

Aplicaciones de análisis: Computación Cognitiva

Aspectos organizativos

- Profesores
 - María José Moreno
 - Jenaro Gallego
- Sesiones:
 - 7 Octubre: Introducción
 - 8 Octubre (dos sesiones)
- Práctica a entregar el 13 de Noviembre (**P2**)

Agenda Computación Cognitiva

- Definición de Computación Cognitiva. Cuando tiene sentido utilizarla
- Qué problemas nos encontramos en el hardware y cómo los podemos resolver
- IBM Watson
- Directrices éticas para una Inteligencia Artificial fiable

Agenda Computación Cognitiva

- **Definición de Computación Cognitiva. Cuando tiene sentido utilizarla**
- Qué problemas nos encontramos en el hardware y cómo los podemos resolver
- IBM Watson
- Directrices éticas para una Inteligencia Artificial fiable

Computación Cognitiva

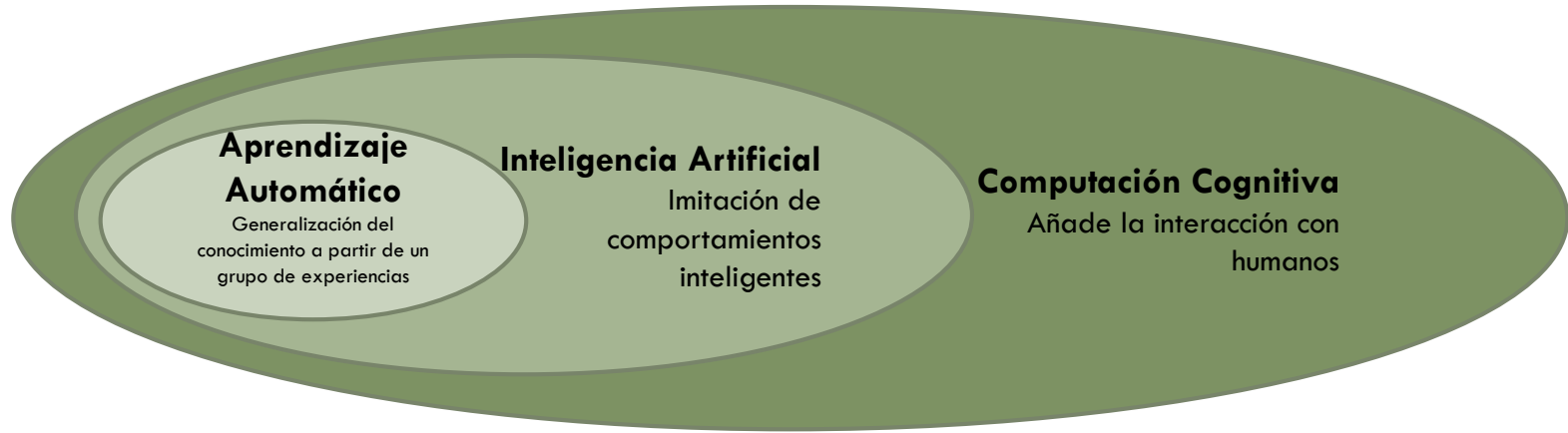
Cognición es la capacidad de recibir una serie de información a través de nuestros sentidos, almacenarla, procesarla y luego aplicarla, lo que nos va a permitir interactuar con el mundo.

Computación cognitiva, diferentes definiciones:

- Cognitive Computing are systems that learn at scale, reason with purpose and interact with humans naturally. It is a mixture of computer science and cognitive science – that is, the understanding of the human brain and how it works.
- Cognitive computing is based on self-learning systems that use machine-learning techniques to perform specific, human-like tasks in an intelligent way.

Computación Cognitiva

- Mezcla de ciencia computacional y neurociencia (entender el cerebro humano y cómo funciona)
- IBM lo describe como: *“Systems that learn at scale, reason with purpose and interact with humans naturally”*



¿Cuándo tiene sentido?

■ Paradigma convencional

- Manejo de números
- Análisis de gran cantidad de datos estructurados
- Totalmente definido por reglas formales



■ El cerebro resuelve de forma muy simple, problemas que son complejos de resolver para un ordenador convencional:

- Percepción
- Toma de decisiones
- Comunicación

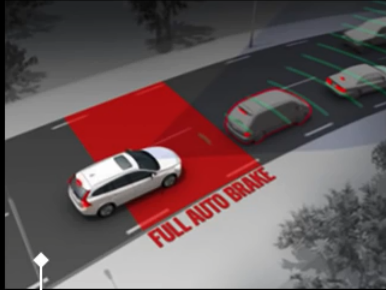


Agenda Computación Cognitiva

- Definición de Computación Cognitiva. Cuando tiene sentido utilizarla
- **Qué problemas nos encontramos en el hardware y cómo los podemos resolver**
- IBM Watson
- Directrices éticas para una Inteligencia Artificial fiable

A qué nos enfrentamos

Sensory data is high velocity, variety, and volume. Its value is maximized with real time reaction



Automotive

- 1GB/s per car fully instrumented
- React in 200 msec for maximum value



Mobile

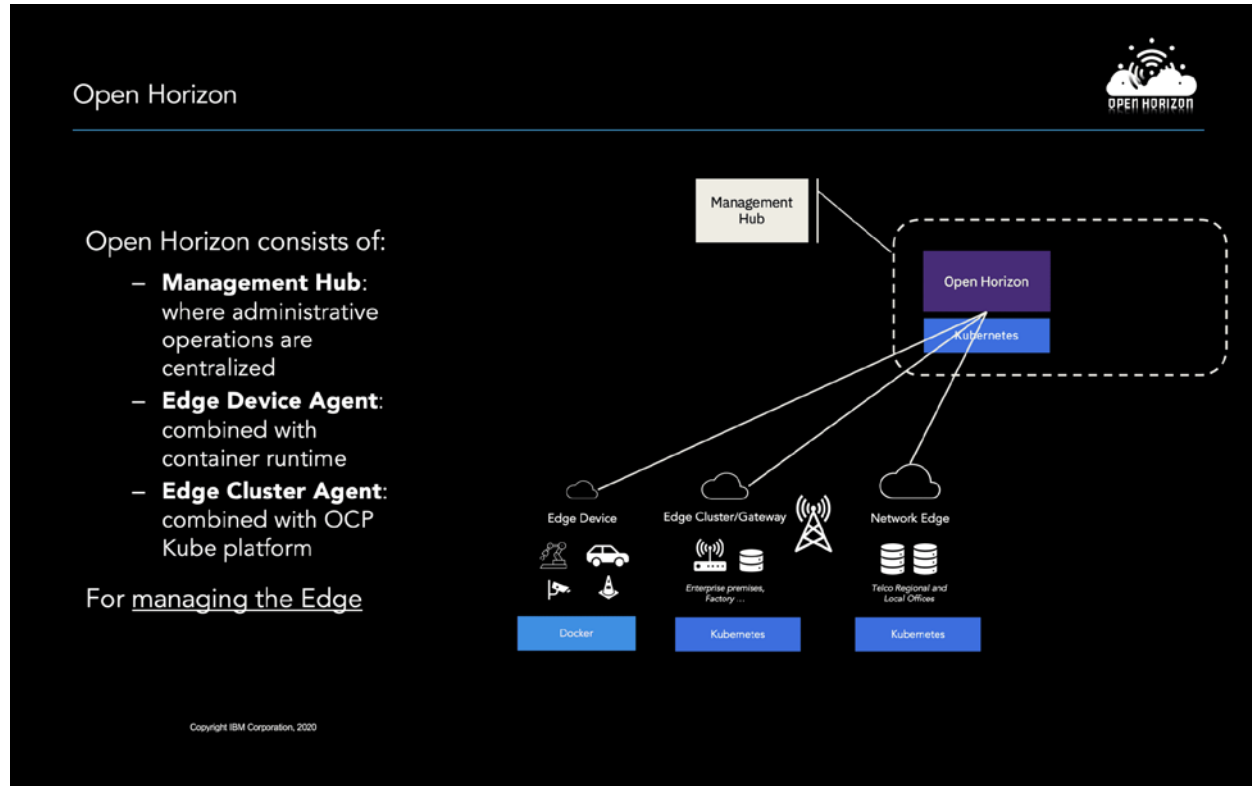
- >10 sensor types
- React in 200 msec for maximum value
- Need to be always on



