



IDAL

Intelligent Data Analysis Laboratory

Ciencia de datos, Big data...

¡Esto no puede pasarme a mí!

Rafael Magdalena

Departamento de Ingeniería Electrónica

ETSE, Avda Universitat s/n, 46100

rafael.magdalena@uv.es

<http://idal.uv.es>

VNIVERSITAT
ID VALÈNCIA

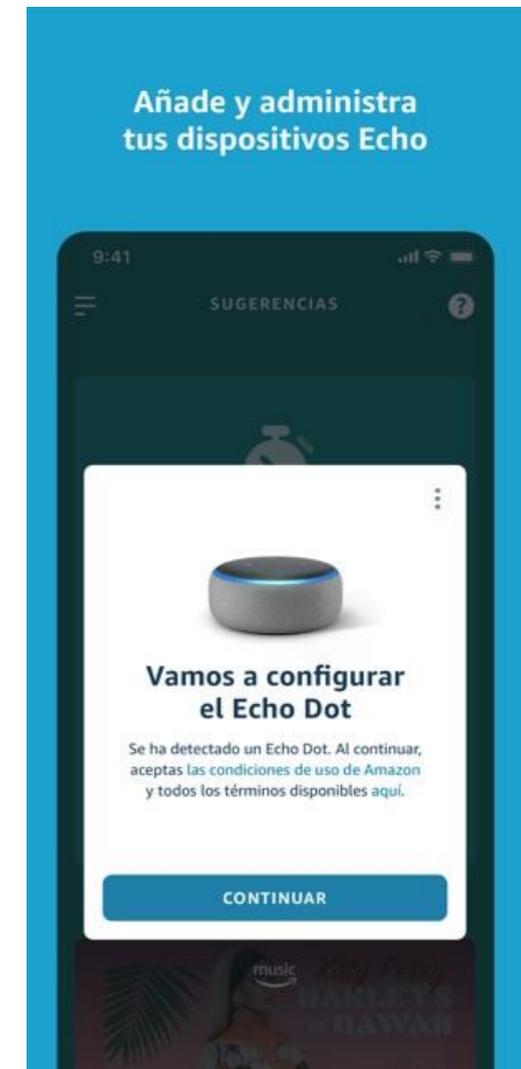


Escola Tècnica Superior d'Enginyeria

Departament d'Enginyeria Electrònica



¿Alguna vez te has preguntado?



Nuestro mundo son datos







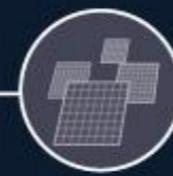
Transportation



Natural Resources



Manufacturing



Renewable Energy



Smart Cities



By 2020, there will be **50 billion devices** connected to the internet.

Things Connected to the Internet

2005

500 million

2010

12.5 billion

2015

25 billion

2020

50 billion

People on Earth

6.3 billion

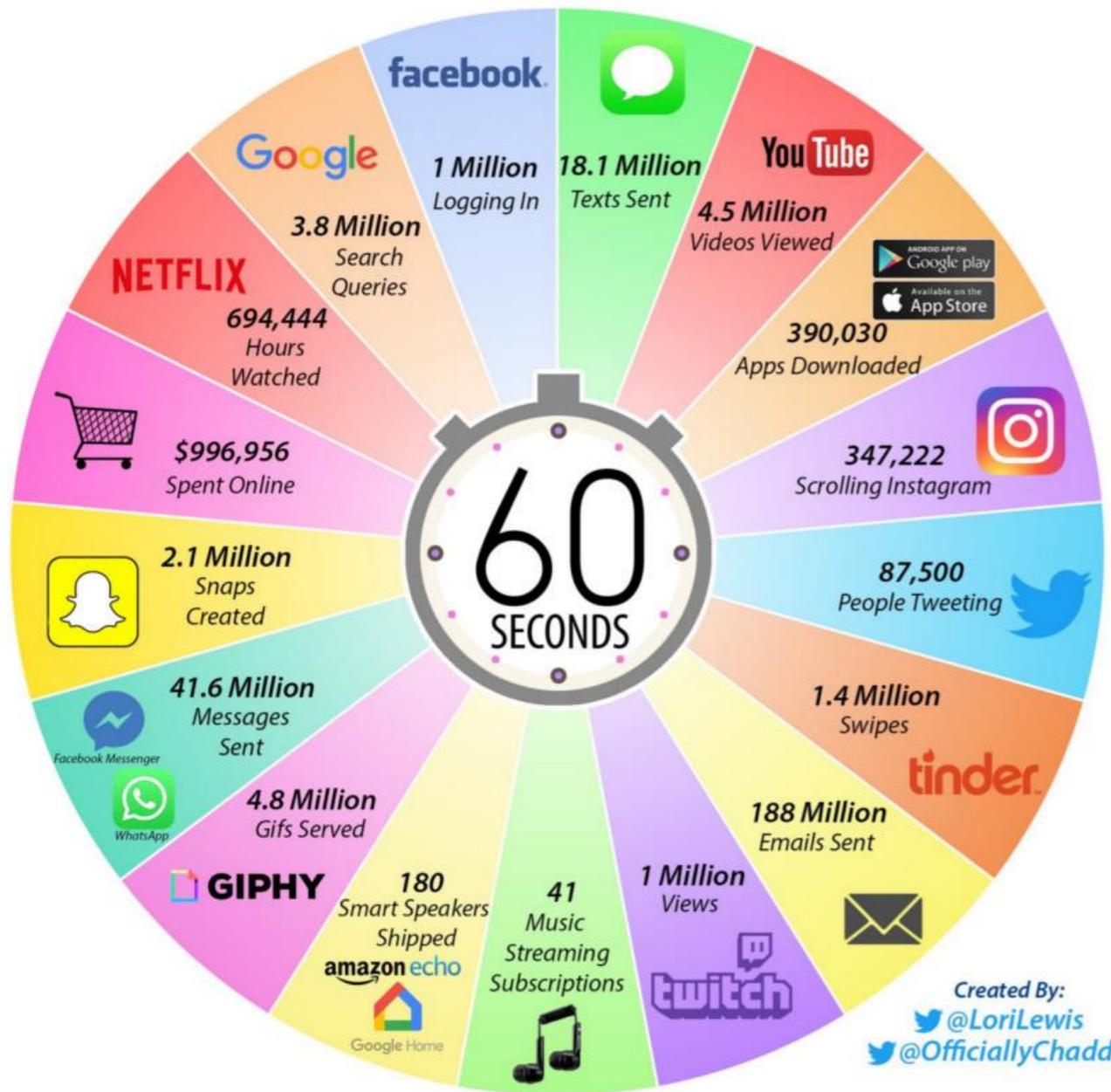
6.8 billion

7.2 billion

7.6 billion

Source: Cisco IBSG

2019 *This Is What Happens In An Internet Minute*



2020 *This Is What Happens In An Internet Minute*

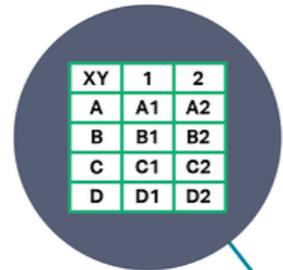


Structured Data

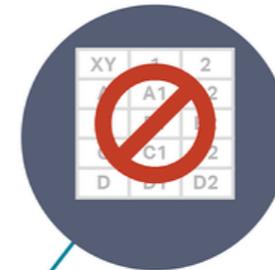
VS

Unstructured Data

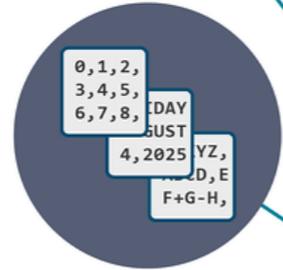
Can be displayed in rows, columns and relational databases



Cannot be displayed in rows, columns and relational databases



Numbers, dates and strings



Images, audio, video, word processing files, e-mails, spreadsheets



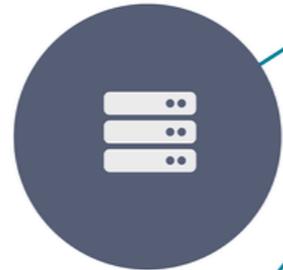
Estimated 20% of enterprise data (Gartner)



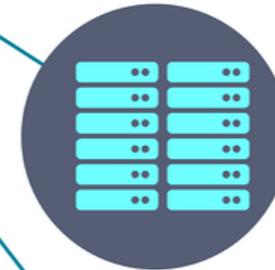
Estimated 80% of enterprise data (Gartner)



Requires less storage



Requires more storage



Easier to manage and protect with legacy solutions



More difficult to manage and protect with legacy solutions



Bruce Schneier

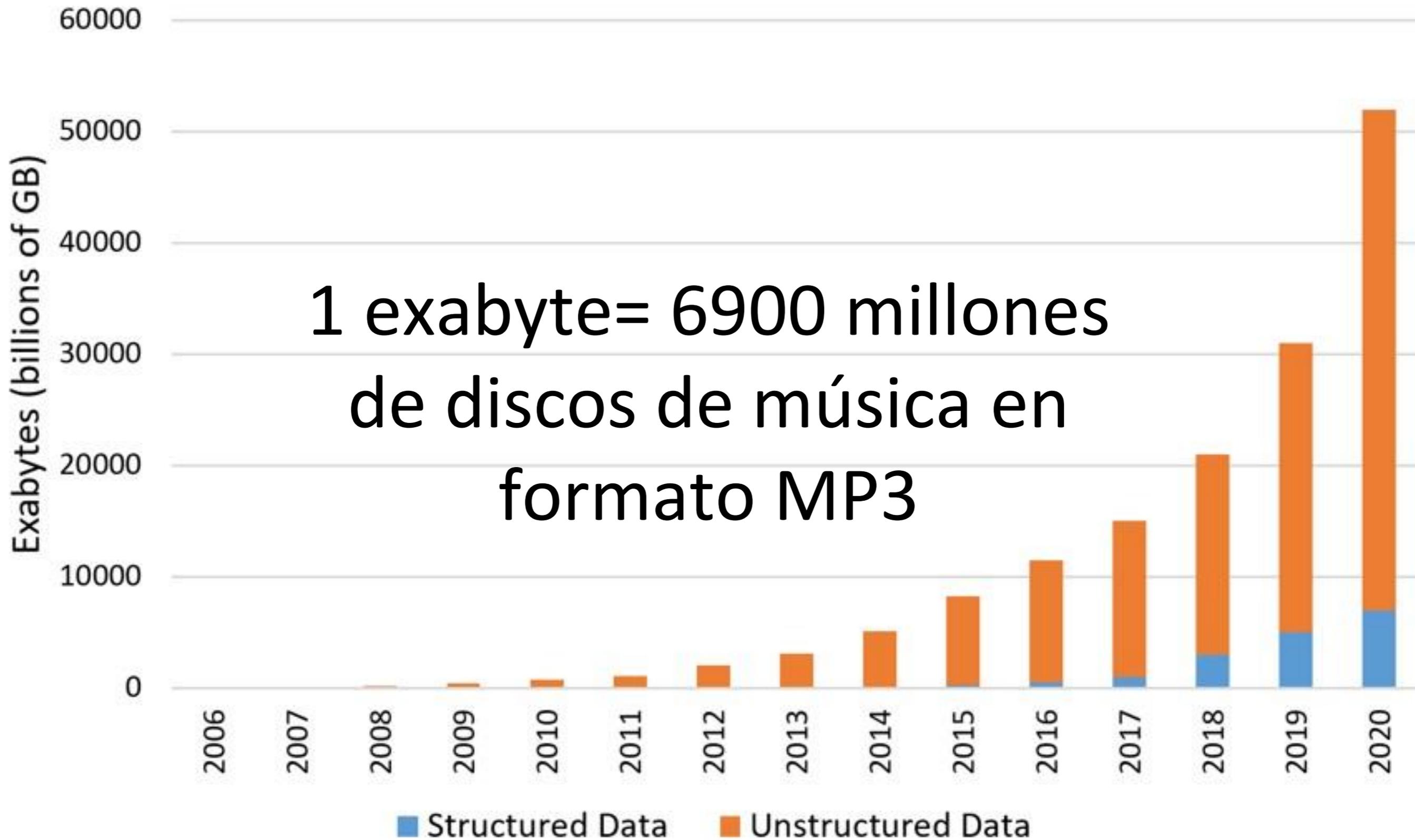
HAZ CLIC AQUÍ PARA
**MATARLOS
A TODOS**

UN MANUAL DE SUPERVIVENCIA

IBLACK
MIRROR
YA ESTÁ
AQUÍ!



