿Es **implementable la inteligencia social** y los recursos emocionales asociados **en un dispositivo robótico con IA?**

Ligero cambio en las perspectivas viables de mejora

La era de las máquinas inteligentes ha comenzado



Machine intelligence

https://en.wikipedia.org/wiki/Watson_%28computer%29

- **La revolución de las máquinas inteligentes**
 - Induce transformaciones a gran escala: comunicaciones, transporte, computación, medicina,
 - Herramienta para los servicios de inteligencia
 - Análisis estructurado sin los sesgos del razonamiento humano
- **Posibilidades exhibidas por el ordenador Watson, de IBM**
 - Ganó el concurso *Jeopardy!* (\$1 mill.).
 - Fue entrenado mediante técnicas de aprendizaje profundo.
 - Sirve como sistema experto para la toma de decisiones en medicina.

La era de las máquinas inteligentes ha comenzado:

How safe can artificial intelligence be?

<http://www.bbc.com/news/science-environment-34249500>

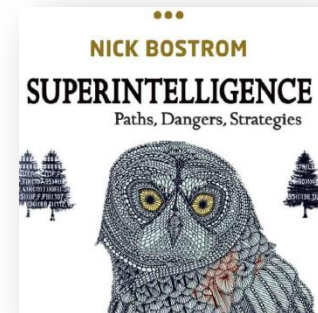
- **“General” or “Strong” AI**
 - Ability to do things that go beyond the original human intentions, not to “think” but to improvise.
- **Brett Kennedy**
 - “For the foreseeable future I am not concerned [...] nor do I expect to see a robot as intelligent as a human. I have first-hand knowledge of how hard it is for us to make a robot that does much of anything.
 - backed up by one of Britain's leading figures in AI, Prof Alan Winfield of the Bristol Robotics Lab.
- **Alan Winfield, a “voice of calm”**
 - Fears of future super intelligence - robots taking over the world - are greatly exaggerated”.
 - But innovations should be carefully handled - and **he was among 1,000 scientists and engineers who signed an appeal for a ban on AI in weaponry.**

La era de las máquinas inteligentes ha comenzado:

How safe can artificial intelligence be?

<http://www.bbc.com/news/science-environment-34249500>

- **Winfield:** Robots and intelligent systems must be engineered to very high standards of safety for exactly the same reasons that we need our washing machines, cars and airplanes to be safe.
 - Predicting the future pace of technology is impossible
 - ¿Every researcher in every part of the world will take a responsible approach?
 - ¿Defense industries would cooperate?
- **Nick Bostrom suggests a new frame: existential risks**
 - ***Superintelligence: Paths, dangers, strategies* (2014):**
 - The definitive texts laying out very clearly why we need to worry
 - **The human-machine parity (Artificial General Intelligence) as a threat**
 - Recent surveys of experts: **there's a 50% chance that computers could reach human-level intelligence as soon as 2050** - just 35 years away.
 - There's **a 90% chance of machine-human parity by 2075.**
 - "We really don't realise the power of this thing we are creating."



La era de las máquinas inteligentes ha comenzado

- ***Superintelligence is worth a closer look*** (Bostrom, p. 38)
 - It may be reasonable to believe that **human-level machine intelligence has a fairly sizeable chance of being developed by mid-century**,
 - that it might **perhaps fairly soon thereafter result in superintelligence**;
 - **a wide range of outcomes** may have a significant chance of occurring, **including extremely good outcomes and outcomes that are as bad as human extinction**.
- **Superintelligence like a tool, rather than an agent**
 - Ordinary software does not raise any safety concerns
 - Might one not create “tool-AI” that is like such software—like a flight control system, or a virtual assistant—only more flexible and capable?
 - Why build a superintelligence that has a will of its own?
 - **The agent paradigm is fundamentally misguided**. Instead of creating an AI that has beliefs and desires and that acts like an artificial person, we could build regular software that “simply does what it is programmed to do”
 - → ¿**Learning machines?**

Narrow AI - general AI

Ho, S.-B., & Liasvia, F. (2013). Knowledge Representation, Learning, and Problem Solving for General Intelligence. In K.-U. Kühnberger, S. Rudolph, & P. Wang (Eds.), *Artificial General Intelligence SE - 7* (Vol. 7999, pp. 60–69). Springer

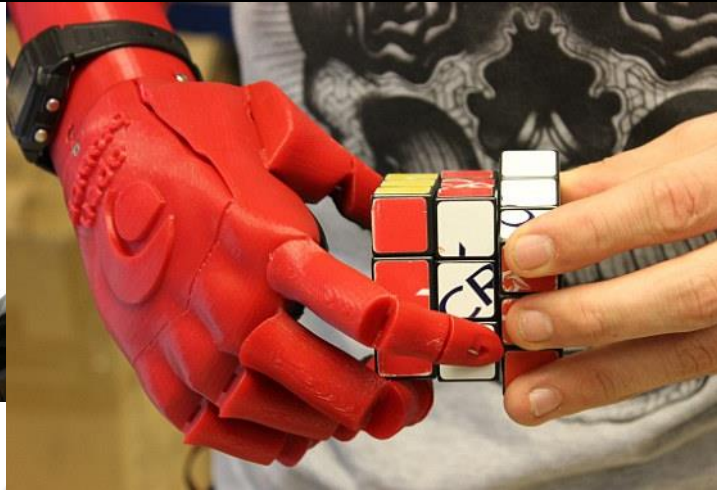
- **Bostrom (2014): nuevas demandas ligadas al uso de agentes dotados de Inteligencia Artificial General (AGI)**
 - “Even if an AI system is designed with a user override, one must consider the career incentive of a bureaucrat who will be personally blamed if the override goes wrong, and who would much prefer to blame the AI for any difficult decision with a negative outcome.”
- **Los investigadores en IA han desarrollado algoritmos que consiguen un rendimiento equivalente al humano (o superior) en dominios restringidos**
 - **Deep Blue** batió al campeón del mundo en ajedrez (aunque no sirva para conducir un automóvil o deducir leyes físicas).
 - Muchos algoritmos de la IA moderna emulan a la vida biológica en una amplia diversidad de aspectos (Homo sapiens es un desafío considerable).
 - Abejas coordinando tareas como construir colmenas, castores construyendo presas...
 - Tb emulan la capacidad humana de aprender de observando otras formas de vida

Casos de interés

Convergencia e integración de tecnologías avanzadas de IA en robótica industrial

Desarrollos limitados en prótesis avanzadas

La biónica no avanza al mismo ritmo que la IA



The cheap robotic hand 3D-printed (£1,000)

- Set to revolutionise prosthetics: 3D-printed device performs advanced tasks for a fraction of the cost.
- Robotic hands are becoming increasingly responsive and dexterous, but their cost is one of the reasons why they're not more widely available.