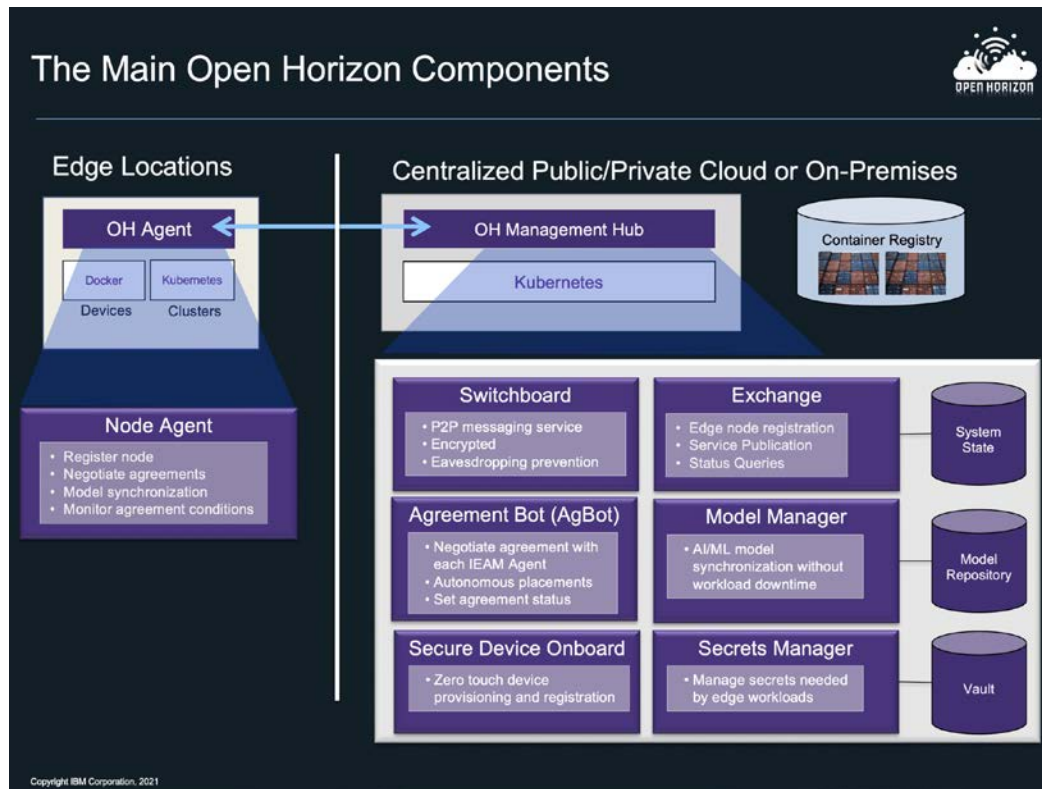
Cameras

- Continuous video streams, always on.
- React 1-5s for maximum value

Linux Foundation: Open Horizon



Componentes de Open Horizon



A qué nos enfrentamos...

Nos enfrentamos a operaciones más complejas que requieren mayor capacidad de procesamiento de datos y mayor complejidad en las operaciones. El objetivo es reducir el tiempo de proceso y/o el consumo de energía

Vídeo visión de IBM:

<https://youtu.be/uhgY3MeTdLo>

<https://www.research.ibm.com/artificial-intelligence/hardware/>

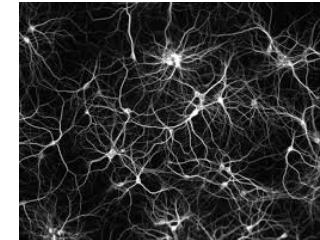
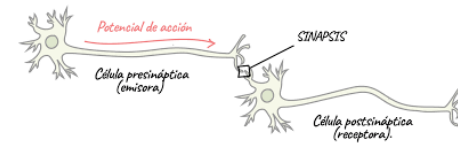
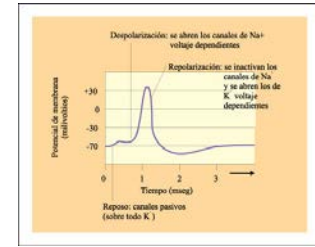
¿Cómo podemos resolverlos?

Utilizando las diferentes soluciones que mejor se ajusten al problema:

- Involucrando procesadores especializados
- Utilizando computación neuromórfica: reproduciendo la estructura y arquitectura del cerebro
- Computación cuántica

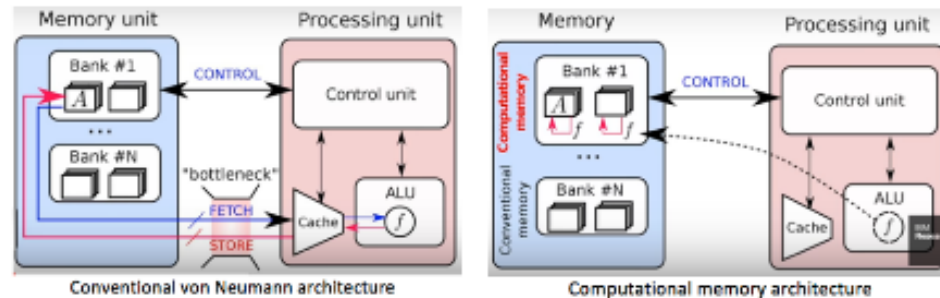
Procesadores neuromórficos:

- Transmisión de señales en las neuronas y potenciales de acción
- Características que nos interesan:
 - Las señales (tanto excitadoras como inhibitoras) recibidas en las dendritas se **integran** y se genera (en su caso) un **potencial de acción** que se propaga a través del axón
 - Características de los potenciales de acción:
 - Son impulsos nerviosos rápidos, con la misma amplitud (**100mV**) y duración aproximada de 1 ms.
 - Velocidad de transmisión no muy alta
 - Su amplitud a lo largo del axón permanece constante
 - La información transmitida por un potencial de acción no está determinada por la forma de la señal sino por la vía que recorre el cerebro.
 - Es una señal todo o nada: los estímulos subumbrales no producen señales y todos los estímulos que superen el umbral, aunque varíen en intensidad o duración, producen la misma señal en cuanto a amplitud y duración se refiere.
 - La característica que transmite información es la frecuencia de los potenciales de acción



De von Neumann a nuevas arquitecturas

- La memoria computacional ("In-memory computing" o "computational memory") es una nueva tecnología que pretende usar las propiedades físicas de la memoria tanto para almacenar como para procesar datos.
- Supone un gran cambio respecto a la arquitectura Von Neumann (utilizada en la mayoría de dispositivos) en la que el trasiego de datos entre la memoria y el procesador ralentiza el proceso y hace que aumente el consumo de energía.



Física involucrada (I)

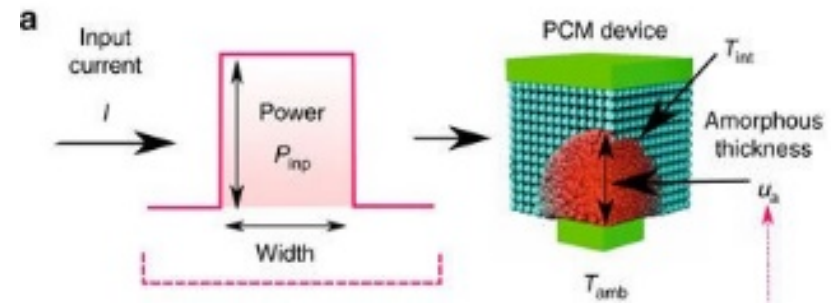
La Memoria de Cambio de Fase (Phase Change Memory o PCM) se basa en las propiedades de ciertos materiales en los que dos estados metaestables o fases de los mismos se usan para almacenar datos. En la actualidad se usa en los dispositivos de almacenamiento óptico (Blue-ray) en los que el cambio de fase se induce calentando el material mediante un haz láser.

Los materiales usados en las PCM presentan dos estados estables, uno amorfo (sin una estructura clara definida) y otro cristalino (con estructura) con conductividades diferentes (baja y alta respectivamente).

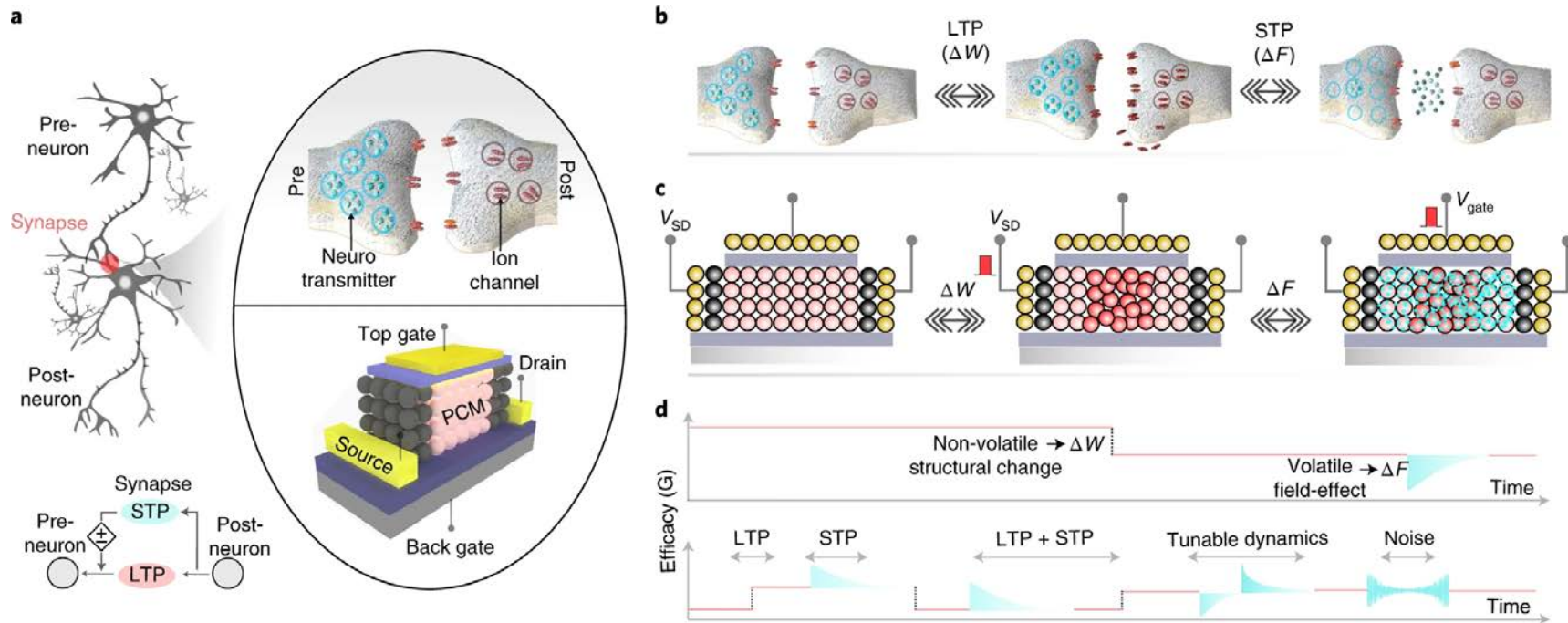
Física involucrada (II)

Con la memoria de cambio de fase (PCM), cuando se aplica un pulso eléctrico al material, cambia la conductividad del dispositivo debido al cambio entre las fases cristalina y amorfa. Un pulso eléctrico bajo hará que el dispositivo pase a la fase cristalina (con menor resistencia, mayor conductividad) y un pulso eléctrico alto hará que el dispositivo esté en fase amorfa (mayor resistencia, menor conductividad).