The Cambrian Explosion...of Data



La explosión informática



La ley de Moore

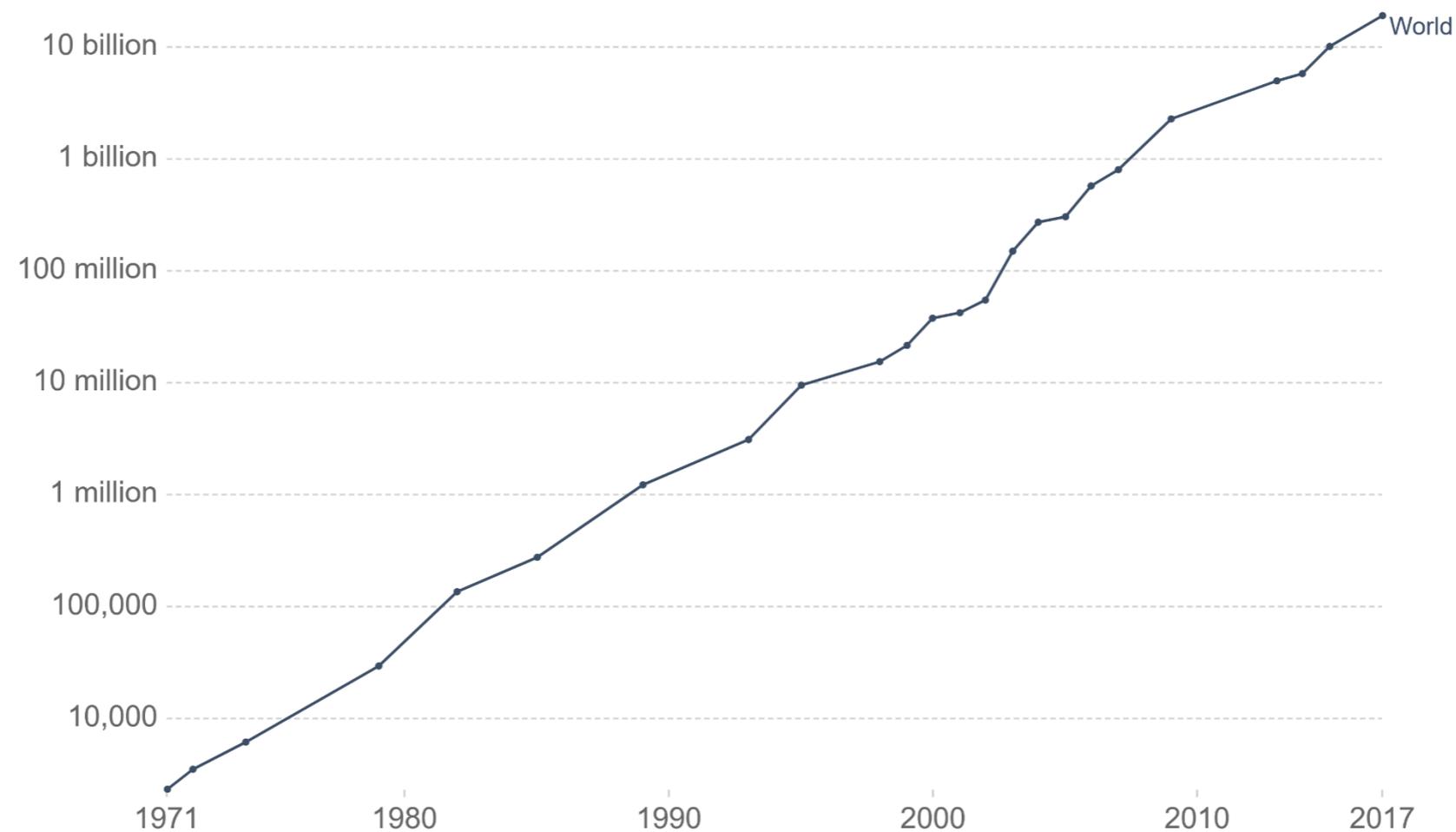


La ley de Moore es la observación (empírica) de que cada dos años se dobla el número de transistores incluido en un procesador

Moore's Law: Transistors per microprocessor

Number of transistors which fit into a microprocessor. This relationship was famously related to Moore's Law, which was the observation that the number of transistors in a dense integrated circuit doubles approximately every two years.