Nigel Ackland, with bebionic3 myoelectric hand (carbon fibre mechanical hand)

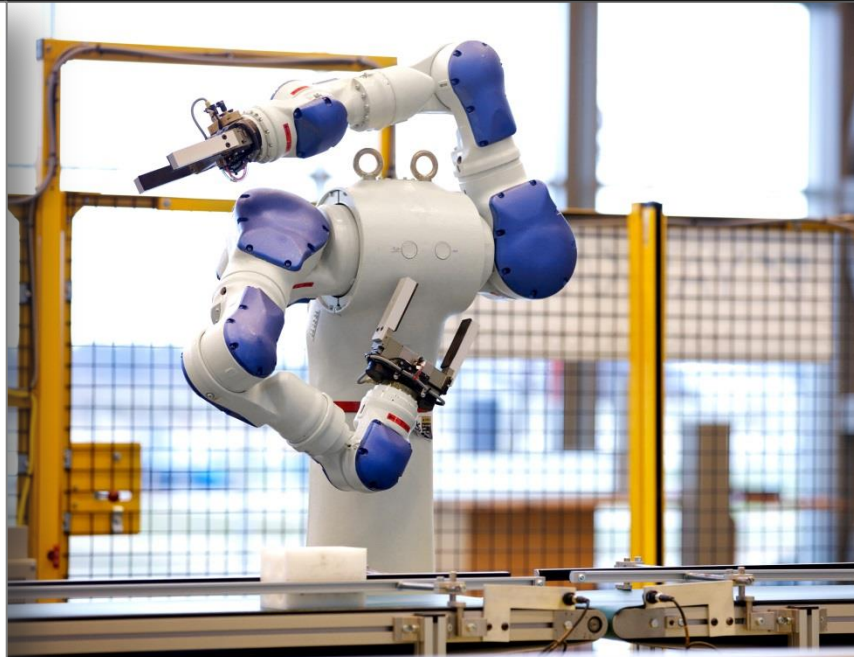
- Controlled with movements in his upper arm.
- Made from aluminium and alloy knuckles, moves like a real human limb by responding to Nigel's muscle twitches.
- The robotic arm is so sensitive that can type on a computer keyboard and peel vegetables.

Robots versátiles de uso industrial



Versatile Yaskawa
Motoman MH24 Robot

<http://www.motoman.com/media/pr/201503-MH24.php>



Midea invested \$160 million on automatic production from 2012 to 2014 (expanded to over \$800 million dollar in 2015-2020).

- 2,800 workers in 2005 (air cond. workshop)
 - 900 on Dec. 2015
 - 50 in the next three years.
- en.ofweek.com, August 14, 2015



Robot Model JX2050

- Arm, Rotating Supply Platform, Spaying and Control System and Oil Supply System.
- With Japan ASAHI-AGB40 automatic spraying guns.
- A Freedom of motion 6Shaft Drive style Holoaxial AC Servomotor +Absolute value coder Placement style Standing or hoisting

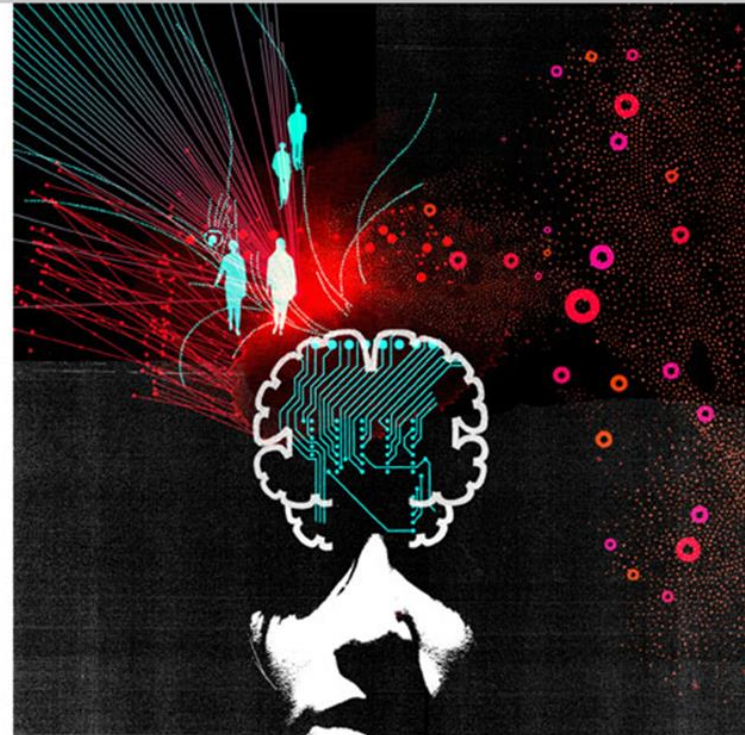
<http://en.ofweek.com>

10 BREAKTHROUGH TECHNOLOGIES 2013

[Introduction](#) [The 10 Technologies](#) [Past Years](#)

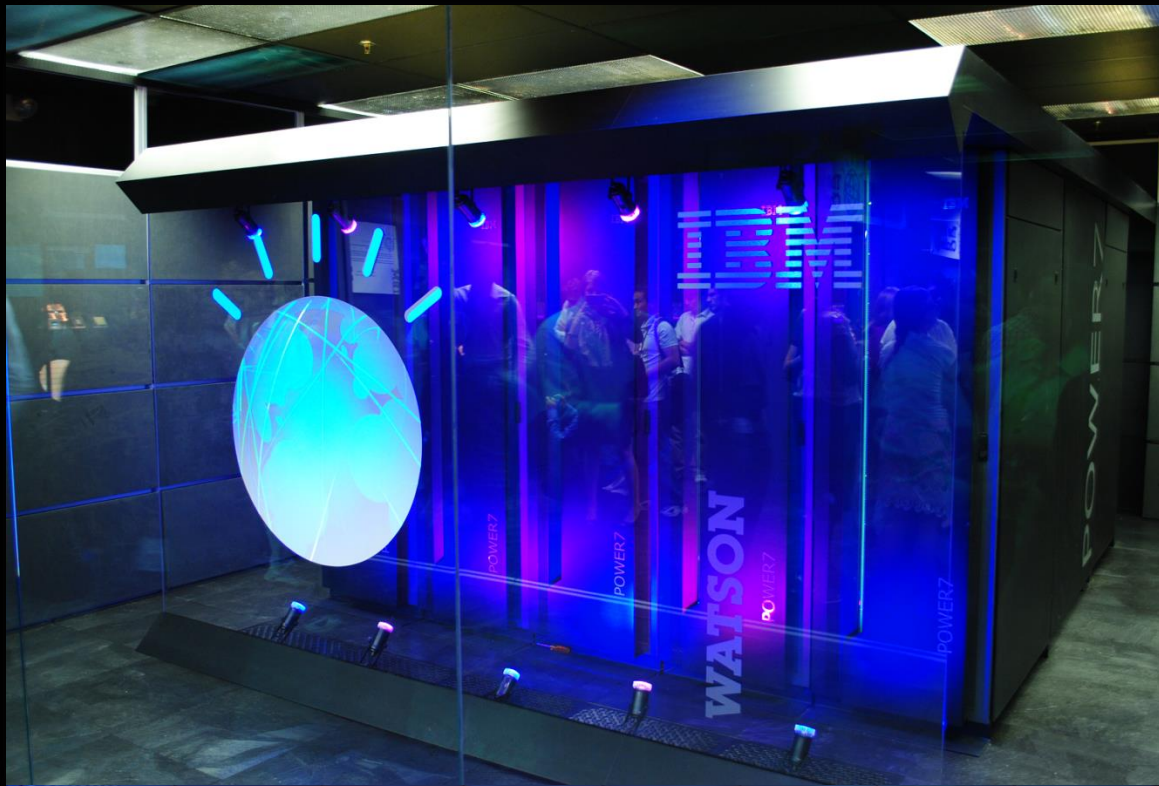
Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.