Por lo tanto, en lugar de situarse en dos posibles estados (0 ó 1), puede situarse en un continuo de estados (pasamos así a las posibilidades analógicas).

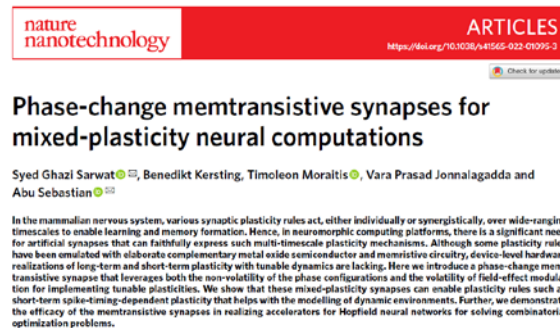


Intentando replicar la neurona



Nuevas sinapsis tecnológicas

- En un reciente artículo IBM detalla el uso de una sinapsis que denomina “*memotransistiva*”, que combina las características de la “memoria” y el “transistor”, el “*memristor*”
- El *memristor* es un elemento de memoria electrónica no volátil. Tanto los *memristores* como los transistores se encuentran en el mismo dispositivo de bajo consumo



Computación cuántica

Nuevo paradigma en computación

Vídeos de Darío Gil (Director IBM Research):

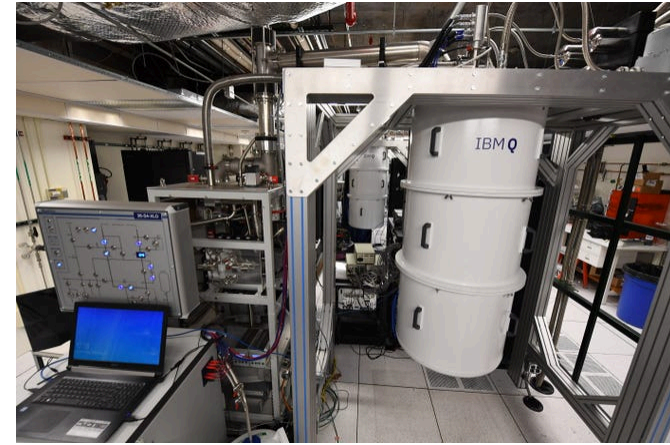
Mapping the future of technology:

- <https://youtu.be/ky32162U06M>

Quantum Summit 2021

- <https://youtu.be/kR0y5Ps-7Kw>

“Convergencia de bits, neuronas y qubits orquestados por una cloud híbrida”





Overview

Learn

Community

Documentation

Start building with Qiskit Runtime. Leverage the new programming model and execution framework to efficiently execute circuits.

[Learn more](#)



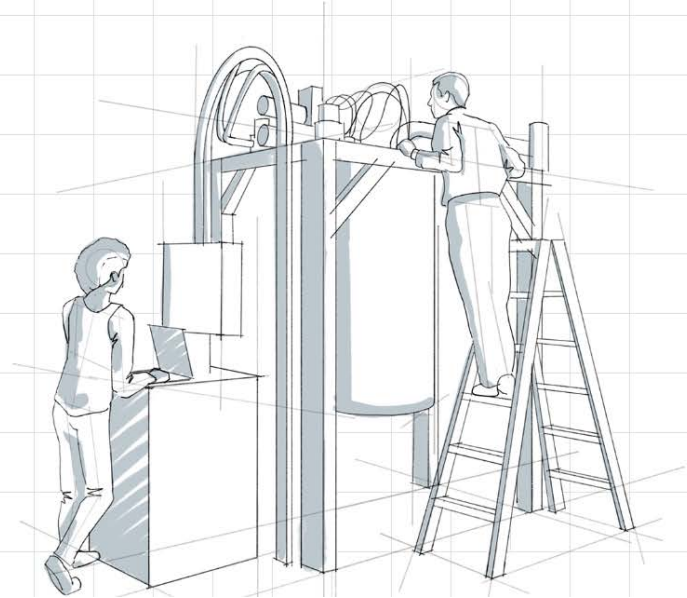
qiskit 0.38.0

[see release notes](#)

Open-Source Quantum Development

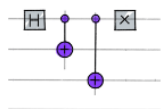
Qiskit [kiss-kit] is an open-source SDK for working with quantum computers at the level of pulses, circuits, and application modules.

Get started



Distintas opciones de programación

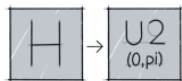
| Get started | Core | Providers | Applications | Extensions | Contributing | Help and Learning |
|--------------------------|------------------------------|---|---|---|---|---|
| Overview | Qiskit Terra | Qiskit Aer Qiskit IBM Runtime Qiskit IBM Provider Partners | Qiskit Nature Qiskit Finance Qiskit Optimization Qiskit Machine learning | Qiskit Experiments Qiskit Dynamics Qiskit Metal | Contributor Guide Deprecation Policy Maintainer Guide | Stack Exchange Qiskit Feedback Slack Qiskit Textbook |



Circuit Library

Qiskit includes a comprehensive set of quantum gates and a variety of pre-built circuits so users at all levels can use Qiskit for research and application development.

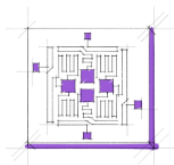
[Learn more](#)



Transpiler

The transpiler translates Qiskit code into an optimized circuit using a backend's native gate set, allowing users to program for any quantum processor or processor architecture with minimal inputs.

[Learn more](#)



Run on real hardware

Users can run and schedule jobs on real quantum processors, and employ Qiskit Runtime to orchestrate quantum programs on cloud-based CPUs, QPUs, and GPUs.

[Learn more](#)



Agenda Computación Cognitiva

- Definición de Computación Cognitiva. Cuando tiene sentido utilizarla
- Qué problemas nos encontramos en el hardware y cómo los podemos resolver
- **IBM Watson**
- Directrices éticas para una Inteligencia Artificial fiable

Leonardo Torres Quevedo

En 1914, publica “Ensayos sobre Automática.—Su definición. Extensión teórica de sus aplicaciones”.

“Además, se necesita -y éste es el principal objeto de la Automática—que los autómatas **tengan discernimiento**, que puedan en cada momento, teniendo en cuenta las **impresiones que reciben**, y también, a veces, **las que han recibido anteriormente, ordenar la operación deseada**.

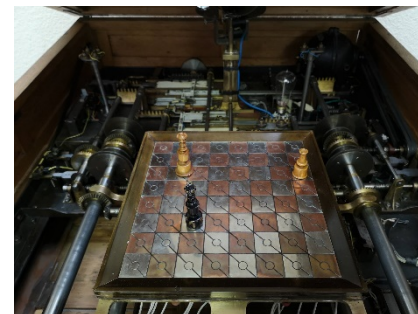
Es necesario que los autómatas **imiten a los seres vivos, ejecutando sus actos con arreglo a las impresiones que reciban y adaptando su conducta a las circunstancias**”

Leonardo Torres Quevedo

Primer ajedrecista: 1912



Segundo ajedrecista: 1920



Haciendo historia

Este tipo de problemas nos han llamado la atención desde el principio de la computación:

- IBM 701(1952):
 - En 1952 IBM puso a Gene Amdahl a trabajar en una red neuronal en la que se “simulaban” mil neuronas. Trabajó en el área de reconocimiento de caracteres
 - En 1954, programa traductor del ruso al inglés
 - En 1956, programa para jugar al ajedrez
- En 1962 se crea una voz sintética que canta “Daisy Bell” en un IBM 7094

<https://www.youtube.com/watch?v=41U78QP8nBk>



Distintos retos en el tiempo

IBM se plantea retos en la investigación

➤ 1997



➤ 2011



➤ 2019



Deep Blue (1997)

- Sistema híbrido: procesadores de propósito general y procesadores específicos para jugadas de ajedrez (mejora de 1996 a 1997: 100 millones de posiciones a 200 millones por segundo)
- Los maestros de ajedrez colaboraron en mostrar jugadas de aperturas adecuadas y como sparrings

