

Soft computing y análisis inteligente de datos guiado por ingeniería del conocimiento

José A. Olivas

viu

Universidad
Internacional
de Valencia



*“Actualmente es muy frecuente **vincular la Inteligencia Artificial solo al análisis (¿inteligente?) de datos**, y éste a las técnicas de 'aprendizaje automático' o 'machine learning', en su mayoría provenientes de la estadística, más o menos básica. Pero la Inteligencia Artificial es mucho más que eso, más aún, algo muy diferente: **dotar a sistemas computacionales con la capacidad de simular comportamientos inteligentes humanos, tanto en raciocinio como en actuación.***

*Por otra parte, asistimos al fenómeno del 'endiosamiento' de los datos, cuando sabemos que son representaciones muy **incompletas, mayoritariamente numéricas, con ruido, imprecisión, incertidumbre y cada vez más, deliberadamente sesgadas o falsas, habitualmente no con las mejores intenciones.***

*En esta charla se abordarán todos estos aspectos y se establecerán **puentes entre 'correlación' y 'causalidad'** a través de la ingeniería del conocimiento, proponiendo finalmente el desarrollo de sistemas que se apoyen **tanto en el aprovechamiento inteligente de datos como en el del conocimiento** sobre el fenómeno a ser abordado.”*

José A. Olivas

Agenda

1. **¿Inteligencia Artificial?...** Ni tecnología ni cerebro: mitos y malentendidos sobre la inteligencia artificial.
2. ***Soft Computing:*** ¿Qué es el *Soft Computing*?, Implicaciones del *Soft Computing*.
Computing With Words (and Perceptions)
3. ***Soft Computing, Aprovechamiento y Análisis inteligente de datos en entornos Big Data:*** Ejemplos y propuestas.
4. **Ingeniería de Conocimiento y Machine Learning:** Origen, definición.
5. ***Análisis Inteligente de Datos guiado por Ingeniería de Conocimiento:*** De la correlación a la causalidad...
6. ***Resumen y reflexiones:*** Desarrollo software en IA, Formación en IC.

1. ¿Inteligencia Artificial?...

¿Inteligencia Artificial?...

Ni tecnología ni cerebro: mitos y malentendidos sobre la inteligencia artificial*

- La inteligencia artificial busca **explicar el funcionamiento del cerebro**.
“Los cerebros se dan **solo en las criaturas vivas** y no en las máquinas artificiales: **neurociencia** y la **neurobiología**”.
- Los **métodos de inteligencia artificial** funcionan como cerebros: las estructuras de conectividad que están presentes en estos modelos **no son** "biológicamente plausibles ni realistas".

*Montse Hidalgo

https://retina.elpais.com/retina/2020/08/20/tendencias/1597921218_374623.html

Frank Emmert-Streib, Olli Yli-Harja y Matthias Dehmer.

¿Inteligencia Artificial?...

Ni tecnología ni cerebro: mitos y malentendidos sobre la inteligencia artificial*

- El propósito de los métodos de inteligencia artificial es **distinto del aprendizaje automático o la estadística**. "No. El propósito general de todos los métodos de estos campos es **analizar datos**". Cambian las **metodologías** y los **objetivos finales**.
- La inteligencia artificial es una **tecnología**. En rigor, estaríamos hablando de una **metodología**. "Si **desescalamos un problema**, las necesidades de hardware se reducen, demostrando los **principios por los que se rige el método**."

*Montse Hidalgo

https://retina.elpais.com/retina/2020/08/20/tendencias/1597921218_374623.html

Frank Emmert-Streib, Olli Yli-Harja y Matthias Dehmer.

¿Inteligencia Artificial?...

Ni tecnología ni cerebro: mitos y malentendidos sobre la inteligencia artificial*

- La inteligencia artificial hace que **los ordenadores piensen**. Complejidad de **acotar lo que implica la inteligencia**. ¿Qué es pensar? “En general, está asociado con humanos y seres biológicos más que con máquinas”.
- La inteligencia artificial **es más mítica** que el *machine learning* y la estadística. Conjeturas e ilusiones y grandes esperanzas futuristas que caracterizan al campo: máquinas que piensan, máquinas humanas...

*Montse Hidalgo

https://retina.elpais.com/retina/2020/08/20/tendencias/1597921218_374623.html

Frank Emmert-Streib, Olli Yli-Harja y Matthias Dehmer.

¿Inteligencia Artificial?...

Ni tecnología ni cerebro: mitos y malentendidos sobre la inteligencia artificial*

- Crear máquinas que se comporten como humanos es óptimo. “El **comportamiento** de los humanos en general en una tarea **no es perfecto** comparado con el del humano que alcanza **mejor rendimiento** en ella”.
- ¿Cuándo alcanzaremos las **metas finales de la inteligencia artificial**?
“Esto empezó en los principios de la era de la inteligencia artificial, en los cincuenta la gente decía que en cuestión de 10, 15, 25 años íbamos a tener una inteligencia artificial completa y capaz de remplazar al ser humano en todos los empleos...”

*Montse Hidalgo

https://retina.elpais.com/retina/2020/08/20/tendencias/1597921218_374623.html

Frank Emmert-Streib, Olli Yli-Harja y Matthias Dehmer.

¿Inteligencia Artificial?...

“El público general y los intelectuales todavía sobrestiman la inteligencia artificial”*

- “Mi mayor miedo es que confiemos en una inteligencia artificial que **no es tan lista**”
- Descarta que seamos capaces de **crear máquinas que nos sometan** y advierte que estamos **infravalorando la complejidad de la mente humana**:
 - Frank Rosenblatt: "Se espera que el ordenador sea capaz de caminar, hablar, ver escribir, reproducirse y ser consciente de su existencia".
 - Claude Shannon (1961): "Espero con seguridad que en cuestión de diez o quince años algo no lejano de los robots de ciencia ficción emergerá del laboratorio".
 - Herbert Simon (1965): "En veinte años, las máquinas serán capaces de hacer cualquier trabajo que pueda hacer un humano".