Watson (IBM)

- A question answering computer system capable of answering questions posed in natural language, developed in IBM's DeepQA project by a research team led by principal investigator David Ferrucci.
- Watson was named after IBM's first CEO and industrialist Thomas J. Watson.[
- Specifically developed to answer questions on the quiz show Jeopardy!.
- In 2011, Watson competed on Jeopardy! against former winners Brad Rutter and Ken Jennings.
- Watson received the first place prize of \$1 million.
- Accessed to 200 million pages of structured and unstructured content consuming four terabytes of disk storage, including the full text of Wikipedia. Not connected to the Internet during the game.
- For each clue, Watson's three most probable responses were displayed on the television screen.
- **Watson consistently outperformed its human opponents on the game's signaling device,**
- **But had trouble responding to a few categories, notably those having short clues containing only a few words.**



https://en.wikipedia.org/wiki/Watson_%28computer%29

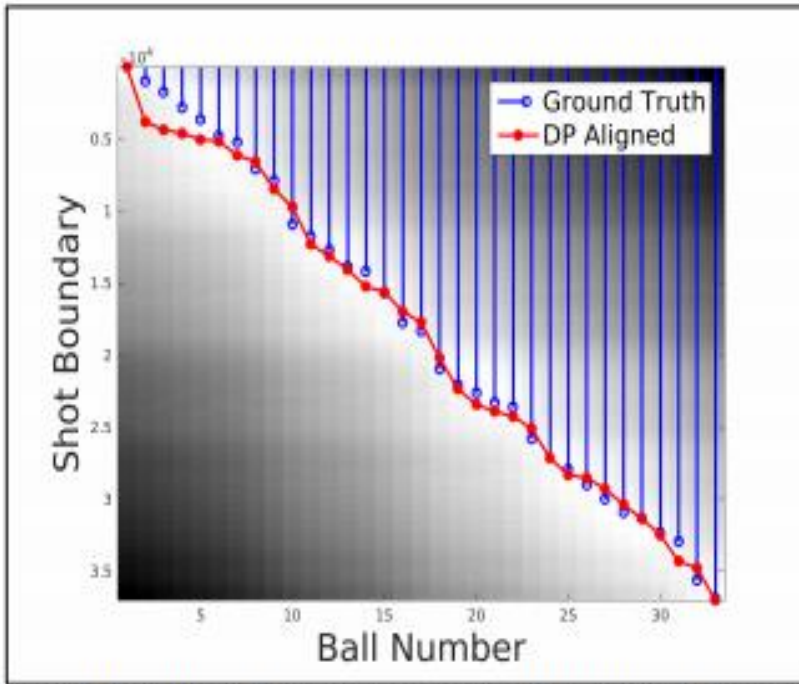


Figure 5. Results of Scene Segmentation depicted over five-overs.



on the pads once more David looks to nudge it to square leg off the pads for a leg-bye



that is again slammed towards long-on



shortish and swinging away lots of bounce

Figure 6. Examples of shots correctly labeled by the batsman or bowler actions.

Algoritmos capaces de comentar jugadas deportivas en directo

- Un sistema informático desarrollado en India es capaz de ofrecer comentarios textuales de partidos de críquet con un acierto del 90% en la retransmisión de las jugadas.
- Combina aprendizaje automático (*machine learning*) y un proceso tedioso para que las máquinas procesen horas y horas de juego.
- En deportes como el críquet ahora, pronto en el tenis...



Cornell University
Library

arXiv.org > cs > arXiv:1511.07607

Computer Science > Multimedia

Fine-Grain Annotation of Cricket Videos

Rahul Anand Sharma, Pramod Sankar K, CV Jawahar

(Submitted on 24 Nov 2015)

Precisión del algoritmo para cricket: 70-85%

<http://arxiv.org/pdf/1511.07607v1.pdf>

- **Methodology**

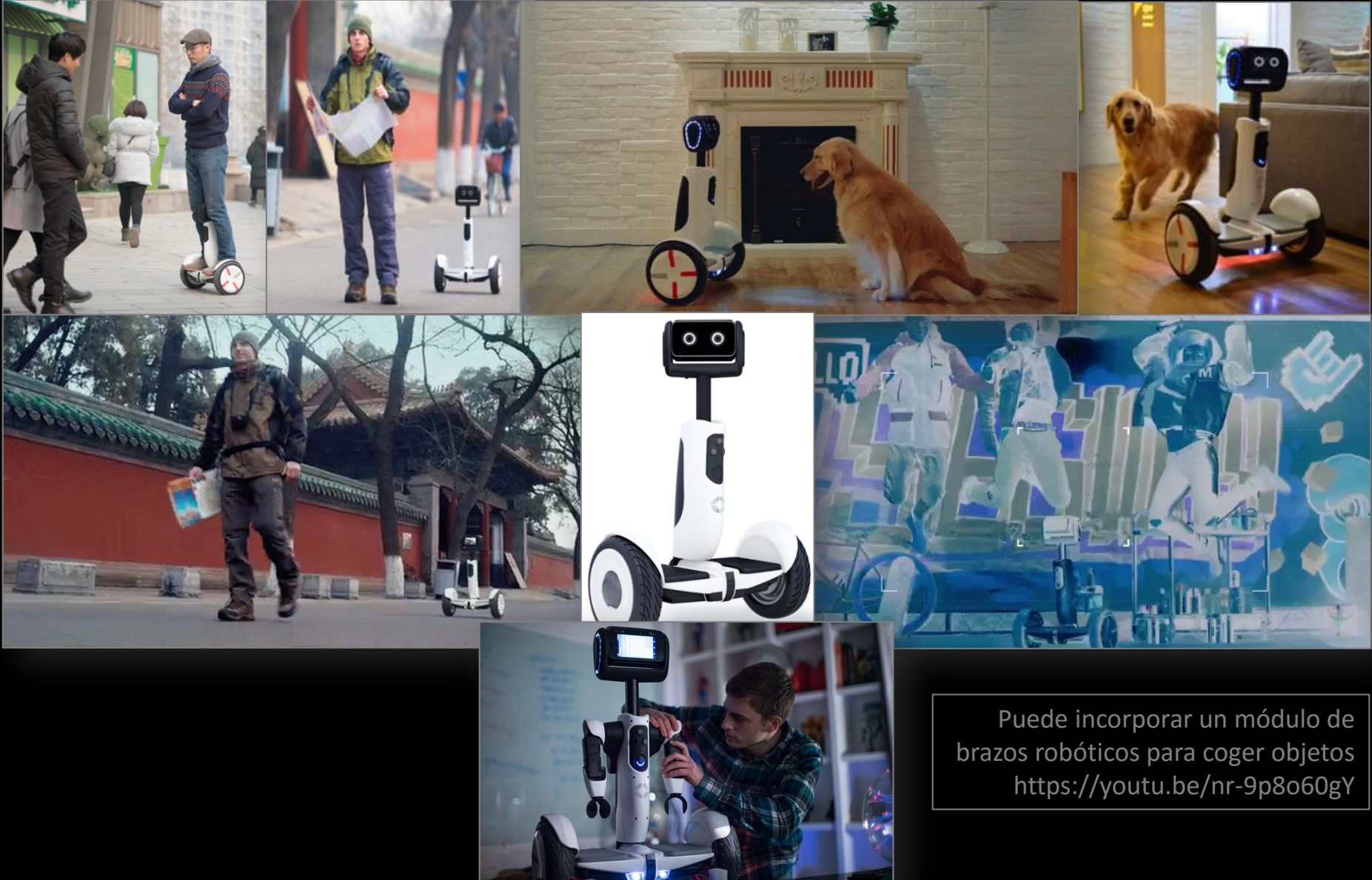
- Their approach is based solely on the team compositions.
- First, this method calculates a score for each player of the team based on their past performances. The team with the higher win probability is the predicted winner of that match.
- They enabled retrieval of specific actions that last only a few seconds, from several hours of video. This solution yields a large number of labeled exemplars, with no manual effort, that could be used by machine learning algorithms to learn complex actions.