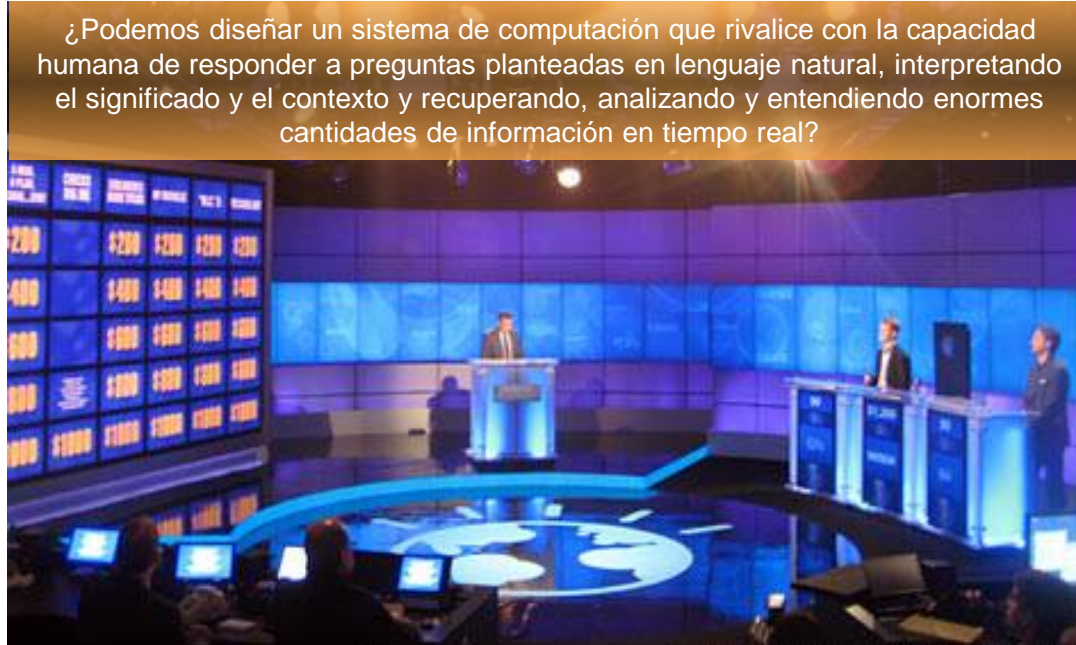


[Artículo sobre 20 años de Deep Blue](#)

Jeopardy (2011)

➤ ¿Cómo funciona el juego? Apuestas y tipos de preguntas

¿Podemos diseñar un sistema de computación que rivalice con la capacidad humana de responder a preguntas planteadas en lenguaje natural, interpretando el significado y el contexto y recuperando, analizando y entendiendo enormes cantidades de información en tiempo real?



Jeopardy en contexto

- Primer programa emitido en 1964

PREMIOS MILLONARIOS EN TELEVISIÓN

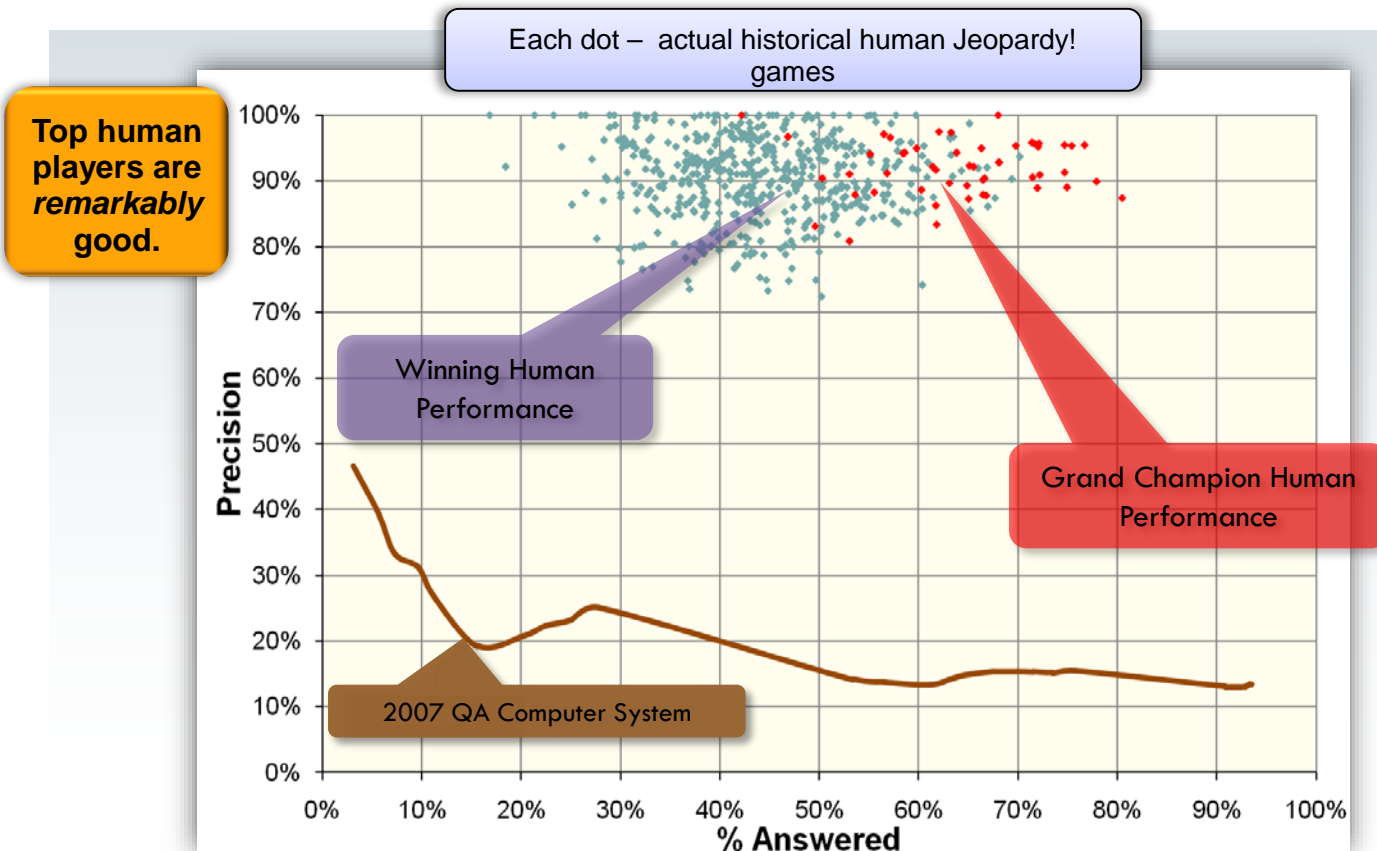
En millones de euros

| Año | Concursante | Programa | País | Premio |
|------|--------------------|--------------------------------|----------|--------|
| 2019 | Los Lobos | <i>iBoom!</i> | España | 6,6 |
| 2001 | Arno Woesthoff | <i>Miljoenenjacht</i> | Holanda | 4,99 |
| 2001 | Brad Rutter | <i>Jeopardy!</i> | EE UU | 4,34 |
| 2004 | Ken Jennings | <i>Jeopardy!</i> | EEUU | 3,52 |
| 2009 | Nino Haase | <i>Schlag den Raab</i> | Alemania | 3 |
| 2016 | Rockampers | <i>iBoom!</i> | España | 2,33 |
| 2013 | Andrew Kravis | <i>The Million Second Quiz</i> | EE UU | 2,31 |
| 2006 | Eduardo Benito | <i>Pasapalabra</i> | España | 2,2 |
| 2019 | James Holzhauer | <i>Jeopardy!</i> | EE UU | 2,19 |
| 2010 | Hans Martin Wismar | <i>Schlag den Raab</i> | Alemania | 2 |

Fuente: elaboración propia. EL PAÍS

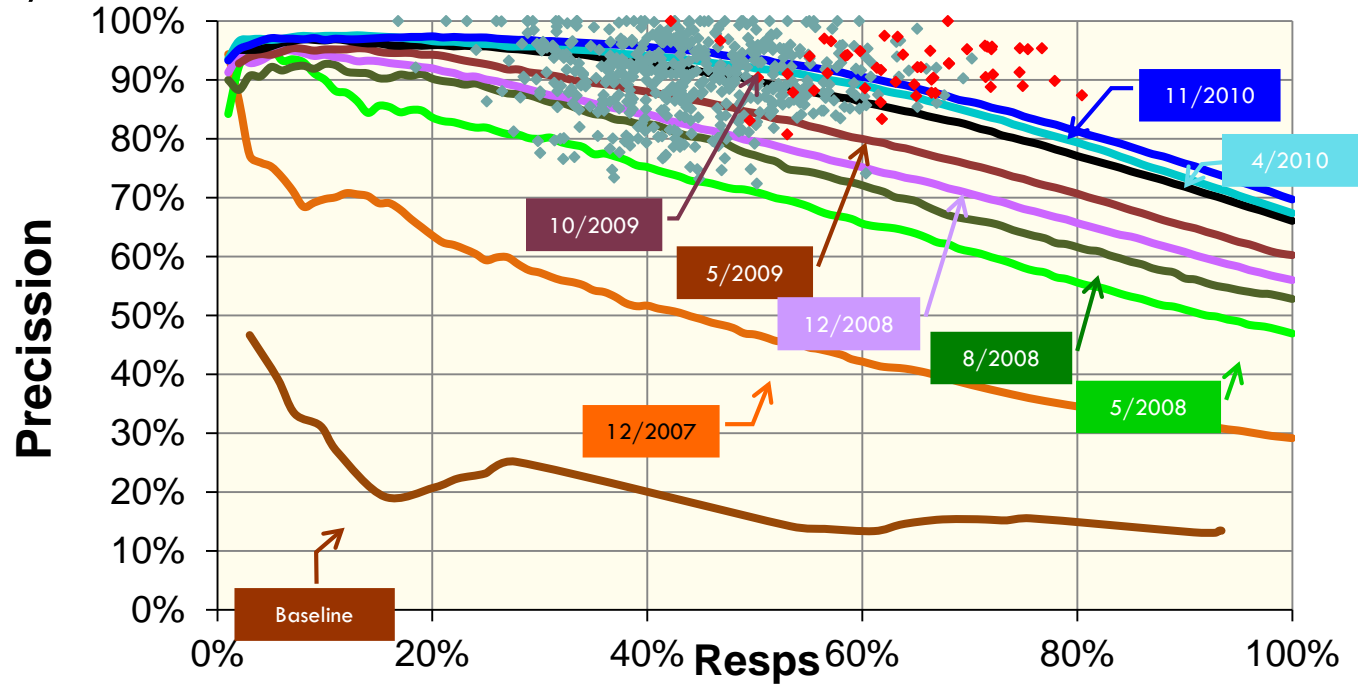
Fuente: https://elpais.com/cultura/2019/07/08/television/1562593572_018549.html