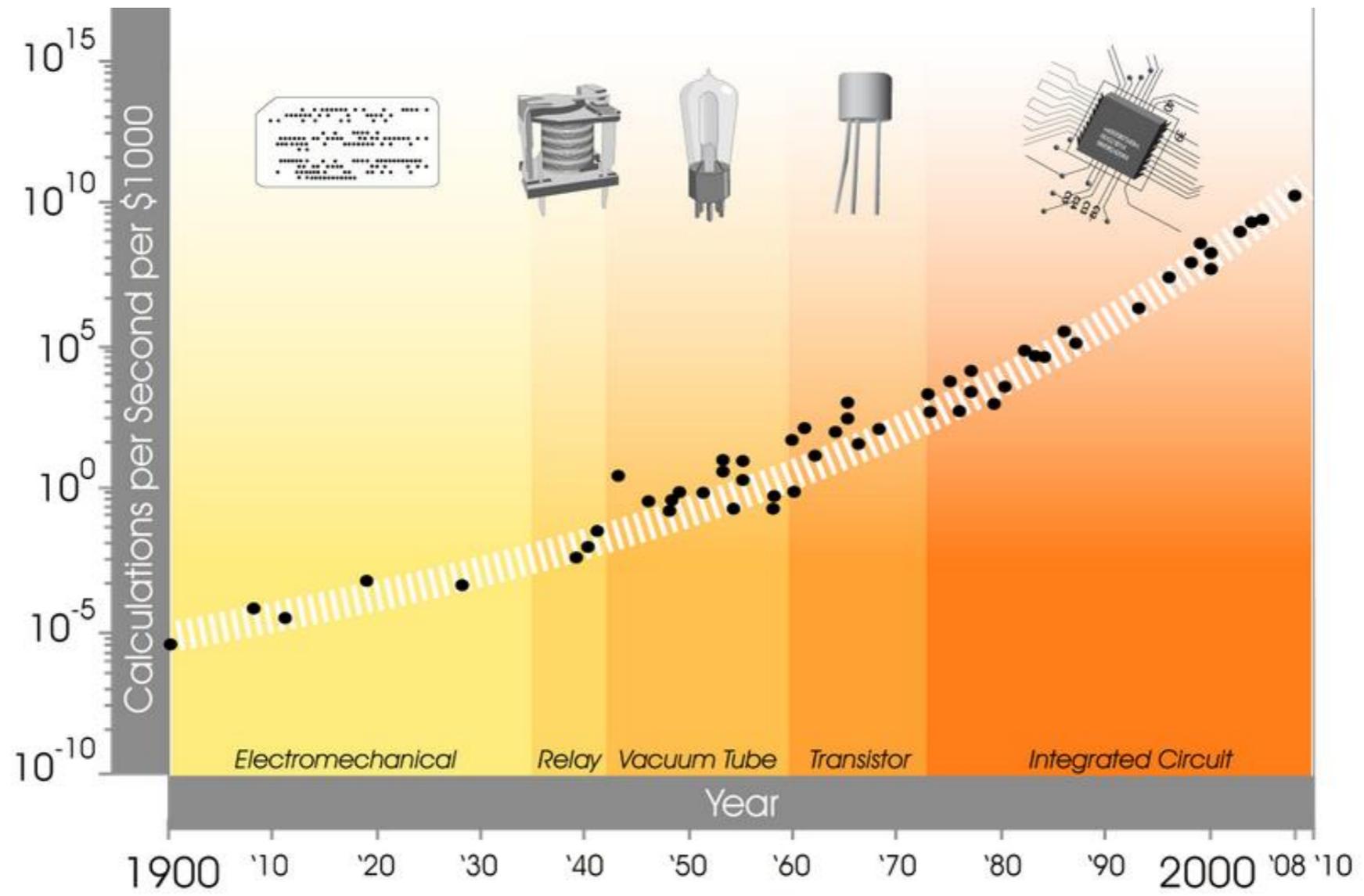
Our World
in Data



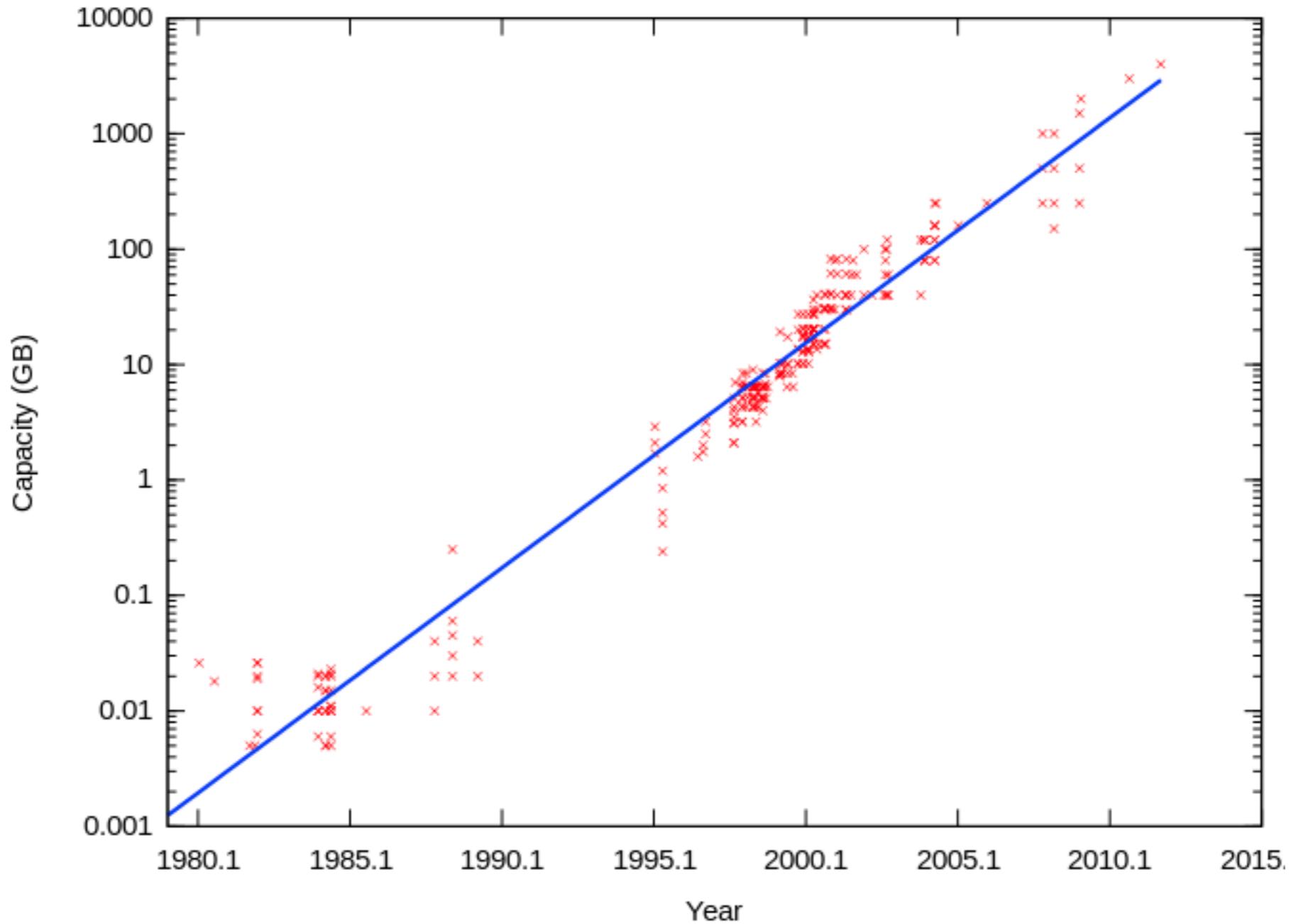
Source: Karl Rupp. 40 Years of Microprocessor Trend Data.

CC BY

...pero el precio no sube



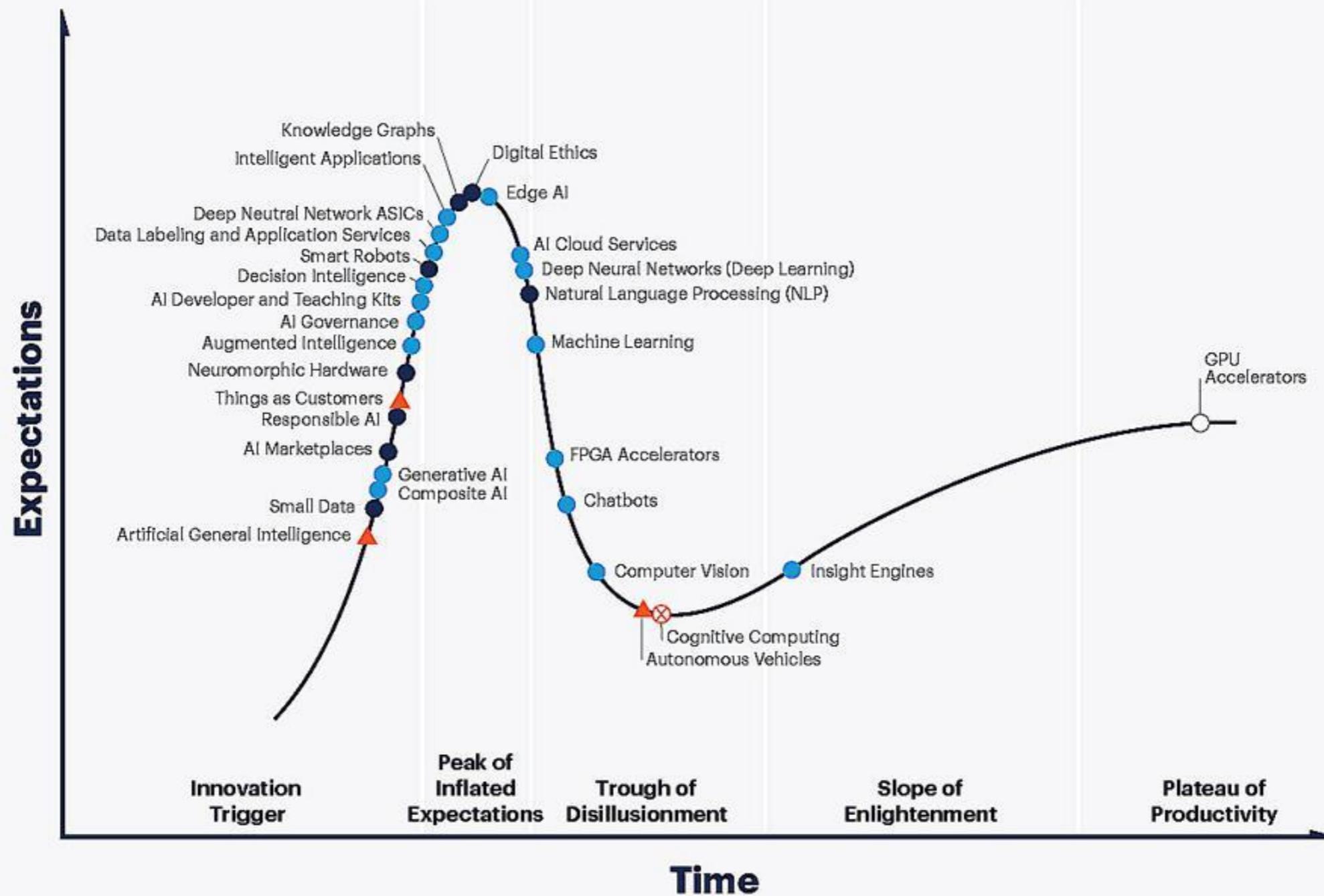
Lo mismo pasa con el almacenamiento



Internet de las Cosas (*Internet of Things, IoT*)



Hype Cycle for Artificial Intelligence, 2020



Plateau will be reached:

○ less than 2 years

● 2 to 5 years

● 5 to 10 years

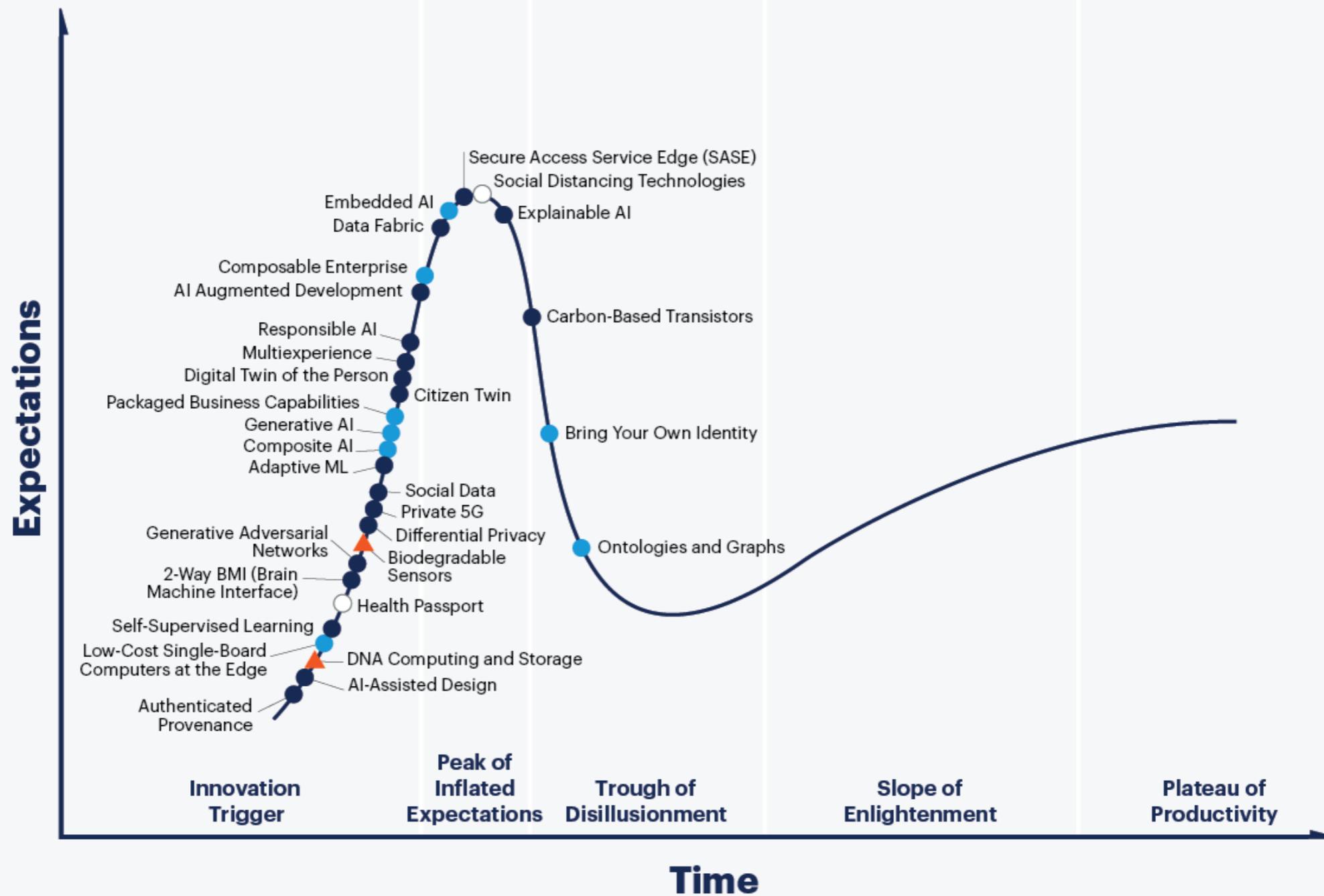
▲ more than 10 years

⊗ obsolete before plateau

As of July 2020

gartner.com/SmarterWithGartner

Hype Cycle for Emerging Technologies, 2020



Plateau will be reached:

○ less than 2 years

● 2 to 5 years

● 5 to 10 years

▲ more than 10 years

⊗ obsolete before plateau

As of July 2020

gartner.com/SmarterWithGartner

Article Full-text available

How Many Variables Can Humans Process?