- **Prediction accuracy**

- World T20 tournament , 2014
 - correctly predicted 15 out of 23 matches for a success rate of 65%.
- For the Cricket ODI World Cup 2015
 - correctly predicted 21 out of the first 31 matches (70-85%).

Adaptaciones de un dispositivo de ocio (Segway)

Sistema de transporte Segway (18 km/h, autonomía hasta 30 km), con sistema de reconocimiento de voz



Puede incorporar un módulo de brazos robóticos para coger objetos
<https://youtu.be/nr-9p8o60gY>

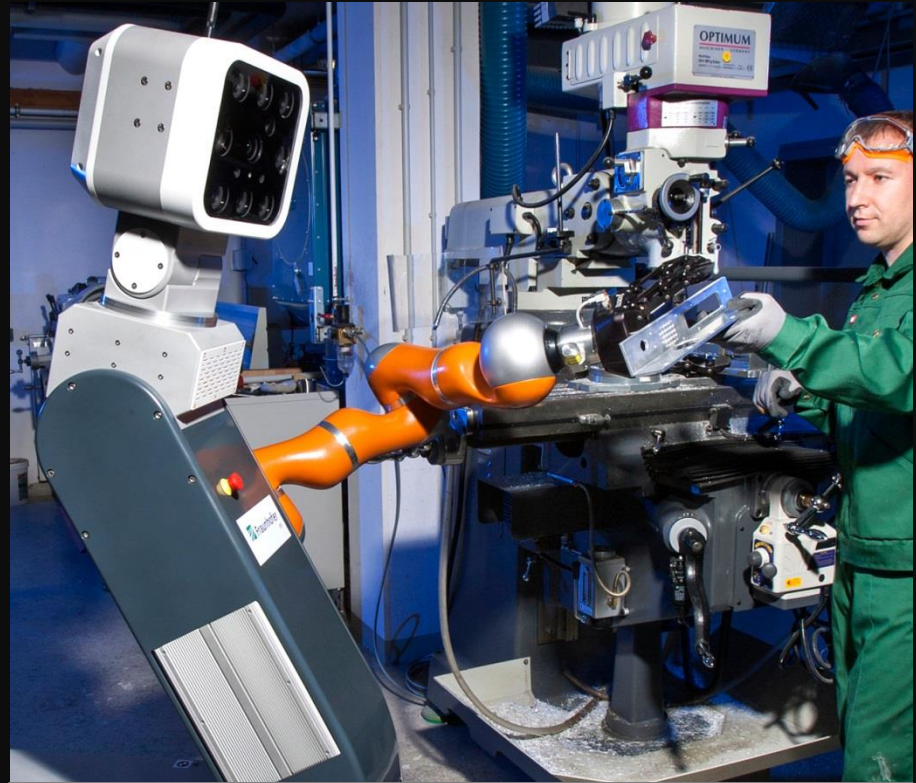
Norbert Elkmann: “Robotics is evolving away from classic industrial robots toward advanced assistive robots.”

Human-Robot Cooperation Conference, Magdeburg, Jun 25, 2013



KUKA presenting its newest lightweight robot at the Human-Robot Cooperation Conference at the Fraunhofer IFF.

© Viktoria Kühne
<http://www.iff.fraunhofer.de>



© Fraunhofer IFF

- Autonomous robot “Annie” to demonstrate present technical possibilities in the field of assistive robotics (Fraunhofer IFF).
- Its “artificial skin”, a tactile sensor system, enables the robot to “feel” the exact location and intensity of contact almost like a human and thus to stop its movement immediately whenever unplanned contact occurs.

Robots diseñados para jugar a un nivel aceptable

La cooperación humano-robot implica:

- Compartir la meta común de seguir jugando y hacer juicios calculados sobre cómo puede lograrse ese objetivo cooperando con el humano”,
- Es importante que la máquina *entienda* las condiciones de la persona y decida actuar para apoyarla.
- No se busca ganarle al jugador humano, sino el compartir metas y retos.



Timo Boll vs. KUKA Robot | 10 mar. 2014

KUKA robot faces off against one of the best table tennis players of all time

<https://youtu.be/tIJJME8-au8>

(sólo puede devolver las pelotas sin comenzar sacando).

<http://ef7bbb00.linkbucks.com>

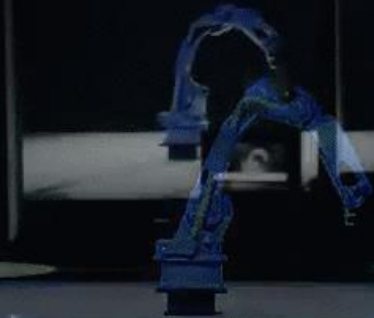
Adaptaciones de un robot industrial

Yaskawa Bushido Project | <https://youtu.be/l8xg9njJTaU>

Six-axis robot excels in assembly, dispensing, material handling, packaging, collaborative tasks.

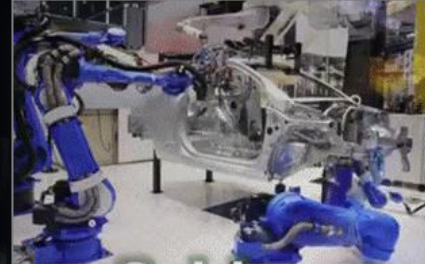
Motoman MH24

robot industrial



Isao Machii

maestro samurai



Soldar



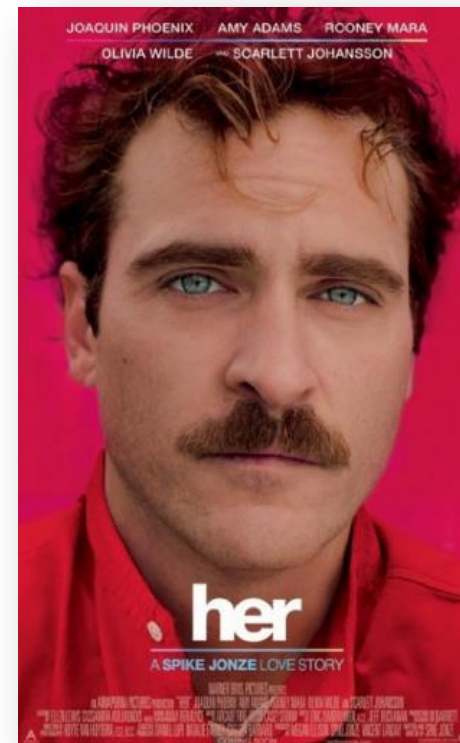
Optional features available



Asistentes personales virtuales

- **Siri** (motor de respuestas: Wolfram Alpha)
- **Google Now**
- **Cortana** (Bing)
- **Sherpa**
- **Speaktoit**