Análisis de los resultados de los ganadores

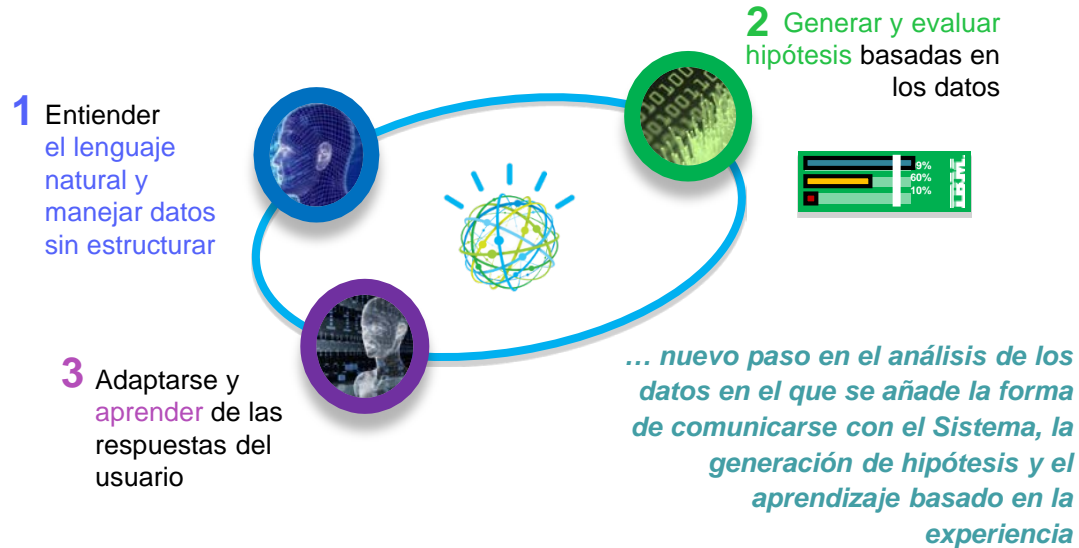


Evolución

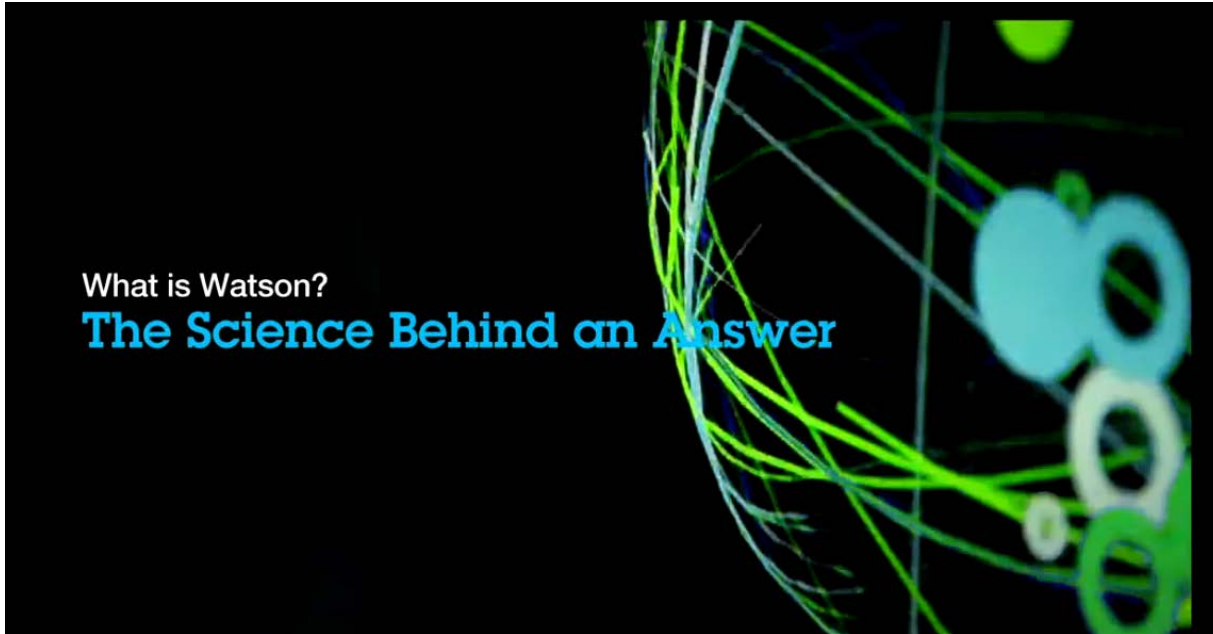
- Progreso en precisión y confianza en la respuesta: 6/2007-11/2010



Qué capacidades cognitivas necesita IBM Watson



Ejemplo



https://youtu.be/heVE_me5VaQ

Ejemplos de utilización

- **IBM Project Debater**
- IBM Personality Insights
- Watson Health - Merative

IBM Debater

En desarrollo desde 2012, es el primer Sistema de IA que puede debatir con humanos sobre temas complejos

Es capaz de:

- Asimilar gran cantidad de textos
- Construir un discurso sobre un tema
- Emitirlo con claridad y con un objetivo determinado
- Rebatir a sus oponentes
- Comprender el lenguaje oral: identificar los conceptos clave dentro de lenguaje hablado del oponente



<https://research.ibm.com/interactive/project-debater/>

Capaz de funcionar sin *prejuicios*

Demostración de capacidades de IBM Debater

El 11 de febrero de 2019, se enfrentó a Harish Natarajan (record mundial en victorias en competición de debates) sobre el tema “Debemos subvencionar las escuelas infantiles” (preescolar).

El debate tuvo lugar en San Francisco con una audiencia de alrededor de 800 personas. Los organizadores fueron [Intelligence Squared](#) y el moderador [John Donovan](#).

Debate:

<https://www.youtube.com/watch?v=m3u-1yttrVw>

La audiencia vota sobre:

- quién gana el debate: Harish Natarajan
- quién mejora el conocimiento previo sobre sobre el tema tratado: IBM Debater

Algunos conceptos

Tema (topic) : una frase corta que define la materia de interés y la opinion sobre ella. Ej: "Debería prohibirse la venta de videojuegos a menores"

Afirmación dependiente del contexto - Context Dependent Claim (CDC)
– una declaración concisa que soporta o refuta el tema. Ej: "Los videojuegos violentos aumentan la violencia entre los jóvenes"

Evidencia dependiente del contexto - Context Dependent Evidence (CDE) – un texto que soporta directamente una CDC en el contexto de un tema determinado. Ej: "El meta-análisis más reciente (que examinó 30 estudios sobre 130,000 sujetos en todo el mundo) concluyó que la exposición a los videojuegos violentos causa agresividad en los jugadores, tanto a corto como a largo plazo"