February 2005 · Psychological Science 16(1):70-6 · [Follow journal](#)

DOI: [10.1111/j.0956-7976.2005.00782.x](https://doi.org/10.1111/j.0956-7976.2005.00782.x)

Source · [PubMed](#)

Graeme S. Halford · Rosemary Baker · Julie McCredden · John D Bain

Research Interest ⓘ 181.4

Citations ⓘ 327

Recommendations ⓘ 0 new 0

Reads ⓘ 10 new 1,547

[See details](#)

Overview

Stats

Comments

Citations (327)

References (24)

Related research (10+)

Download



Abstract and figures

The conceptual complexity of problems was manipulated to probe the limits of human information processing capacity. Participants were asked to interpret graphically displayed statistical interactions. In such problems, all independent variables need to be considered together, so that decomposition into smaller subtasks is constrained, and thus the order of the interaction directly determines conceptual complexity. As the order of the interaction increases, the number of variables increases. Results showed a significant decline in accuracy and speed of solution from three-way to four-way interactions. Furthermore, performance on a five-way interaction was at chance level. These findings suggest that a structure defined on four variables is at the limit of human processing capacity.

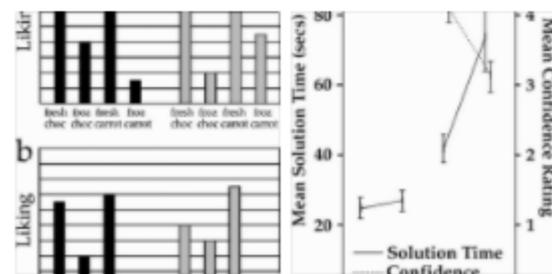


Figure content uploaded by [John D Bain](#) Author content

Content may be subject to copyright.



Ad v

Funding available of up to 200,000 €

Solve our Myeloid-Derived Growth Factor (MYDGF) question!



Apply and Collaborate

How would you identify the molecular targets of MYDGF?

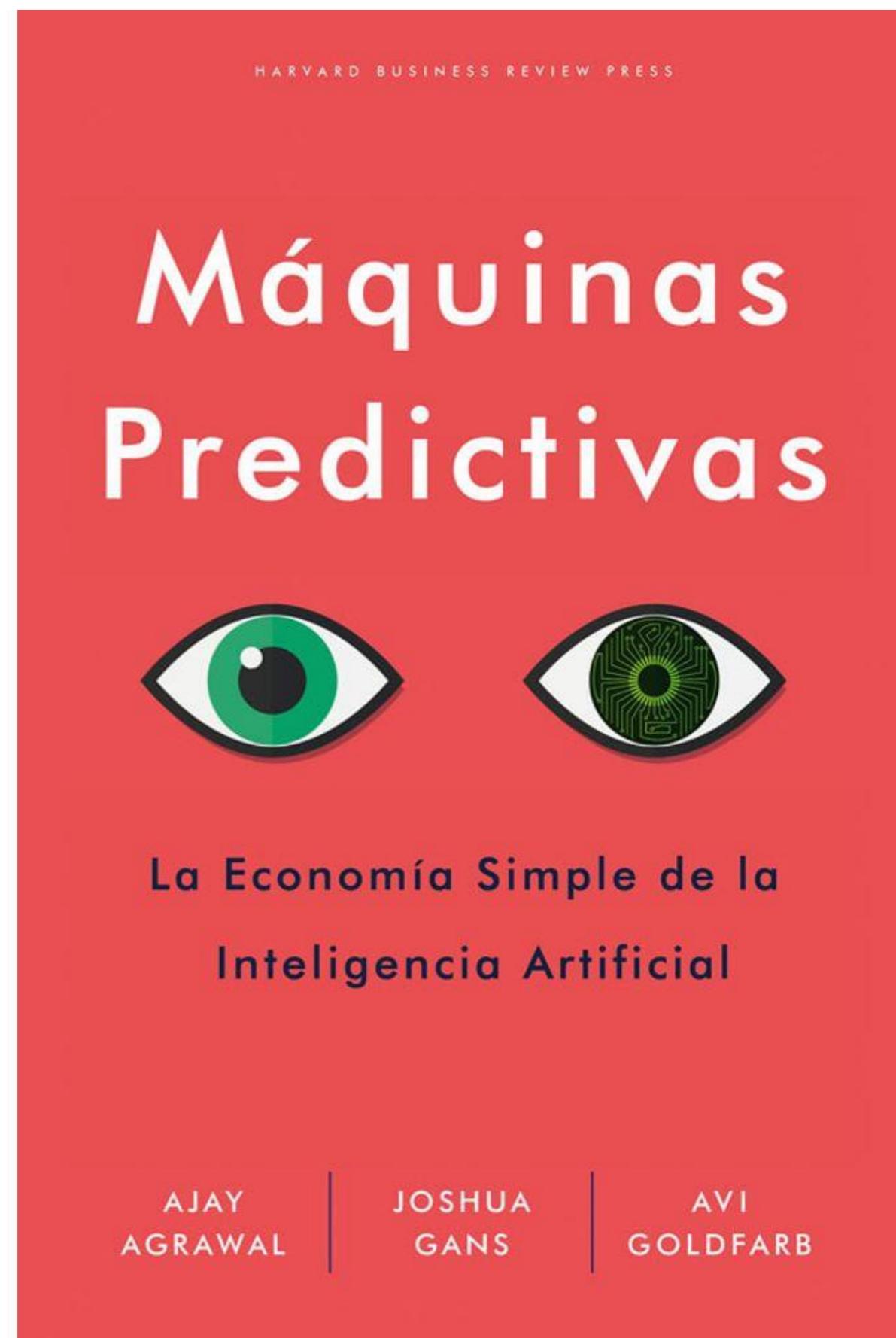
Get started



Related research

Los humanos no damos para más

Los sistemas de Inteligencia Artificial son máquinas que realizan predicciones usando información pasada en forma de datos



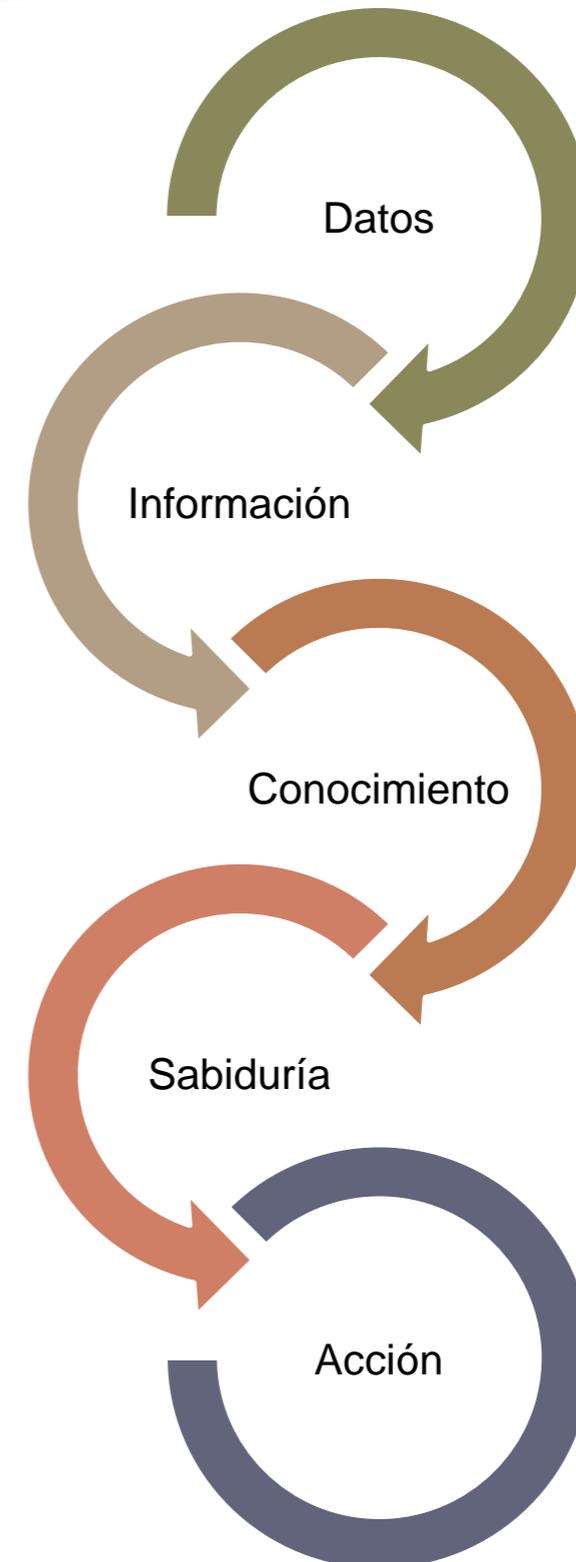
Datos, datos, datos



- Fácil y barato capturar datos
- Fácil y barato almacenar datos
- ¿Fácil y barato procesar datos?