Her (Spike Jonze, 2013)



Escenarios de riesgo

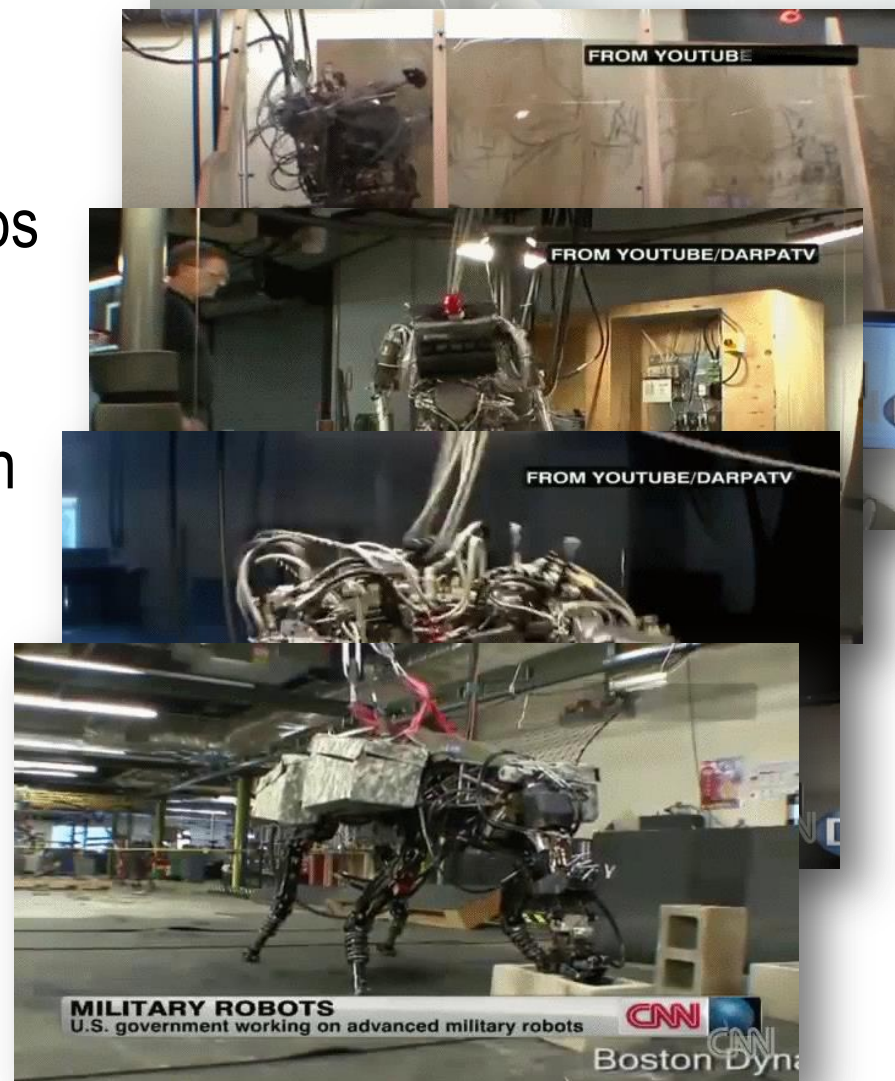
Desarrollo de armas robóticas autónomas

New Military Robots Show Off Strength

<https://www.youtube.com/watch?v=00NlrY3tHE4> (march 2013)

- **Robots de uso militar**
 - Los relatos de ciencia ficción distorsionan la percepción de los desarrollos reales.
 - Los robots en desarrollo con más interés para uso militar son UAV (drones), vehículos autónomos terrestres o marítimos y robots capaces de llevar con rapidez carga a zonas de difícil acceso.

(Cheetah Robot: 45.54 km/h)



Vehículos autónomos terrestres

<https://www.youtube.com/watch?v=PVT5TfFL40s>



Desarrollo de armas robóticas autónomas

<http://www.oxfordmartin.ox.ac.uk/downloads/briefings/Robo-Wars.pdf>

- **Robo-Wars: The Regulation of Robotic Weapons**
(Alex Leveringhaus, Gilles Giaccagive):
 - This report urges governments to recognise the **increasing prominence of robotic weapons in contemporary and future forms of warfare**, and proposes steps towards suitable regulation.
 - Analysis of technological dimensions of these weapons, their treatment under existing international, and legal and ethical frameworks.
 - **Regulatory options currently under discussion**
 - Need of recommendations for states, manufacturers and the military to develop a suitable framework.

Unmanned Aerial Vehicle (UAV)

<http://www.oxfordmartin.ox.ac.uk/downloads/briefings/Robo-Wars.pdf>

1.2 X47-B

U.S. Navy photo by Mass Communication Specialist
2nd Class Timothy Walter/Released.



- Developed by Northrop Grumman
- Uninhabited aerial vehicle
- Once programmed can take off from, and land on, an aircraft carrier without intervention by a human pilot
- Not currently equipped to carry a payload

1.3 Taranis

QinetiQ Group (source: Flickr). See <https://creativecommons.org/licenses/by-nc-nd/2.0/> for image licence.



- Developed by BAE Systems
- Unmanned aerial vehicle
- Stealth plane
- Not currently equipped to carry a payload
- Successfully tested by the UK Ministry of Defence and the UK's RAF in Australia in 2013

Unmanned Aerial Vehicles (UAV)

<https://www.youtube.com/watch?v=PVT5TfFL40s>



Vehículos autónomos de superficie

<http://www.w54.biz/showthread.php?1314-Unmanned-Surface-Vessels/page6>



NEANY's DragonSpy - UAV Boat
<https://www.youtube.com/watch?v=bb0A-uhaUqY>



Air-Droppable SeaScout UAV-USV (\$25M SBIR Contract for)



Unmanned Surface Vessels 2013
General Dynamics - Anti Submarine Warfare



B-7 unmanned surface vessel (USV)
Fully capable commercial as well as military applications

Las armas autónomas incrementan el peligro para civiles

Ej. de sistema autónomo utilizado en vigilancia de fronteras:

<https://www.hrw.org/news/2012/11/19/ban-killer-robots-its-too-late>



South Korea Defense Super aEgis II Robot Turret-Sentry
<https://www.hrw.org/news/2012/11/19/ban-killer-robots-its-too-late> (alcance: 3 km)