Componentes de IBM Debater

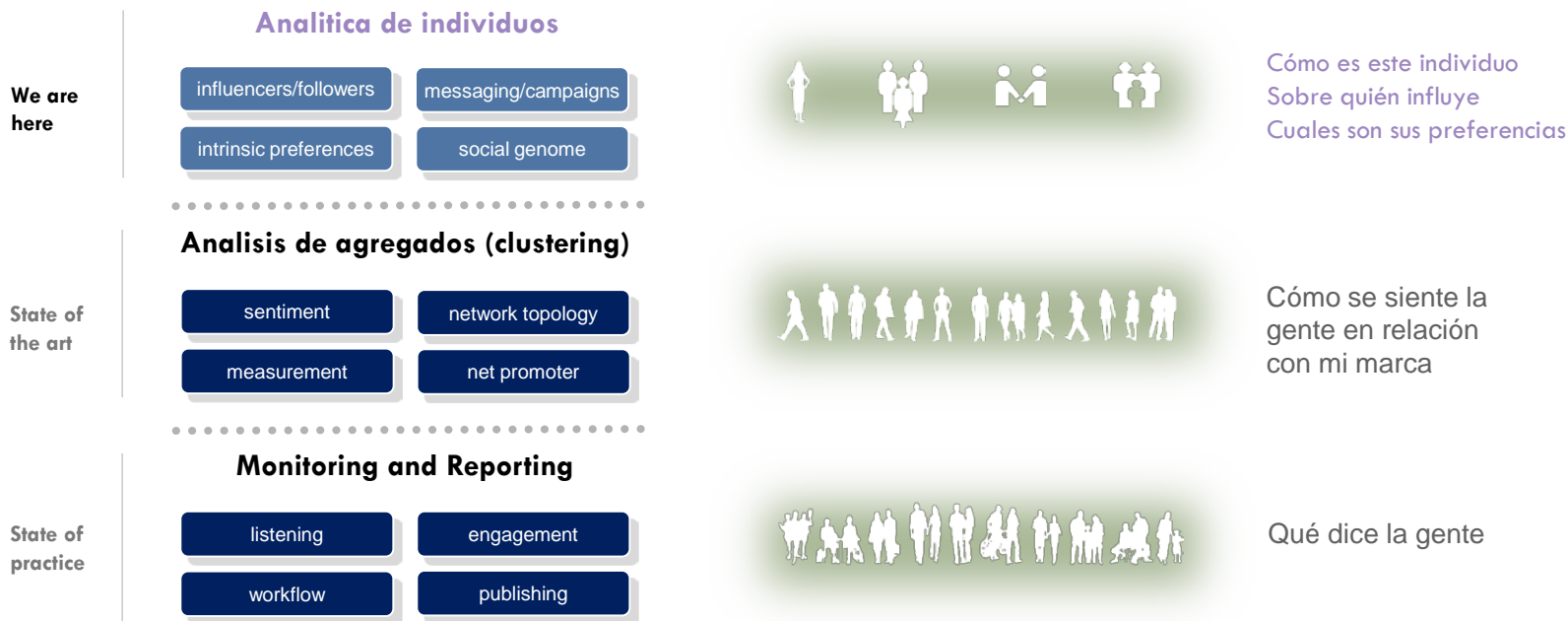
- **Análisis del tema:** identificar los conceptos importantes mencionados en un tema y las opiniones (“sentiments”) hacia cada uno de ellos
- **Recuperación de artículos:** busca en la Wikipedia artículos con una probabilidad alta de contener CDCs (afirmaciones dependientes del contexto).
- **Detección de CDCs:** analiza los artículos para detectar afirmaciones en este contexto
- **Pro/con de las afirmaciones:** juzga la polaridad del CDC en relación con el tema en cuestión
- **Equivalencia de afirmaciones:** trata de evitar la redundancia determinando qué CDCs son semanticamente equivalentes
- **Refinamiento de afirmaciones:** mejora la precisión de la salida eliminando los CDCs con poca confianza
- **Texto a habla:** presenta en modo hablado las afirmaciones que son mejores candidatas dentro de un discurso coherente.

Ejemplos de utilización

- IBM Project Debater
- **IBM Personality Insights**
- Watson Health - Merative

Personalización: servir a las personas de forma individual

- Utilización de las opiniones y comentarios públicos
- La tecnología actual puede soportar, en tiempo real y a gran escala, la posibilidad de entender a un individuo mediante el retrato de su personalidad

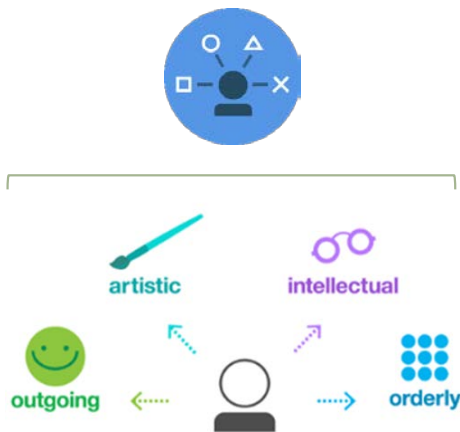


Watson Personality Insights

Nuestra forma de escribir depende de nuestra personalidad

Personality Insights (PI) es un servicio que permite entender a los individuos con los que la empresa quiere comunicarse, permitiendo dirigirse a aquellos de forma única. Personality Insights obtiene las características de personalidad a través del lenguaje utilizado (escrito o hablado)

- El servicio utiliza el **análisis lingüístico** para inferir rasgos de personalidad (a través de correos, forums, tweets, voz, ...)
- El análisis está basado en la psicología del lenguaje y proporciona como resultado datos sobre la personalidad del individuo en tres dimensiones: **Big 5, Valores y Necesidades**
- Los atributos del espectro de personalidad analizado ayudan a **descubrir formas de interacción y productos que se ajustan más a dicho sujeto.**



Infiriendo características de personalidad

El servicio puede derivar del texto métricas relacionadas con tres tipos de características de personalidad



Big Five

5 Dimensiones

Describen cómo una persona se enfrenta al mundo

- Amabilidad
- Responsabilidad
- Extraversión
- Neuroticismo/inestabilidad emocional
- Apertura a nuevas experiencias



Needs

12 Características

Describen qué aspectos de un producto se espera que le llamen la atención

- Excitement
- Harmony
- Curiosity
- Closeness
- Self-expression
- Liberty
- Love
- Practicality
- Stability
- Ideal
- Challenge
- Structure



Values

5 Valores

Describen factores de motivación que influyen en la toma de decisiones

- Self-transcendence
- Tradition
- Hedonism
- Self-enhancement
- Open to change

Indica las preferencias de consumo

Basadas en las características de personalidad inferidas de los textos. Estos resultados indican la tendencia a preferir diferentes productos, servicios y actividades. Las referencias se agrupan en ocho categorías:

| Category | Description | N° Preferences |
|--------------------|---|----------------|
| Shopping | Interest in different types of purchases and author's spending habits | 13 |
| Movie | Interest in different types of movies | 10 |
| Music | Interest in different types of music and if the author enjoys playing music | 9 |
| Reading & Learning | Likelihood to read, author's motivations for reading and types of content | 9 |
| Health & Activity | Interest in healthy foods and physical activity | 5 |
| Volunteering | Interest in volunteering for social causes | 3 |
| Environment | Interest in the environment | 1 |
| Entrepreneurship | Interest in starting a business | 1 |

- The service derives the score from the personality characteristics that it infers from the input text. The score is a double that indicates how likely the author of the text is to prefer the item. It is an indication of preference, not a normalized percentage
- For some preferences, the score represents a binary value (low or high level of interest); for others, is a ternary value (unlikely, neutral and likely)

Indicación de las preferencias de consumo

➤ Ejemplos de preferencias por categoría...

Health & Activity

| Name | Scores |
|------------------------------------|---|
| Likely to eat out frequently | 0.0 (unlikely) 0.5 (neutral) 1.0 (likely) |
| Likely to eat fast food frequently | 0.0 (unlikely) 0.5 (neutral) 1.0 (likely) |
| Likely to have a gym membership | 0.0 (unlikely) 1.0 (likely) |
| Likely to like adventurous sports | 0.0 (unlikely) 1.0 (likely) |
| Likely to like outdoor activities | 0.0 (unlikely) 0.5 (neutral) 1.0 (likely) |

Music

| Name | Scores |
|--------------------------------------|---|
| Likely to like rap music | 0.0 (unlikely) 0.5 (neutral) 1.0 (likely) |
| Likely to like country music | 0.0 (unlikely) 0.5 (neutral) 1.0 (likely) |
| Likely to like R&B music | 0.0 (unlikely) 0.5 (neutral) 1.0 (likely) |
| Likely to like hip hop music | 0.0 (unlikely) 0.5 (neutral) 1.0 (likely) |
| Likely to attend live musical events | 0.0 (unlikely) 1.0 (likely) |

Shopping

| Name | Scores |
|--|---|
| Likely to be sensitive to ownership cost when buying automobiles | 0.0 (unlikely) 0.5 (neutral) 1.0 (likely) |
| Likely to prefer safety when buying automobiles | 0.0 (unlikely) 0.5 (neutral) 1.0 (likely) |
| Likely to prefer resale value when buying automobiles | 0.0 (unlikely) 0.5 (neutral) 1.0 (likely) |
| Likely to prefer quality when buying clothes | 0.0 (unlikely) 1.0 (likely) |
| Likely to prefer style when buying clothes | 0.0 (unlikely) 1.0 (likely) |


Reading & Learning

| Name | Scores |
|--|---|
| Likely to read often | 0.0 (unlikely) 0.5 (neutral) 1.0 (likely) |
| Likely to read for enjoyment | 0.0 (unlikely) 1.0 (likely) |
| Likely to read for information | 0.0 (unlikely) 1.0 (likely) |
| Likely to read entertainment magazines | 0.0 (unlikely) 1.0 (likely) |
| Likely to read non-fiction books | 0.0 (unlikely) 1.0 (likely) |

Nuestra escritura denota nuestra personalidad

- Ejemplo: Partimos del modelo de los Big Five (rasgos de personalidad) y elegimos una forma determinada de medirlos (a través del lenguaje)
- Tened en cuenta que podríamos haber elegido diferentes formas de medida:
 - Tests estandarizados
 - Patrón de contestación (velocidad, ...)
 - Lenguaje utilizado
 - Análisis semántico
 - Análisis sintáctico
 -

Hay un mercado


 Solutions ▾ Product ▾ Get Started ▾ Library ▾ [Start Free Trial](#)

Looking for an IBM Watson Personality Insights API alternative? Look no further.

Big Five personality assessment. DISC personality profile. Workplace Behavioral Factors. All done predictively, via one powerful API.