Datos: nuevo petróleo

Todo dirigido por datos



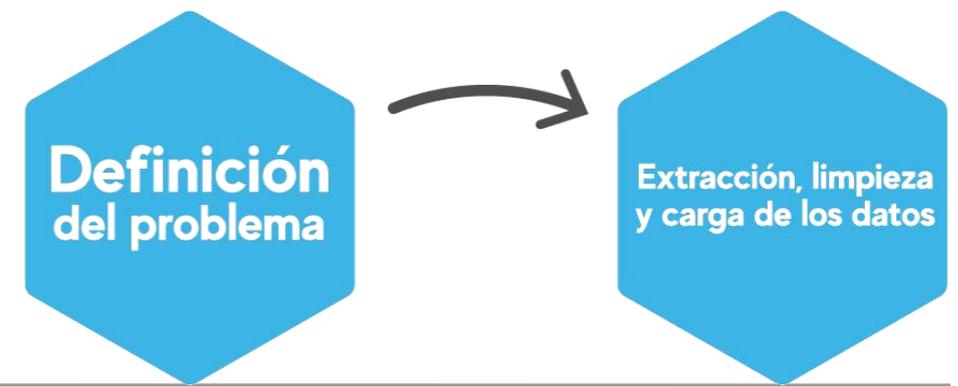
Ciencia de datos



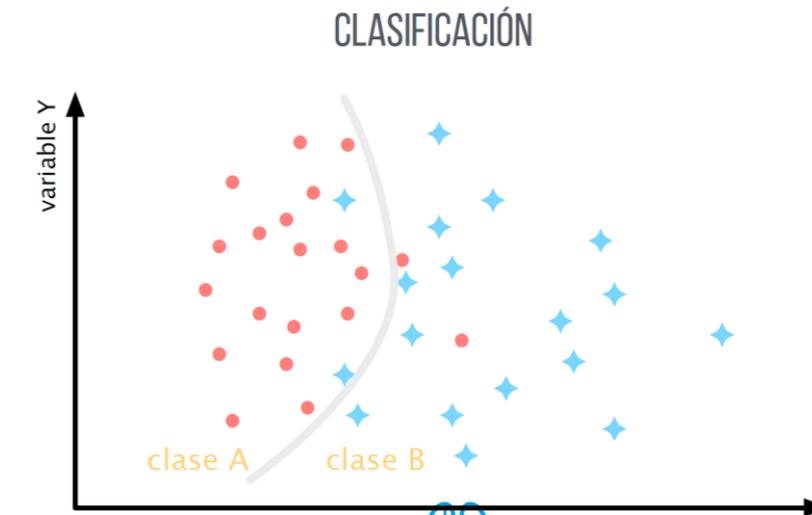
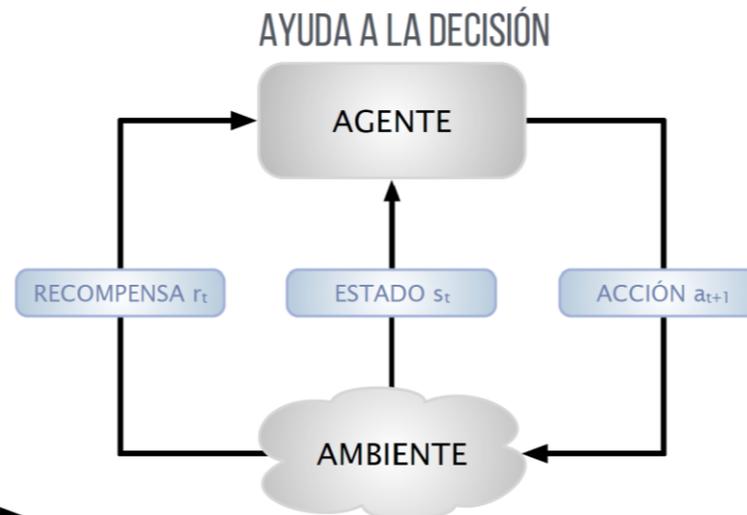
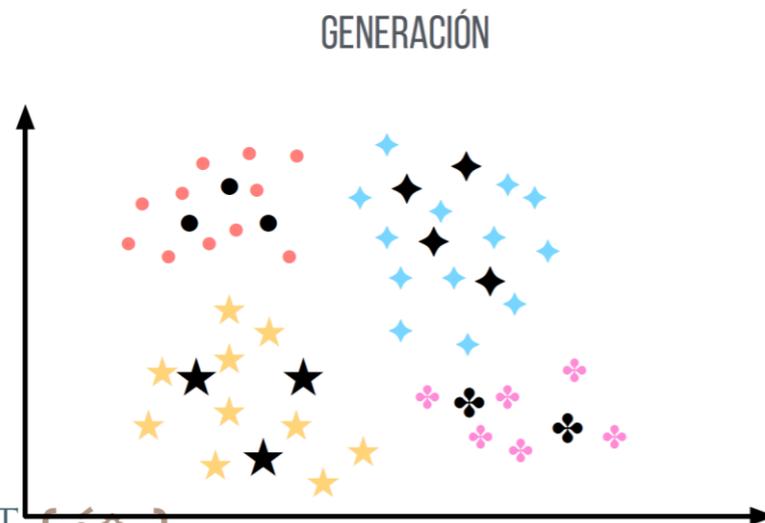
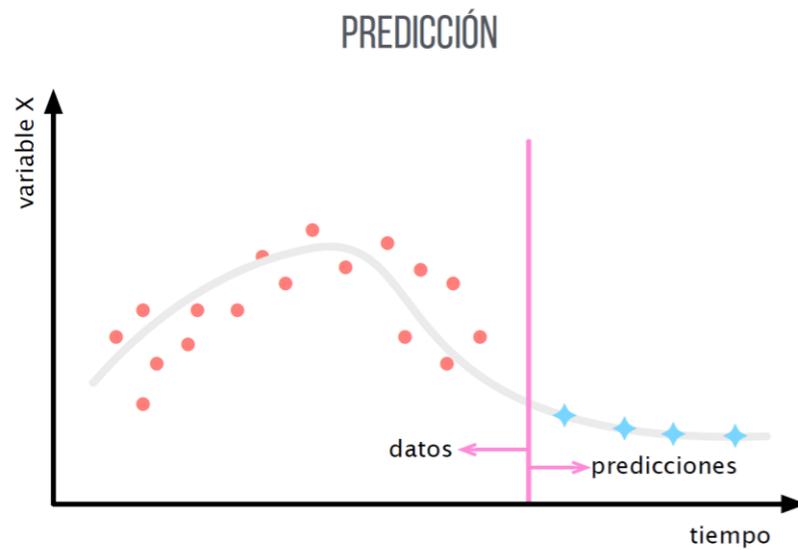
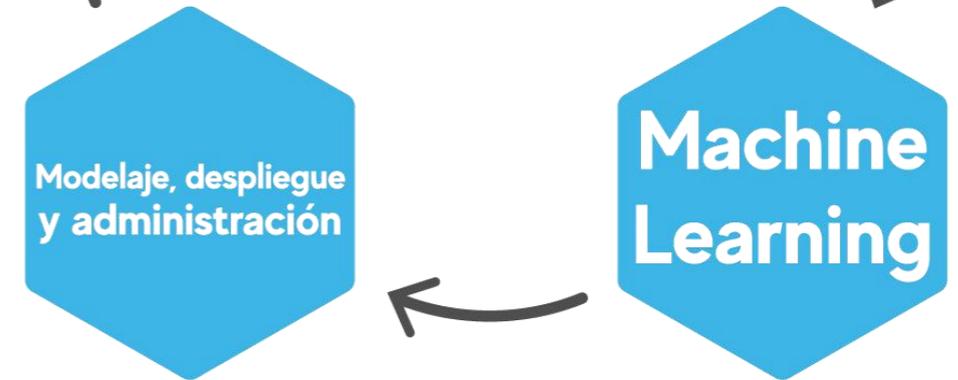
La **ciencia de datos** es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sea estructurados o no estructurados,. La ciencia de datos es, por tanto, una continuación de algunos campos de análisis de datos como la estadística, la minería de datos, el aprendizaje automático y la analítica predictiva.

También se define La ciencia de datos como "Un concepto para unificar estadísticas, análisis de datos, aprendizaje automático y sus métodos relacionados para comprender y analizar los fenómenos reales", empleando técnicas y teorías extraídas de muchos campos dentro del contexto de las matemáticas, la estadística, la ciencia de la información y la informática.

Ciencia de datos



Ciencia de Datos
Ciclo de vida



ERIC NIILER

SCIENCE 01.25.2020 07:00 AM

An AI Epidemiologist Sent the First Warnings of the Wuhan Virus

The BlueDot algorithm scours news reports and airline ticketing data to predict the spread of diseases like those linked to the flu outbreak in China.



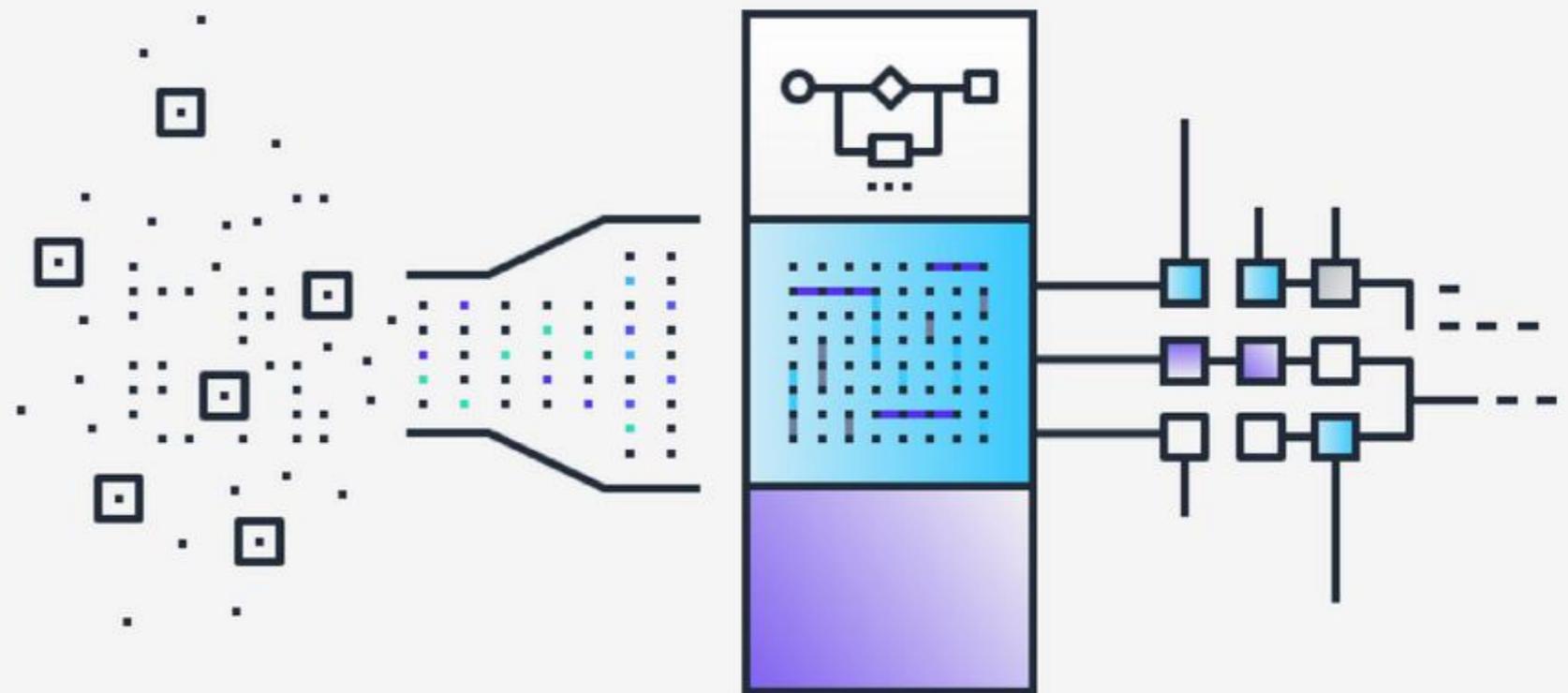
Forbes / Tech

The Little Black Book of Billionaire Secrets

FEB 16, 2012 @ 11:02 AM 2,983,732 VIEWS

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Are Generative Models Good Enough?





Jukebox

We're introducing Jukebox, a neural net that generates music, including rudimentary singing, as raw audio in a variety of genres and artist styles. We're releasing the model weights and code, along with a tool to explore the generated samples.