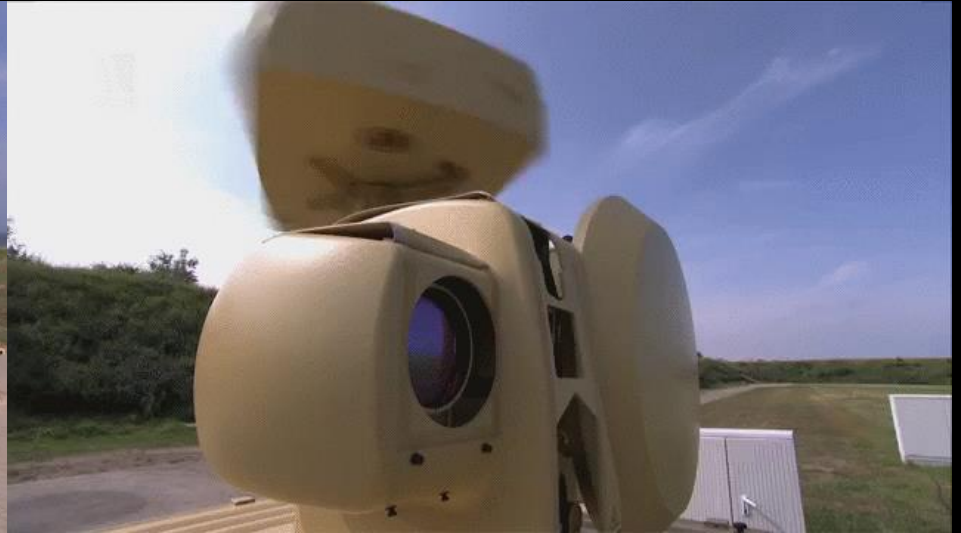
The South Korean SGR-1 sentry robot, a precursor to a fully autonomous weapon, can detect people in the Demilitarized Zone and, if a human grants the command, fire its weapons. The robot is shown here during a test with a surrendering enemy soldier.

© 2007 Getty Images

MANTIS - Sistema de defensa aérea de la Bundeswehr

Cañón de 35mm contra amenazas aéreas (misiles, proyectiles de artillería y morteros)

<https://www.youtube.com/watch?v=PMDJ0bwn4sE> (26 sept. 2012)



Diversos tipos de robots de uso militar

<https://www.youtube.com/watch?v=PVT5TfFL4Os> (25 ago. 2014)



Perspectiva realistas de disponer de drones capaces de tomar decisiones letales sin asistencia humana

<http://www.nationaljournal.com/s/69067/soon-drones-may-be-able-make-lethal-decisions-their-own>

NationalJournal

MONDAY, JANUARY 11, 2016

TWENTYSIXTEEN CONGRESS WHITE HOUSE POLITICS ENERGY HEALTH CARE DEFENSE TECH

DC INFO AGE THE NEXT AMERICA THE NEXT ECONOMY | SPONSORED: AMERICA'S ENERGY | MORE ▶

Soon, Drones May Be Able to Make Lethal Decisions on Their Own

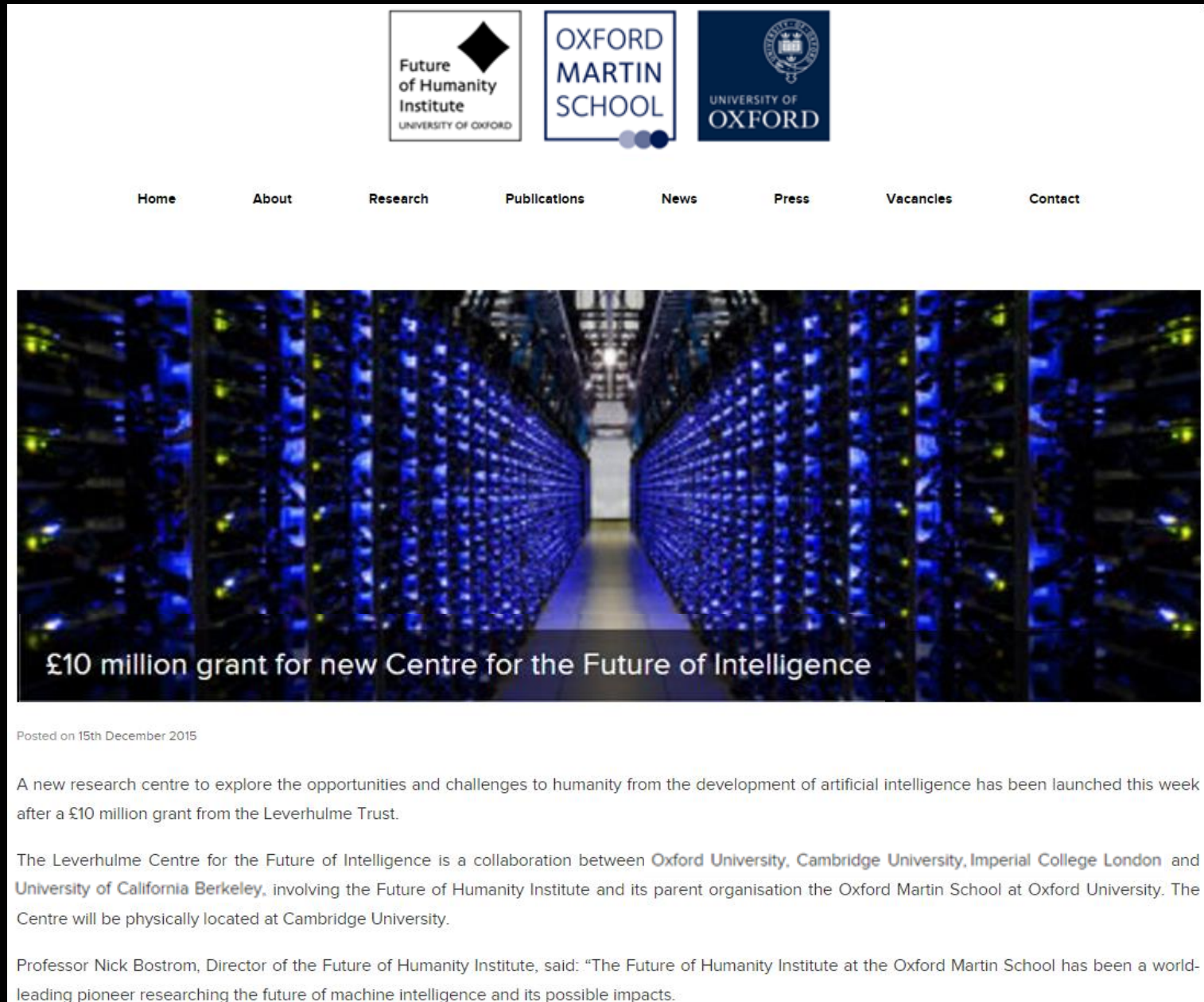
- Dotados de sistemas de AI para responder a una serie programada de entradas, seleccionar objetivos y disparar sus armas sin supervisión humana ni control del resultado.
- **Elementos a favor:** “Un avión no tripulado con un sistema suficientemente sofisticado podría ser **menos emocional, más selectivo y capaz de proporcionar sólo la fuerza necesaria** para lograr un objetivo táctico con el menor daño“
- “Un robot autónomo letal puede apuntar mejor, orientar mejor, seleccionar mejor, y, en general, ser una mejor baza del equipo de ISR [inteligencia, vigilancia y reconocimiento].”
- La decisión de desplegar robots autónomos letales (LAR) no puede dejarse sólo en manos del personal militar.