Business Email*

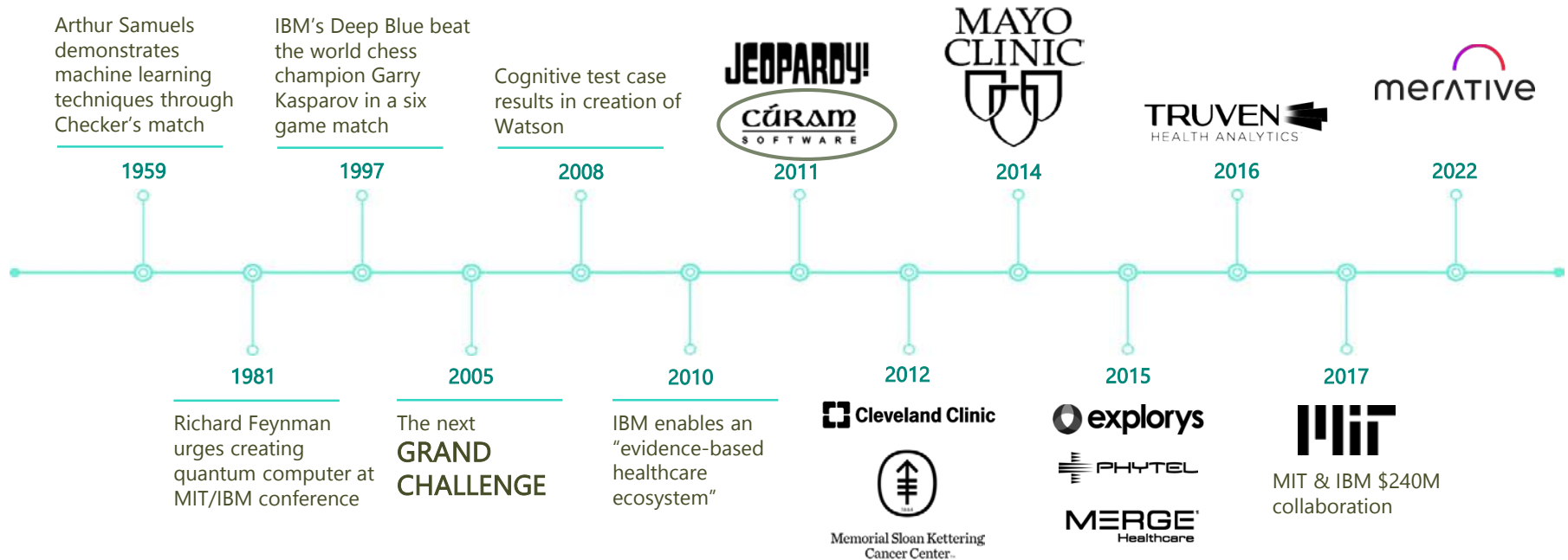
[GET YOUR API KEY](#)



Ejemplos de utilización

- IBM Project Debater
- IBM Personality Insights
- **Watson Health - Merative**

Primeros desarrollos de Watson en el área de la salud

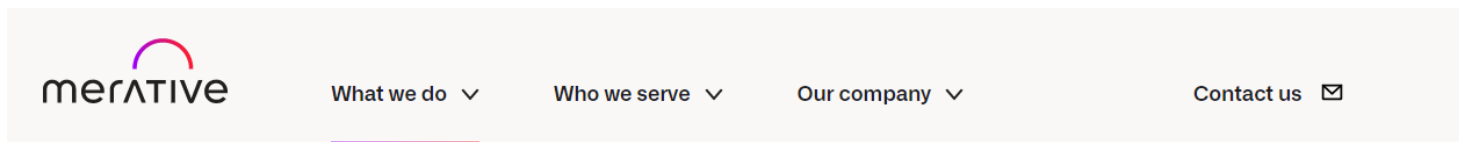


Watson Health es ahora Merative

Armonk, NY and San Francisco, CA – **January 21, 2022 – IBM (NYSE: IBM) and Francisco Partners**, a leading global investment firm that specializes in partnering with technology businesses, today announced that the companies have signed a definitive agreement under which Francisco Partners will acquire healthcare data and analytics assets from IBM that are currently part of the Watson Health business. The assets acquired by Francisco Partners include extensive and diverse data sets and products, including **Health Insights, MarketScan, Clinical Development, Social Program Management, Micromedex, and imaging software offerings.**

The transaction is expected to close in the second quarter of this year and is subject to customary regulatory clearances. Financial terms of the transaction were not disclosed.

“Today’s agreement with Francisco Partners is a clear next step as IBM becomes even more focused on our platform-based hybrid cloud and AI strategy,” said Tom Rosamilia, Senior Vice President, IBM Software. “IBM remains committed to Watson, our broader AI business, and to the clients and partners we support in healthcare IT. Through this transaction, Francisco Partners acquires data and analytics assets that will benefit from the enhanced investment and expertise of a healthcare industry focused portfolio.”



Healthcare analytics

Deliver personalized care experiences by turning health data into actionable insights.

Clinical decision support

Make informed clinical decisions with fast access to the latest evidence-based insights at the point-of-care.

Clinical development

Get to market faster with a unified, cloud-based clinical data acquisition and management system.

Social program management

Connect at-risk populations to health and social services by modernizing social program administration and patient engagement.

Real-world evidence

Support research findings and inform decision-making by generating real-world insights.

Enterprise imaging

Transform at your own pace with a scalable, AI-infused infrastructure that centralizes your imaging data.

<https://merative.com>

Agenda Computación Cognitiva

- Definición de Computación Cognitiva. Cuando tiene sentido utilizarla
- Qué problemas nos encontramos en el hardware y cómo los podemos resolver
- IBM Watson
- **Directrices éticas para una Inteligencia Artificial fiable**

Directrices éticas para una IA fiable

Directrices éticas elaboradas por un grupo de expertos de alto nivel: documento de junio de 2018

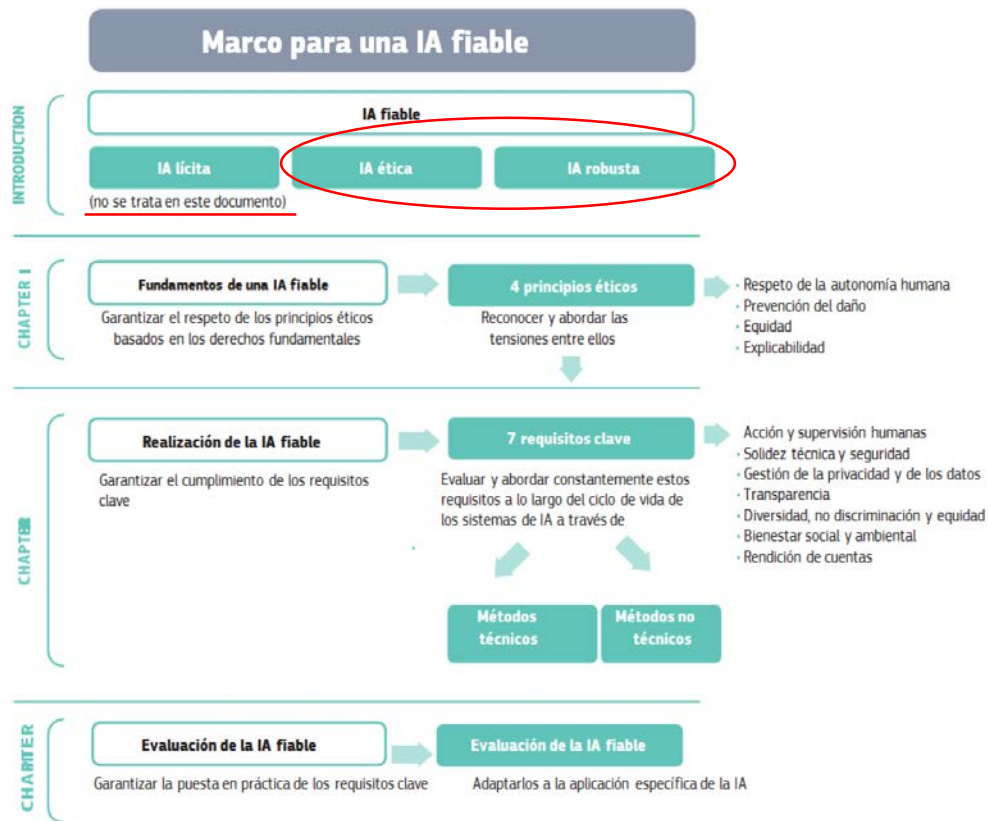
- <https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>

En el grupo de expertos están representados tanto el ámbito académico como el empresarial

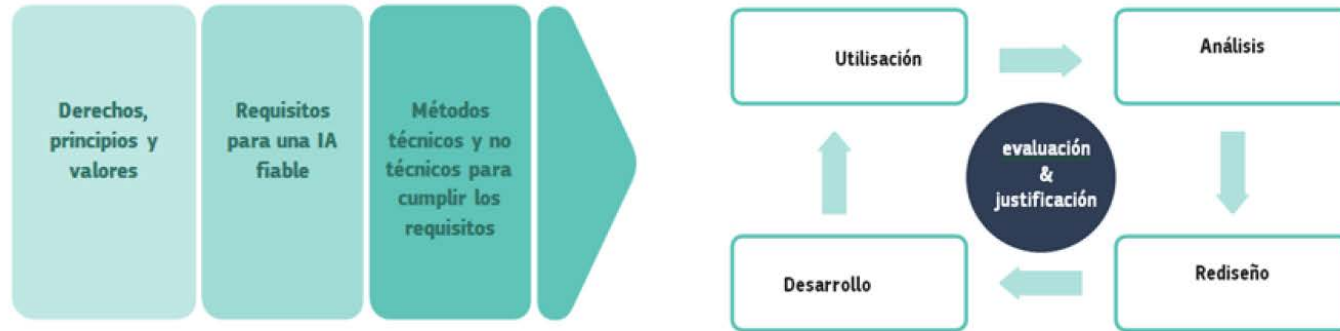
Directrices éticas para una IA fiable

- Se reconoce el efecto positivo tanto comercial como social de la IA
- Preocupación por que los riesgos y otros aspectos adversos se gestionen de manera adecuada
- Es importante desarrollar sistemas de IA merecedores de confianza
- La IA fiable tiene tres componentes:
 - Debe ser lícita de forma que se cumplan todas las leyes y reglamentos aplicables
 - Ha de ser ética, de modo que se garantice el respeto de los principios y valores éticos
 - Debe ser robusta, tanto desde el punto de vista técnico como social, a fin de asegurar que, incluso si las intenciones son buenas, no provoquen daños accidentales