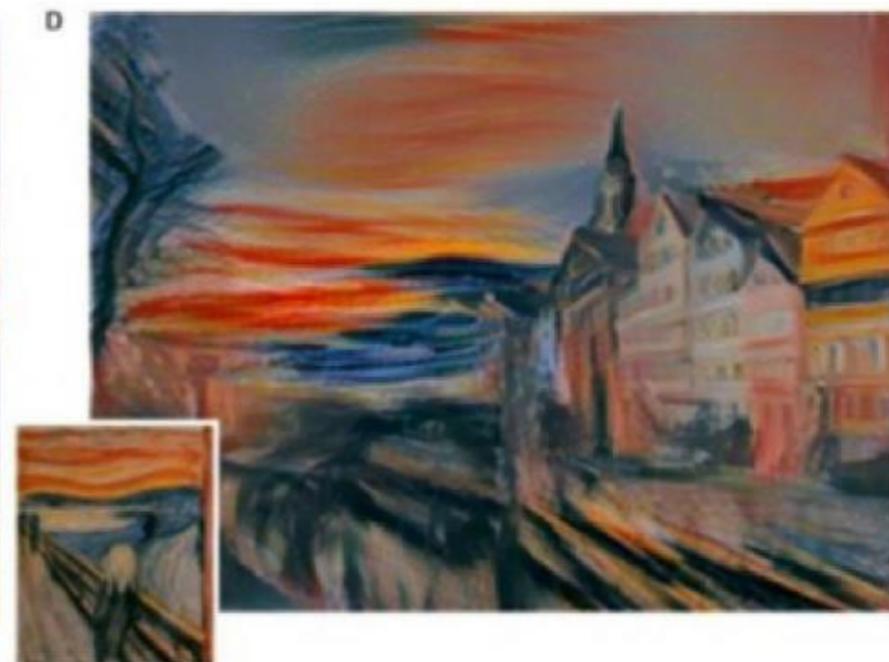
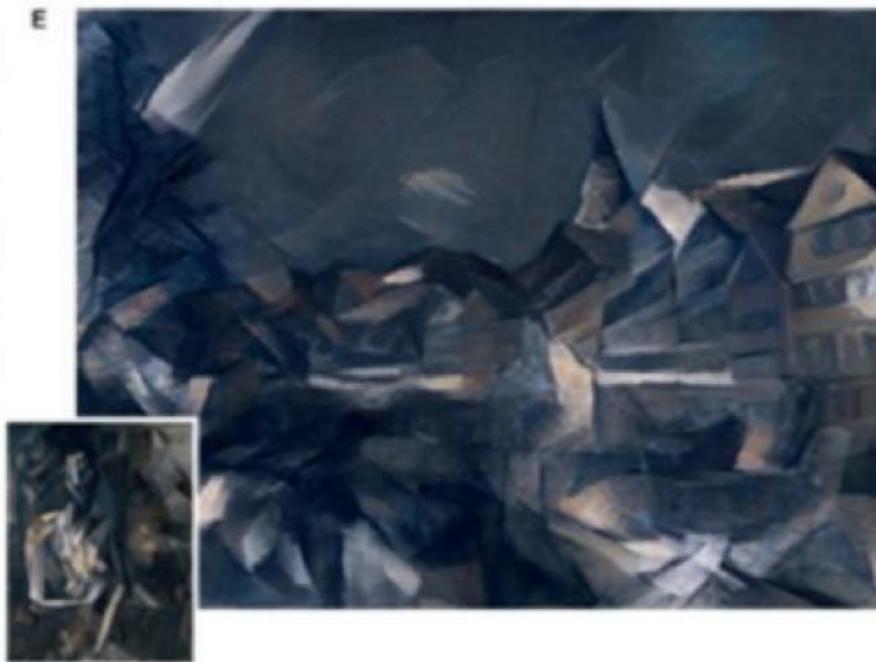
[📄 READ PAPER](#)

[🔗 VIEW CODE](#)

APRIL 30, 2020
12 MINUTE READ, 10 DAY LISTEN



Figure 7: Generated samples

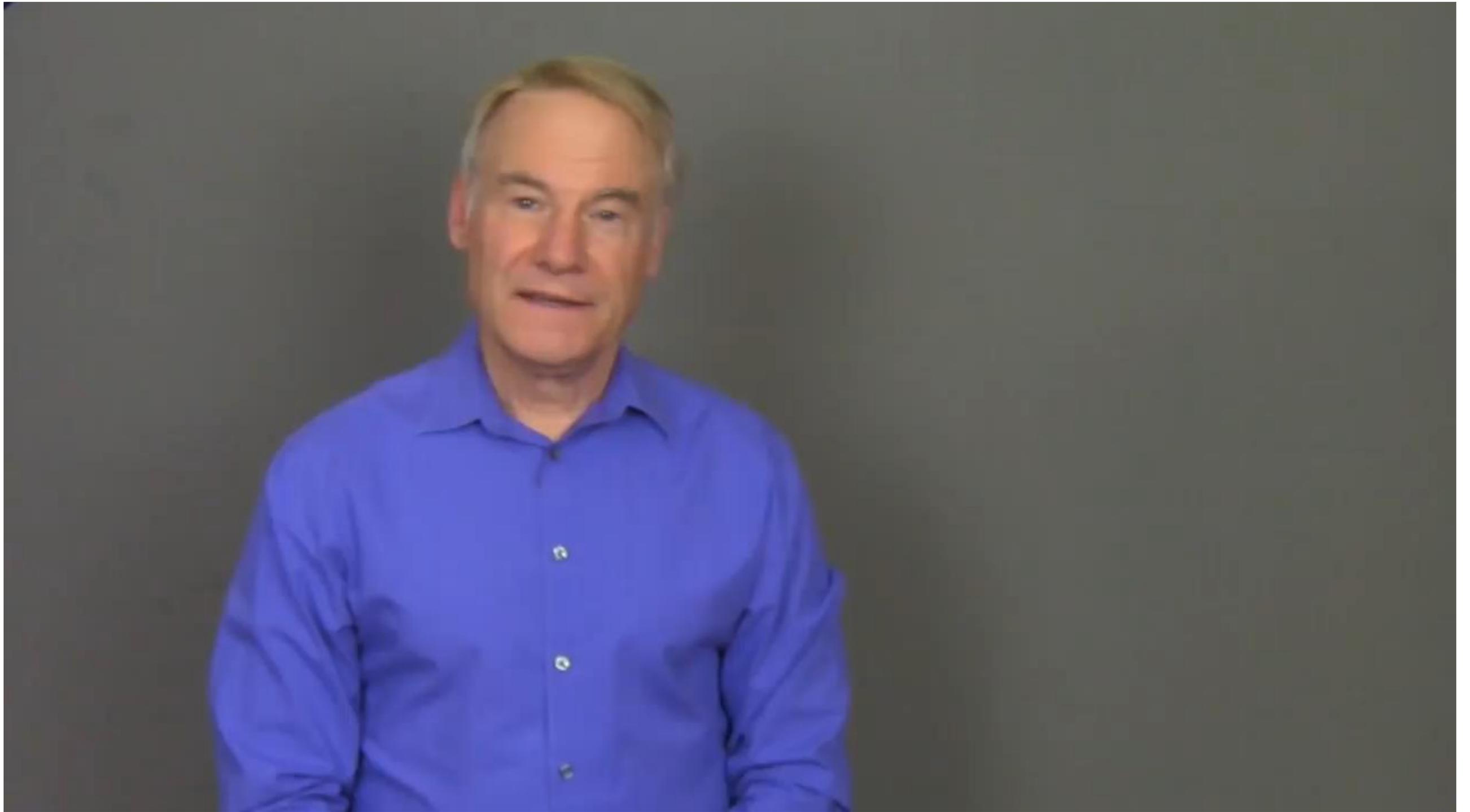




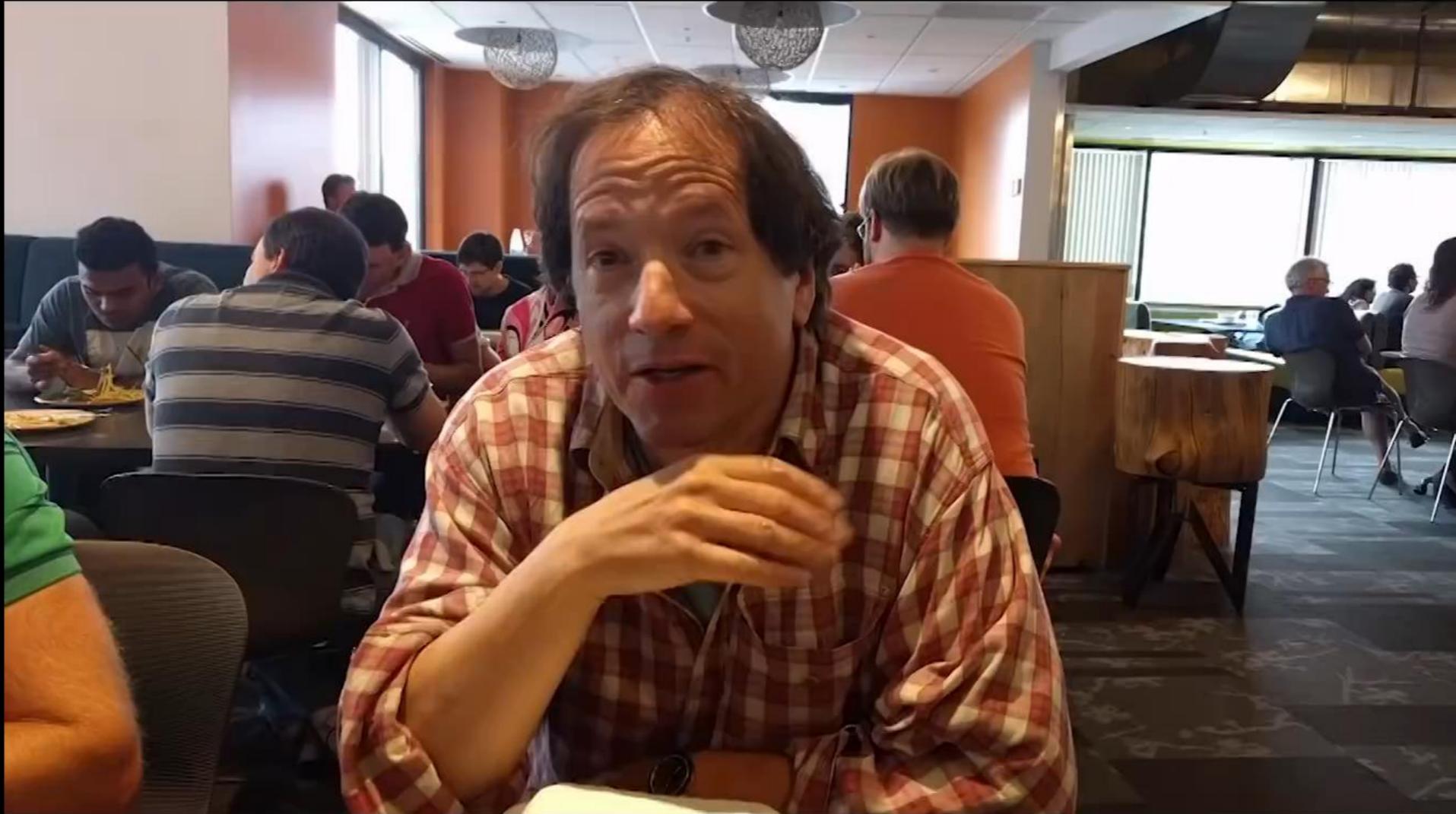
Artificial intelligence / Machine learning

OpenAI's new language generator GPT-3 is shockingly good—and completely mindless

The AI is the largest language model ever created and can generate amazing human-like text on demand but won't bring us closer to true intelligence.