Dilución de responsabilidad usando LAR

<https://collateralmurder.wikileaks.org/>

- **Ni fabricante ni controlador serían responsables en escenario bélico**
 - Ej.: **caso denunciado por wikileaks**:
 - El **artillero de un helicóptero Apache estadounidense** mata con disparos de una ametralladora de 30 mm a dos periodistas de la agencia Reuters en Bagdad, el 12 de julio de 2007. Fueron asesinadas otras nueve personas y heridos dos niños.
 - **Las armas autónomas letales sólo deberían desplegarse en contextos donde la probabilidad de hallar seres humanos sea baja.**
 - A largo plazo, pueden hacer inmanejable la dinámica de los conflictos armados.
 - Plantean **riesgos tecnológicos, políticos y estratégicos de tal magnitud que refuerzan las estrategias viables a largo plazo para preservar y asegurar la paz.**
 - En el contexto actual, el despliegue de armas robóticas autónomas es contrario al derecho internacional: **reglas para entablar combate muy estrictas.**
 - Una vez desplegadas, **es muy probable que los LAR y la IA asociada no sean capaces de discernir adecuadamente entre combatientes y no combatientes, ni de evaluar si el uso de la fuerza es excesiva en un caso particular.**

Área de financiación prioritaria (2015: > 13 millones EUR)



Future of Humanity Institute
UNIVERSITY OF OXFORD

OXFORD MARTIN SCHOOL

UNIVERSITY OF OXFORD

[Home](#) [About](#) [Research](#) [Publications](#) [News](#) [Press](#) [Vacancies](#) [Contact](#)

£10 million grant for new Centre for the Future of Intelligence

Posted on 15th December 2015

A new research centre to explore the opportunities and challenges to humanity from the development of artificial intelligence has been launched this week after a £10 million grant from the Leverhulme Trust.

The Leverhulme Centre for the Future of Intelligence is a collaboration between [Oxford University](#), [Cambridge University](#), [Imperial College London](#) and [University of California Berkeley](#), involving the Future of Humanity Institute and its parent organisation the Oxford Martin School at Oxford University. The Centre will be physically located at Cambridge University.

Professor Nick Bostrom, Director of the Future of Humanity Institute, said: "The Future of Humanity Institute at the Oxford Martin School has been a world-leading pioneer researching the future of machine intelligence and its possible impacts."

Adquisición estratégica de Apple: Emotient Inc.

<http://www.wsj.com/articles/apple-buys-artificial-intelligence-startup-emotient-1452188715>

THE WALL STREET JOURNAL

TECH

Apple Buys Artificial-Intelligence Startup Emotient

Emotient technology is used to assess emotions by reading facial expressions

By **ROLFE WINKLER**, **DAISUKE WAKABAYASHI** and **ELIZABETH DWOSKIN**

Updated Jan. 7, 2016 1:13 p.m. ET

Apple Inc. has purchased Emotient Inc., a startup that uses artificial-intelligence technology to read people's emotions by analyzing facial expressions.

EMOTIENT™

Emotions Predict Attitudes and Actions

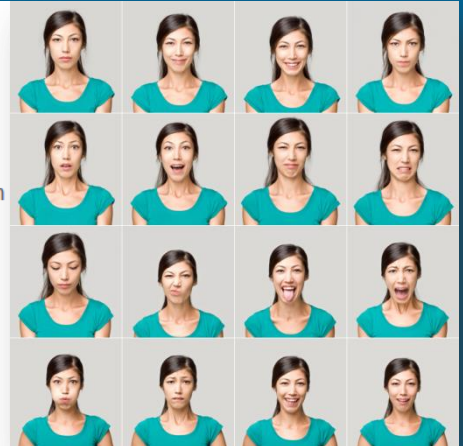
Emotient delivers insights in the form of three key performance indicators (KPIs), and detailed data on emotions, for every face in each video frame:

Emotient KPIs

Attention – Is your advertising or product getting noticed?

Engagement – Are people responding emotionally?

Sentiment – Are they showing positive, negative or no emotion?



Adquisición estratégica de Google: Deep Mind

<http://recode.net/2014/01/26/exclusive-google-to-buy-artificial-intelligence-startup-deepmind-for-400m/>

Exclusive: Google to Buy Artificial Intelligence Startup DeepMind for \$400M January 26, 2014

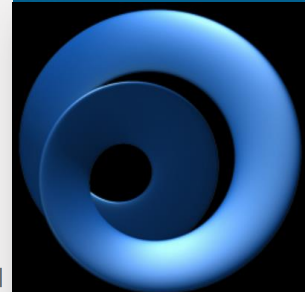
DEEPMIND IS EXCITED TO
HAVE JOINED FORCES
WITH **GOOGLE**.

Founded in 2011 by **Demis Hassabis, Shane Legg** and **Mustafa Suleyman**.
The team is based in London and was supported by some of the most iconic
technology entrepreneurs and investors of the past decade.

OUR MISSION

SOLVE INTELLIGENCE

We combine the best techniques from machine learning and systems neuroscience to build powerful
general-purpose learning algorithms.



DEEPMIND

Adquisición estratégica de Google: Deep Mind

<http://www.deepmind.com/publications.html>



[WHO WE ARE](#) [OUR MISSION](#)

Nature 2015

HUMAN LEVEL CONTROL THROUGH DEEP REINFORCEMENT LEARNING

This paper describes a Deep Q-Network (DQN), which is able to master a diverse range of Atari 2600 games, through combining Deep Neural Networks with Reinforcement Learning.

[READ MORE](#) [NEWS & VIEWS](#)

WE ARE HIRING!

arXiv 2014

NEURAL TURING MACHINES

Neural Turing Machines (NTMs) couple differentiable, external memory resources to neural network controllers. Unlike classical computers, they can be optimized by stochastic gradient descent to infer algorithms from data.

[READ MORE](#)

NIPS 2015

TEACHING MACHINES TO READ AND COMPREHEND

We define a new methodology for capturing large scale supervised reading comprehension data, as well as novel mechanisms for teaching machines to read and comprehend.

ICML 2015

DRAW: A RECURRENT NEURAL NETWORK FOR IMAGE GENERATION

This paper introduces the Deep Recurrent Attentive Writer (DRAW) architecture for image generation with neural networks.

ICML 2015

UNIVERSAL VALUE FUNCTION APPROXIMATORS

UVFAs jointly represent many goals/rewards simultaneously and generalize to unseen ones; a factored embedding approach makes training efficient.

Facebook y Tesla: Vicarious FPC (29 mill. EUR)

<http://blogs.wsj.com/digits/2014/03/21/zuckerberg-musk-invest-in-artificial-intelligence-company-vicarious>

Digits

Tech News & Analysis
From the WSJ

7:19 am ET
Mar 21, 2014 BIG DATA

Zuckerberg, Musk Invest in Artificial-Intelligence Company

ARTICLE VIDEO COMMENTS (18)

ARTIFICIAL INTELLIGENCE ELON MUSK FACEBOOK MARK ZUCKERBERG VICARIOUS

Email Print  1.4k 

By REED ALBERGOTTI [CONNECT](#)

Elon Musk made the electric car cool. **Mark Zuckerberg** created **Facebook**. Ashton Kutcher portrayed **Apple** founder **Steve Jobs** in a movie. Now, the three are joining in a \$40 million investment in Vicarious FPC, a secretive artificial-intelligence company.



- Replicar el neocórtex en un sistema informático capaz de entender el lenguaje y realizar cálculos.
- Diseñar un ordenador que piensa como un ser humano.
- Facilitar el reconocimiento facial en una gran base de datos inteligente.

Debate ético

Algoritmos sustitutivos del juicio humano
Exclusión masiva del mercado laboral
Nuevas obligaciones morales (IAG)

Crerios para algoritmos sustitutivos del juicio humano

Bostrom, N., & Yudkowsky, E. (2013). The ethics of artificial intelligence. In *Cambridge Handbook of Artificial Intelligence* (p. chap. 15).