Marco para la construcción de una IA fiable



Construcción de una IA fiable a lo largo del ciclo de vida del sistema



Evaluación de la IA fiable



Evaluación de la IA Fiable

Involucración de las empresas en el proceso de evaluación:

<https://www.ibm.com/design/ai/ethics/everyday-ethics/>

- IBM Design for AI: 5 áreas éticas de foco
 - Responsabilidad (rendir cuentas)
 - Alineación de valores
 - Explicabilidad
 - Equidad
 - Derechos de los datos de los usuarios

Responsabilidad

“Nearly 50% of the surveyed developers believe that the humans creating AI should be responsible for considering the ramifications of the technology. Not the bosses. Not the middle managers. The coders.”

– [Mark Wilson](#), Fast Company on Stack Overflow's [Developer Survey Results 2018](#)

Ejemplo al desarrollar IA para un hotel:

- El equipo utiliza, en el diseño de la solución, investigadores para ponerse en contacto con huéspedes reales en los hoteles para entender sus deseos y necesidades mediante entrevistas cara a cara con los usuarios.
- El equipo considera que es su propia responsabilidad, cuando los comentarios de un asistente de hotel no satisfacen las necesidades o expectativas de los huéspedes. Han puesto en marcha un bucle de aprendizaje de opiniones para comprender mejor las preferencias y han destacado la posibilidad de que un huésped pueda desactivar la IA en cualquier momento de su estancia

Alineación de valores

- El equipo entiende que para que un asistente por voz funcione correctamente, debe estar "siempre a la escucha". El equipo aclara a los huéspedes que el asistente de hotel de IA está diseñado para no guardar ningún dato, ni monitorizar a los huéspedes, en ambos casos sin su conocimiento, incluso si está escuchando una palabra para activarse.
- El audio recogido durante la escucha de una palabra-despertador se borra automáticamente cada 5 segundos. Incluso si un huésped opta por ello, la IA no escucha activamente a los huéspedes a menos que se le solicite.
- El equipo sabe que este agente se utilizará en hoteles de todo el mundo, lo que requerirá diferentes idiomas y costumbres. Consultan con lingüistas para asegurarse de que la IA sea capaz de hablar en los idiomas respectivos de los huéspedes y respetar las costumbres aplicables.

Derechos sobre los datos de los usuarios

“Individuals require mechanisms to help curate their unique identity and personal data in conjunction with policies and practices that make them explicitly aware of consequences resulting from the bundling or resale of their personal information.”

– The [IEEE Global Initiative](#) on Ethics of [Autonomous and Intelligent Systems](#)

➤ Recomendaciones:

- Emplear prácticas de seguridad que incluyan encriptación, metodologías de control de acceso y módulos de gestión de consentimiento propios para restringir el acceso a los usuarios autorizados y anonimizar los datos de acuerdo con las preferencias de los usuarios.

Education and Toolkits

➤ AI explainability 360 (interpretar, explicar,...)

Objetivo: Comprender por qué un sistema se comportó de una determinada manera y por qué ofreció una interpretación específica

Demo: <https://aix360.mybluemix.net/data>

➤ AI fairness 360 (imparcialidad: estudio de sesgos)

Objetivo: Examinar, informar y mitigar la discriminación y el sesgo en los modelos de aprendizaje automático

(En la mayoría de los casos ciertos sesgos se trasladan a los modelos a través de los datos)

Demo: <https://aif360.mybluemix.net/data>

Información adicional

Big Five

Cómo una persona se enfrenta al mundo

Big Five (Costa & McCrae -1992, Norman - 1963) incluye cinco dimensiones primarias con seis aspectos (facetas) cada una de ellas, que caracterizan al individuo de acuerdo con la dimensión

| Dimension | Description | Facets | |
|---|---|---|--|
| Agreeableness Amabilidad | A person's tendency to be compassionate and cooperative toward others | Altruism Cooperation Modesty | Morality Sympathy Trusting others |
| Conscientiousness Responsabilidad | A person's tendency to act in an organized or thoughtful way | Achievement striving Cautiousness Dutifulness | Orderliness Self-discipline Self-efficacy |
| Extraversion Extraversión | A person's tendency to seek stimulation in the company of others | Activity level Assertiveness Cheerfulness | Excitement-seeking Friendliness Gregariousness |
| Emotional Range Neuroticismo/inestabilidad emocional | The extent to which a person's emotions are sensitive to the individual's environment | Anger Anxiety Depression | Immoderation Self-consciousness Vulnerability |
| Openness Apertura a nuevas experiencias | The extent to which a person is open to experiencing a variety of activities | Adventurousness Artistic interests Emotionality | Imagination Intellectual curiosity Liberalism |

Necesidades

Qué aspectos de un producto son más proclives a llamar su atención

Necesidades (Kotler & Armstrong - 2013, Ford - 2005) sugieren que diferentes tipos de necesidades humanas son universales e influyen directamente el comportamiento del consumidor

| Characteristic | People who score high ... | Characteristic | People who score high ... |
|-----------------|---|----------------|--|
| Excitement | Want to get out there and live life, have upbeat emotions, and want to have fun | Liberty | Have a desire for fashion and new things, as well as the need for escape |
| Harmony | Appreciate other people, their viewpoints, and their feelings | Love | Enjoy social contact, whether one-to-one or one-to-many. Any brand that is involved in bringing people together taps this need |
| Curiosity | Have a desire to discover, find out, and grow. | Practicality | Have a desire to get the job done, a desire for skill and efficiency, which can include physical expression and experience |
| Ideal | Desire perfection and a sense of community | Stability | Seek equivalence in the physical world. They favor the sensible, the tried and tested |
| Closeness | Relish being connected to family and setting up a home | Challenge | Have an urge to achieve, to succeed, and to take on challenges |
| Self-expression | Enjoy discovering and asserting their own identities | Structure | Exhibit groundedness and a desire to hold things together. They need things to be well organized |

Valores

Factores de motivación que influyen en la toma de decisiones

Values (Schwartz - 2006) expresan qué es lo más importante para un individuo. Expresan las metas que sirven como principios que guían la vida de las personas

| Value | People who score high ... |
|--------------------------------------|--|
| Self-transcendence / Helping others | Show concern for the welfare and interests of others |
| Conservation / Tradition | Emphasize self-restriction, order, and resistance to change |
| Hedonism / Taking pleasure in life | Seek pleasure and sensuous gratification for themselves |
| Self-enhancement / Achieving success | Seek personal success for themselves |
| Open to change / Excitement | Emphasize independent action, thought, and feeling, as well as a readiness for new experiences |

Personality Insights

1- Big Five: modelo de personalidad que analiza ésta como la composición de cinco dimensiones de personalidad que describen la forma en la que una persona se enfrenta al mundo. El modelo incluye cinco dimensiones (OCEAN):

- *O (Openness o apertura a nuevas experiencias): imaginación activa, sensibilidad estética, latencia a las vivencias internas, gusto por la variedad, curiosidad intelectual e independencia de juicio.*
- *C (Conscientiousness o responsabilidad): auto-control no sólo de impulsos sino también en la planificación, organización y ejecución de tareas. Por esta razón a este factor también se le ha denominado como "voluntad de logro", ya que implica una planificación cuidadosa y persistencia en sus metas. Está asociado además con la responsabilidad, confiabilidad, puntualidad y escurpulosidad.*
- *E (Extraversion o extraversión): alta sociabilidad, tendencia a la compañía de otros, atrevimiento en situaciones sociales, tendencia a evitar la soledad. Existe una tendencia alta a experimentar emociones positivas tales como alegría, satisfacción, excitación, etc.*
- *A (Agreeableness o amabilidad): Refleja tendencias interpersonales. En su polo positivo, el individuo es altruista, considerado, confiado y solidario.*
- *N (Neuroticism o inestabilidad emocional): inestabilidad emocional, ansiedad, alta preocupación, etc*

Personality Insights

2 –Valores: describen factores de motivación que influyen en las decisiones de las personas. El modelo incluye cinco dimensiones de valores humanos:

Self-transcendence / Helping others, Conservation / Tradition, Hedonism / Taking pleasure in life, Self-enhancement / Achieving success, and Open to change / Excitement.

3- Necesidades Ciertas investigaciones sugieren que varios tipos de necesidades humanas son universales y que influyen directamente el comportamiento de los consumidores. Las doce categorías citadas en el servicio de Personality Insights, se describen en la literatura de marketing como deseos que una persona espera satisfacer cuando está considerando un producto. El modelo incluye doce necesidades características:

Excitement, Harmony, Curiosity, Ideal, Closeness, Selfexpression, Liberty, Love, Practicality, Stability, Challenge, and Structure.