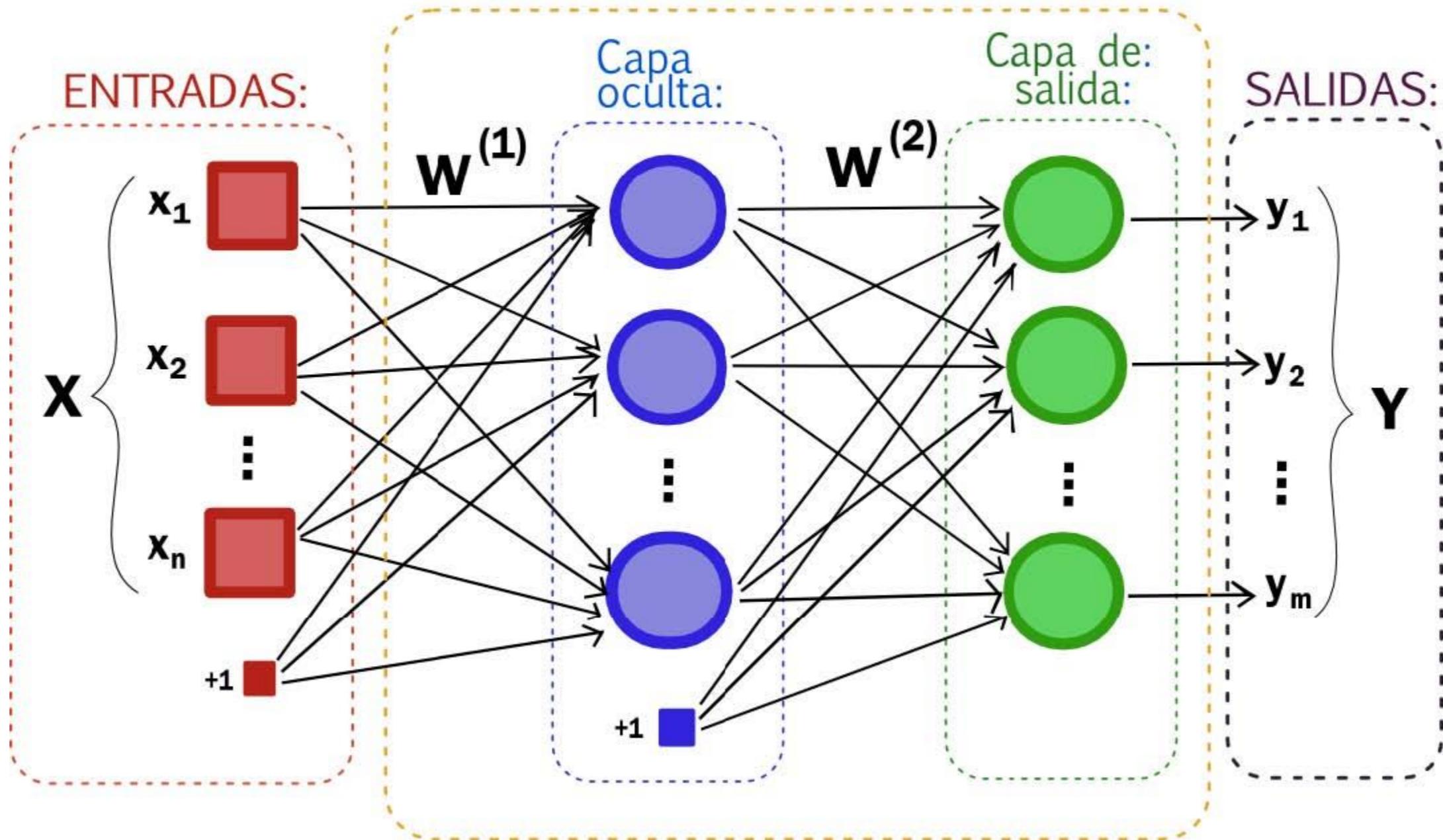
Input video



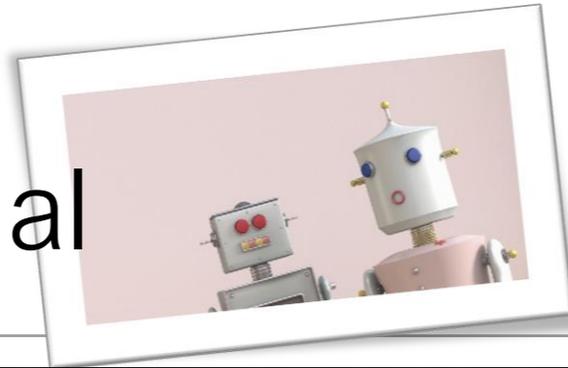
Red neuronal



RED NEURONAL:



Inteligencia artificial



A brief history of AI

ARTIFICIAL INTELLIGENCE

Intelligent algorithms defined and coded by people into machines



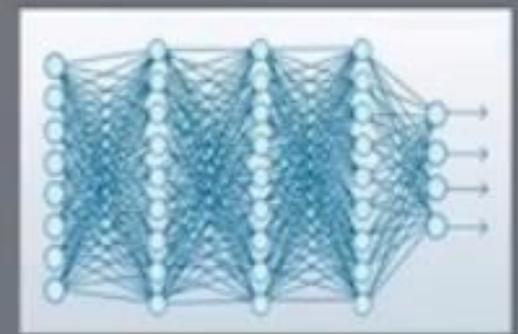
MACHINE LEARNING

Ability to learn without being explicitly programmed



DEEP LEARNING

Learning based on Deep Neural Networks



1950's

1960's

1970's

1980's

1990's

2000's

2006's

2010's

2012's

2017's

Visualización

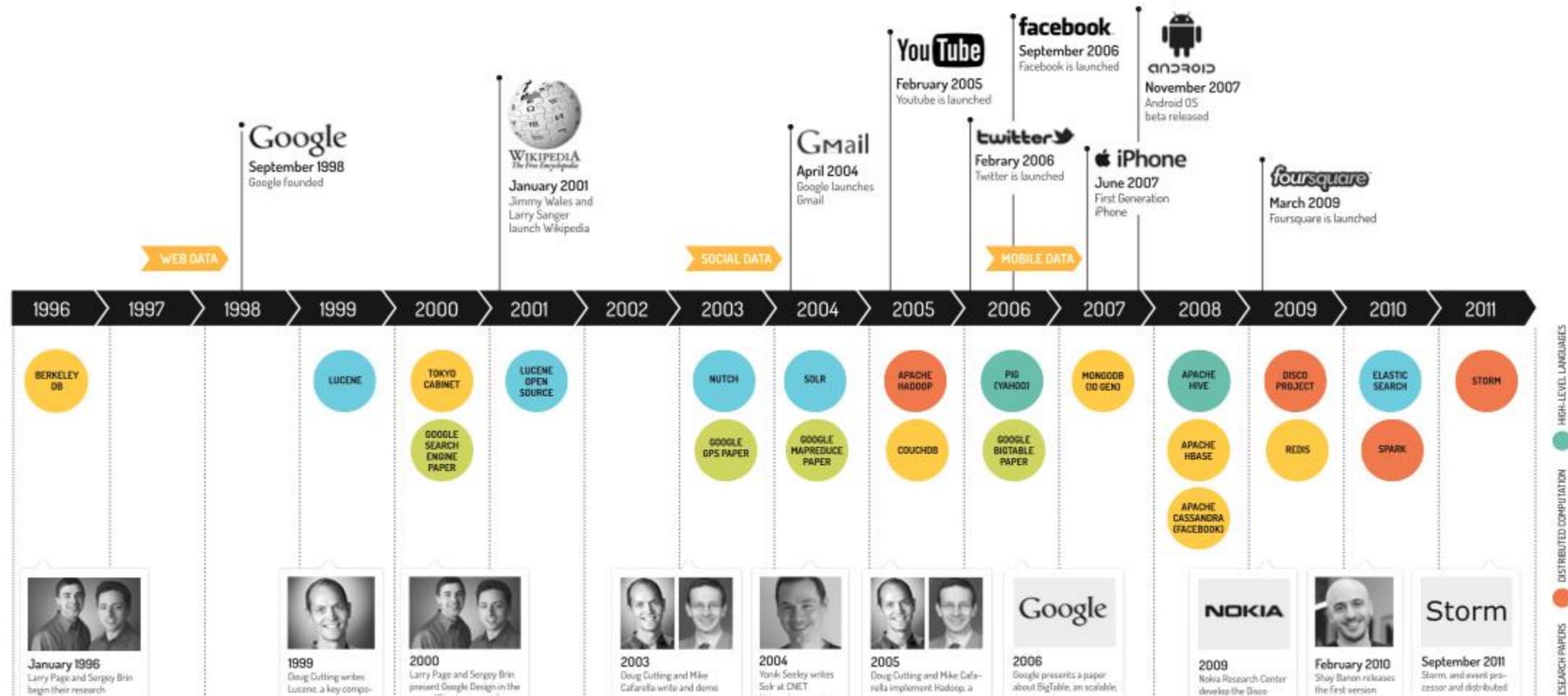


Big Data



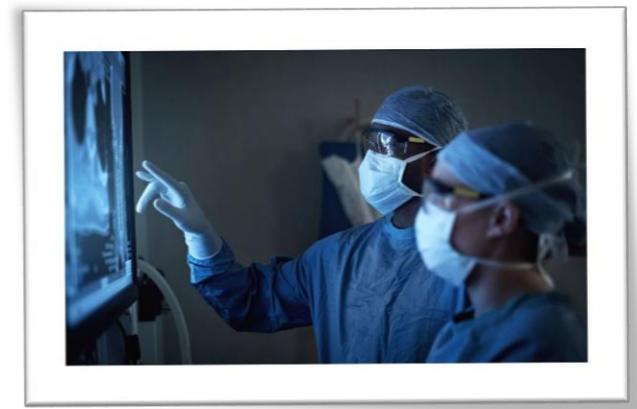
BIG DATA A BRIEF HISTORY

Conjuntos de datos tan grandes, complejos o rápidos que no pueden ser tratados de las maneras habituales