- **Exigencias aplicables a un algoritmo destinado a sustituir al juicio humano (tareas sociales)**
 - Responsabilidad (garantías de no ofuscación)
 - Transparencia
 - Auditabilidad
 - Incorruptibilidad
 - Previsibilidad
 - Override/Procedimientos fiables para evitar víctimas inocentes
- **Debates pendientes** en una sociedad cada vez más informatizada

Quienes financian establecen los objetivos

- **Financiación pública de la I+D en robótica e IA**
 - Es importante no dejar a la iniciativa privada el diseño, desarrollos y aplicaciones en robótica avanzada e IA.
 - Los estados terminan pagando doblemente por sus aplicaciones y servicios: tanto la investigación y capacitación de personal cualificado como los derechos por patentes industriales de terceros.
- **La sociedad industrial que hemos conocido desde el s. XIX llega a su fin**
 - El poder deriva hoy de la capacidad que aportan máquinas complejas, sistemas avanzados de IA y los equipos de ingeniería necesarios para diseñarlos y mejorarlos.
 - Es previsible que el mercado laboral desplace masivamente incluso a trabajadores muy cualificados (será el escenario y trasfondo de conflictos importantes).

Referencias y bibliografía

Frankish, K., & Ramsey, W. M. (2014). *The Cambridge Handbook of Artificial Intelligence*. Cambridge University Press.

Bostrom, N. (2014). *Superintelligence: paths, dangers, strategies*. Oxford University Press.

Agar, N. (2010). *Humanity's End. Why We Should Reject Radical Enhancement*. Cambridge, Massachusetts / London: A Bradford Book - The MIT Press.

Referencias y bibliografía

- Ashrafian, H. (2015a). AlonAI: A Humanitarian Law of Artificial Intelligence and Robotics. *Science and Engineering Ethics*.
- Ashrafian, H. (2015b). Artificial Intelligence and Robot Responsibilities: Innovating Beyond Rights. *Science and Engineering Ethics*.
- Azofra, D., Martínez, E., Jiménez, E., Blanco, J., Azofra, F., & Saenz-Díez, J. C. (2015). Comparison of the influence of photovoltaic and wind power on the Spanish electricity prices by means of artificial intelligence techniques. *Renewable and Sustainable Energy Reviews*, 42(0), 532–542. doi:<http://dx.doi.org/10.1016/j.rser.2014.10.048>
- Baum, S. D., Denkenberger, D. C., & Haqq-Misra, J. (n.d.). Isolated refuges for surviving global catastrophes. *Futures*. doi:<http://dx.doi.org/10.1016/j.futures.2015.03.009>
- Dirican, C. (2015). The Impacts of Robotics, Artificial Intelligence On Business and Economics. *Procedia - Social and Behavioral Sciences*, 195, 564–573. doi:<http://dx.doi.org/10.1016/j.sbspro.2015.06.134>
- Fahn, M., & Hadjer, T. (2015). Optimal contracting with private military and security companies. *European Journal of Political Economy*, 37, 220–240. doi:<http://dx.doi.org/10.1016/j.ejpoleco.2014.10.004>
- Farish, M. (2015). Canons and wars: American military geography and the limits of disciplines. *Journal of Historical Geography*, 49, 39–48. doi:<http://dx.doi.org/10.1016/j.jhg.2015.04.012>
- Fong, S., Deb, S., & Chaudhary, A. (2015). A review of metaheuristics in robotics. *Computers & Electrical Engineering*, 43, 278–291. doi:<http://dx.doi.org/10.1016/j.compeleceng.2015.01.009>.

Referencias y bibliografía

- Ghiassi, M., Lio, D., & Moon, B. (2015). Pre-production forecasting of movie revenues with a dynamic artificial neural network. *Expert Systems with Applications*, 42(6), 3176–3193. doi:<http://dx.doi.org/10.1016/j.eswa.2014.11.022>
- Gonzalez, D. (2015). Chapter 10 - The Future of Online Security. In D. B. T.-M. O. R. Gonzalez (Ed.), (pp. 237–258). Boston: Butterworth-Heinemann. doi:<http://dx.doi.org/10.1016/B978-0-12-420055-5.00010-4>
- Gribaudo, M., Iacono, M., & Marrone, S. (2015). Exploiting Bayesian Networks for the Analysis of Combined Attack Trees. *Electronic Notes in Theoretical Computer Science*, 310, 91–111. doi:<http://dx.doi.org/10.1016/j.entcs.2014.12.014>
- Huijbrechts, B., Velikova, M., Michels, S., & Scheepens, R. (2015). Metis1: An Integrated Reference Architecture for Addressing Uncertainty in Decision-support Systems. *Procedia Computer Science*, 44, 476–485. doi:<http://dx.doi.org/10.1016/j.procs.2015.03.007>
- Kant, G., & Sangwan, K. S. (2015). Predictive Modeling for Power Consumption in Machining Using Artificial Intelligence Techniques. *Procedia CIRP*, 26(0), 403–407. doi:<http://dx.doi.org/10.1016/j.procir.2014.07.072>
- Knowles, W., Prince, D., Hutchison, D., Disso, J. F. P., & Jones, K. (2015). A survey of cyber security management in industrial control systems. *International Journal of Critical Infrastructure Protection*, 9, 52–80. doi:<http://dx.doi.org/10.1016/j.ijcip.2015.02.002>
- Luxton, D. D. (2016). Chapter 1 - An Introduction to Artificial Intelligence in Behavioral and Mental Health Care. In D. D. B. T.-A. I. in B. and M. H. C. Luxton (Ed.), (pp. 1–26). San Diego: Academic Press. doi:<http://dx.doi.org/10.1016/B978-0-12-420248-1.00001-5>

Referencias y bibliografía

- Majot, A., & Yampolskiy, R. (2015). Global catastrophic risk and security implications of quantum computers. *Futures*, 72, 17–26. doi:10.1016/j.futures.2015.02.006
- Neuman, Y. (2016). Chapter 10 - Artificial Intelligence in Public Health Surveillance and Research. In D. D. B. T.-A. I. in B. and M. H. C. Luxton (Ed.), (pp. 231–254). San Diego: Academic Press. doi:http://dx.doi.org/10.1016/B978-0-12-420248-1.00010-6
- Poulin, C., Thompson, P., & Bryan, C. (2016). Chapter 9 - Public Health Surveillance: Predictive Analytics and Big Data. In D. D. B. T.-A. I. in B. and M. H. C. Luxton (Ed.), (pp. 205–230). San Diego: Academic Press. doi:http://dx.doi.org/10.1016/B978-0-12-420248-1.00009-X
- Silverman, B. G., Hanrahan, N., Huang, L., Rabinowitz, E. F., & Lim, S. (2016). Chapter 7 - Artificial Intelligence and Human Behavior Modeling and Simulation for Mental Health Conditions. In D. D. B. T.-A. I. in B. and M. H. C. Luxton (Ed.), (pp. 163–183). San Diego: Academic Press. doi:http://dx.doi.org/10.1016/B978-0-12-420248-1.00007-6
- Veruggio, G. (n.d.). Roboethics. Retrieved April 6, 2015, from file:///C:/Users/mig/Documents/My Dropbox/_Ahora/AI/veruggio_Roboethics_2005.pdf

Gracias

mm3@ugr.es