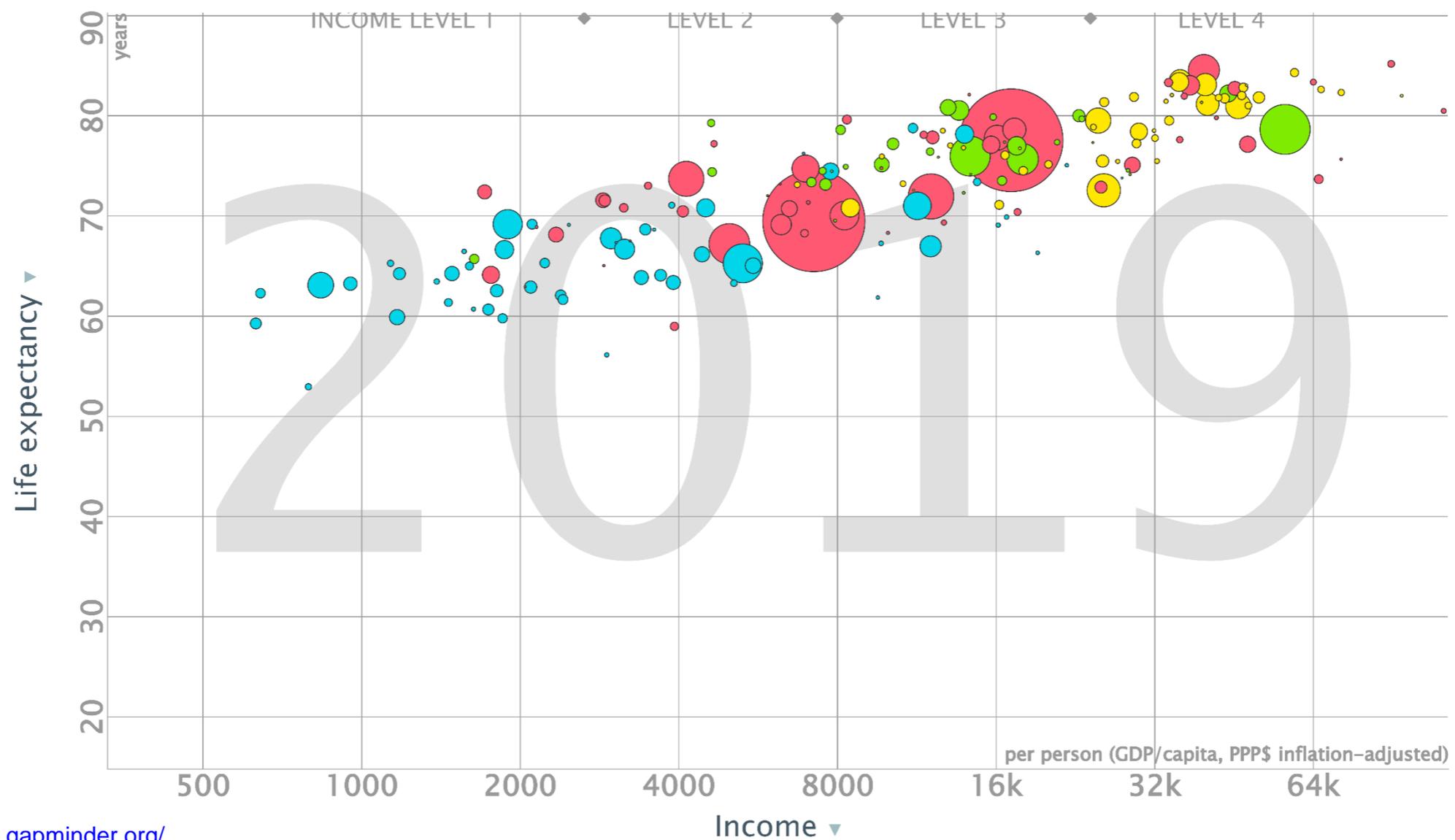
¿Cómo me afecta esto a mí?



Ética, sociedad y globalización



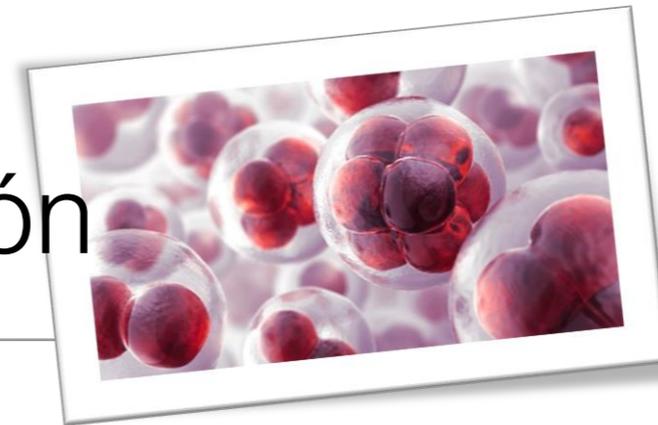
La globalización es un fenómeno imparable, y no tiene por qué ser pernicioso



<https://www.gapminder.org/>

Vivimos en un mundo mejor

Ética, sociedad y globalización





Informe Empleos Emergentes 2020

#1 76% crecimiento anual

Especialista en Inteligencia Artificial

Lo que deberías saber:

El número de profesionales que se presentan como Especialistas en Inteligencia Artificial ha crecido un 75,97% con respecto al año pasado. Este perfil emergente caracterizado por habilidades específicas como el conocimiento de Machine Learning o Data Science se está incorporando de manera rápida a sectores tan consolidados como el de las Tecnologías de la información o la Programación informática que ven en estos perfiles no solo un potencial para mejoras operativas sino para el desarrollo de nuevos productos, servicios y mercados.

Habilidades requeridas para el trabajo:

Machine Learning, Python, Inteligencia Artificial, Visión por ordenador, Data Science, R, Git, Deep Learning, C++ y Java.

Dónde están los empleos:

Barcelona, Madrid y Valencia.

Sectores que contratan estos perfiles :

Servicios y tecnologías de la información, Programación informática, Investigación, Educación, Internet, Telecomunicaciones y Consultoría de gestión.

Las tecnologías disruptivas dirigirán la innovación



Business is business



Las reglas han cambiado, nos guste o no, y no se va a volver atrás

La información era negocio. Ahora los datos también son negocio

- Nuevos modelos de negocio
- Competitividad
- Ahorro de costes
- Necesidad de modernización ante unos clientes que lo demandan
- Quizá no haga falta “ciencia de datos”, pero indudablemente hace falta “digitalización”

Business is business



Las tecnologías disruptivas impulsarán la innovación en todo tipo de industrias y servicios

Comunicaciones

Medios

Gestión de publicidad
Venta inteligente de publicidad

Energía y servicios

Smart Grid
Redes de distribución de servicios

Distribución

Fabricación

Planificación y temporización de la producción
Optimización de la producción

Minorista

Gestión de inventario, distribución y procesos

Gestión de almacenamiento

Gestión de RRHH

Transporte y viajes

Gestión de contenedores

Gestión de recursos turísticos

Financieras

Banca

Gestión del dinero
Optimización de campañas

Optimización de la carta de servicios

Servicios financieros

Gestión de bolsa y mercados

Industria

Automoción

Planificación de ventas y operaciones
Gestión de producción

Aeroespacial y defensa

Planificación de la producción

Gestión de proyectos
Gestión de recursos

Electrónica

Procesado en tiempo real

Química y petróleo

Gestión de redes y procesos

Administración

Salud

Gestión de plantilla
Planificación de producción

Gestión de la red de distribución e inventarios

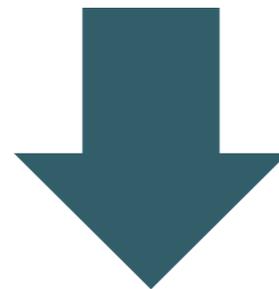
Administración

Optimización de la red de puntos de atención
Gestión del transporte

Renovarse o ¿morir?



- La sociedad demanda herramientas digitales
- Nuevos clientes/usuarios con nuevas necesidades. Y nuevos productos (incluso no tangibles)
- Los proveedores van a implantar nuevos métodos de trabajo y ciencia de datos (también la administración)



- Nuevos modelos de negocio y obsolescencia de otros modelos

Renovarse o ¿morir?



- La información sigue siendo un producto en sí mismo y un valor añadido para las empresas
 - Importancia de los datos. Lo que no se mide no existe
 - Imprescindible el uso de las TIC (Tecnologías de la Información y Comunicaciones)
 - Evaluación
-
- Ciencia de Datos e IA: nuevas herramientas competitivas

Renovarse o ¿morir?



Servicios industriales



Agricultura/Ganadería



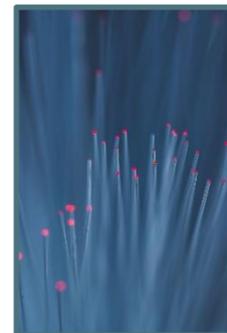
Turismo



Pequeño comercio



Producción industrial



Tecnológicas



Bibliografía divulgativa

- **Miquel Barceló** (2008). *Una historia de la informática*. Volumen 107 de Informática (Editorial UOC). ISBN 978-84-9788-709-0
- **Hans Rosling** (2018). *Factfulness: Diez razones por las que estamos equivocados sobre el mundo. Y por qué las cosas están mejor de lo que piensas*. ISBN 978-84-2342-996-7
- **Nicholas Negroponte** (1997). *El mundo digital*. ISBN 978-84-4065-925-5. (Existe una versión de 2015, pero está sólo en inglés)
- **Juan Manuel López Zafra, Ricardo A. Queralt** (2019). *Alquimia: Cómo los datos se están transformando en oro*. ISBN 978-84-2343-080-2
- **Enrique Dans** (2019). *Viviendo en el futuro: Claves sobre cómo la tecnología está cambiando nuestro mundo*. ISBN 978-84-2343-084-0
- **Bruce Schneier y Álvaro Robledo** (2019). *Haz clic aquí para matarlos a todos: Un manual de supervivencia*. ISBN 978-84-9998-753-8
- **Kai-Fu Lee** (2020). *Superpotencias de la inteligencia artificial: China, Silicon Valley y el nuevo orden mundial*. ISBN 978-84-2343-131-1
- **Herbert Jones** (2019). *Ciencia de los datos: La guía definitiva sobre análisis de datos, minería de datos, almacenamiento de datos, visualización de datos, Big Data para empresas y aprendizaje automático para principiantes*. ISBN 978-10-9080-296-5
- **Lasse Rouhiainen** (2018). *Inteligencia artificial: 101 cosas que debes saber hoy sobre nuestro futuro*. ISBN 978-84-1756-808-5
- **Hannah Fry** (2019). *Hola mundo: Cómo seguir siendo humanos en la era de los algoritmos*. ISBN 978-84-1755-241-1
- **Max Tegmark** (2018). *Vida 3.0*. ISBN 978-84-3061-962-7
- **Kurzweil Ray** (2019) *La singularidad está cerca*. ISBN 978-39-4420-301-0
- **Klaus Schwab** (2016) *La cuarta revolución industrial*. ISBN 978-84-9992-694-0



IDAL

Intelligent Data Analysis Laboratory

Ciencia de datos, Big data...

¡Esto no puede pasarme a mí!

Rafael Magdalena

Departamento de Ingeniería Electrónica

ETSE, Avda Universitat s/n, 46100

rafael.magdalena@uv.es

<http://idal.uv.es>

VNIVERSITAT
ID VALÈNCIA



Escola Tècnica Superior d'Enginyeria

Departament d'Enginyeria Electrònica



Licencia

Esta presentación está disponible bajo la licencia CC-BY [Creative Common Attribution License 4.0](https://creativecommons.org/licenses/by/4.0/)



Lo que significa: ¡Úsalo y difúndelo como quieras!

Se puede

- **Reusar** – Puedes cambiar estas transparencias, e incluir otro material de otros trabajos.
- **Difundir** – Se pueden hacer copias, distribuir, publicar y transmitir el material.
- **Vender** – Se puede incluir este material en productos comerciales o servicios que luego se cobren.

Reconocimiento requerido: “Material docente gratuito de <http://idal.uv.es>”

Las únicas condiciones cumplir son:

- **Marcas** – No incluir las palabras “IDAL”, “ETSE-UV” o “Universitat de València” ni los logotipos de estas instituciones en tu producto o servicio comercial.
- **Atribución** – Debe quedar claro para los demás la licencia de este trabajo e incluir la cita de reconocimiento requerido.