*Montse Hidalgo

https://retina.elpais.com/retina/2019/12/11/tendencias/1576063514_700068.html

Melanie Mitchell: *Artificial Intelligence: A Guide for Thinking Humans*.

¿Inteligencia Artificial?...

“El público general y los intelectuales todavía sobrestiman la inteligencia artificial”*

- “John McCarthy le puso un nombre tan ambicioso como ‘**inteligencia artificial**’. Herbert Simon propuso otro nombre: ‘**Procesamiento de información compleja**’. Es mucho menos sexi pero tal vez **es más preciso**. Hay un atractivo en ‘inteligencia’ que no ves en ‘procesamiento de información’.”
- “Ocurre lo mismo con el ***deep learning***, que suena muy importante y profundo, pero la palabra ‘hondo’ (*deep*) no hace referencia a ninguna idea de comprensión o inteligencia profunda, sino al **tamaño de la red neuronal**, a cuantas capas de neuronas simuladas tiene.

*Montse Hidalgo

https://retina.elpais.com/retina/2019/12/11/tendencias/1576063514_700068.html

Melanie Mitchell: *Artificial Intelligence: A Guide for Thinking Humans*.

¿Inteligencia Artificial?...

“El público general y los intelectuales todavía sobrestiman la inteligencia artificial”*

- “Además, utilizamos la palabra ‘aprendizaje’ (*learning*) para estos sistemas, que **distan mucho del aprendizaje humano**. Pero bueno, si los llamamos así, suenan mucho más inteligentes de lo que realmente son.”
- “Tenemos tecnologías de inteligencia artificial a las que estamos dando **autonomía**. Las ponemos a **tomar decisiones** sobre **gente, cárceles**, solicitudes de **préstamos**... La cuestión es: ¿**confiamos** en ellas?, ¿estamos **aplicándolas** como deberíamos?, ¿reconocemos sus **limitaciones**?, ¿necesitamos **regularlas**?”

*Montse Hidalgo

https://retina.elpais.com/retina/2019/12/11/tendencias/1576063514_700068.html

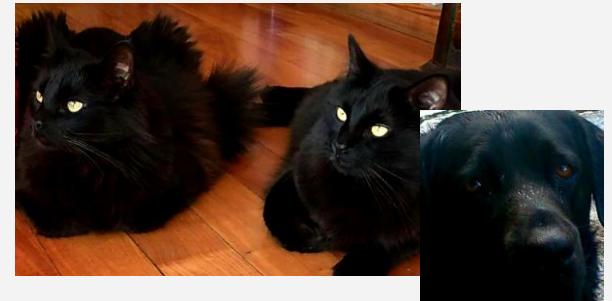
Melanie Mitchell: *Artificial Intelligence: A Guide for Thinking Humans*.

¿Inteligencia Artificial?...

“La inteligencia artificial es muy burra”*

- “Y, aún así, **los animales son más listos...** vuelo sincronizado de las **bandadas de pájaros**, **visión** de las serpientes, capacidad de **orientación** de los gatos, los loros pueden **replicar el sonido**, los perros... Y sin embargo, ninguno de esos

animales tiene las **habilidades cognitivas** y la inteligencia que típicamente **atribuimos a los humanos.**”



- “El problema es el potencial de la inteligencia artificial, que puede acabar por **desorientar** a jóvenes generaciones de investigadores e incluso a una **asignación inadecuada de recursos** a nivel **académico, industrial y gubernamental.**”

*https://retina.elpais.com/retina/2018/11/14/innovacion/1542207383_538303.html

Adnan Darwiche, UCLA.

¿Inteligencia Artificial?...

“La inteligencia artificial es muy burra”*

- “Trabajo de una **comunidad investigadora más amplia**, menos centrada en el ‘mira lo que podemos hacer’ y más orientada a la consecución de **objetivos conjuntos**.”
- “Este objetivo exige el **entrenamiento** de una **nueva generación** de investigadores que conozcan y aprecien **distintos métodos** de inteligencia artificial -no solo redes neuronales- y estén mejor informados sobre la **historia** de este campo”.

*https://retina.elpais.com/retina/2018/11/14/innovacion/1542207383_538303.html

Adnan Darwiche, UCLA.

¿Inteligencia Artificial?...

Por lo tanto...

IA \neq ML \neq DL

- ¡No sólo se hace IA a partir de datos!
- Modelos de simulación de la inteligencia humana:
 - Razonamiento, inferencia.
 - Incertidumbre, imprecisión.
 - Causalidad, heurística, categorización conceptual, almacenamiento de experiencias.
 - Abstracción, lenguaje natural.
 - Sentimientos, emociones.

¿Inteligencia Artificial?...

IA ≠ ML ≠ DL

- **Otras áreas** de la IA en las que se debe profundizar...
 - Ingeniería del conocimiento.
 - PLN.
 - Reconocimiento de patrones.
 - Razonamiento, inferencia.
 - Incertidumbre, imprecisión.
 - Causalidad, heurística, categorización conceptual, almacenamiento de experiencias.
- ...que no dejan de estar **profundamente vinculadas** con el 'Big Data' (Aprovechamiento inteligente de datos masivos)

2. Soft Computing

Soft Computing

¿Qué es el *Soft Computing*?

- El profesor **Zadeh** acuñó el término de ***soft-computing***, que se puede traducir al español como **computación suave o blanda**.
- La computación **suave** se diferencia de la computación **convencional (dura)** en que, a diferencia de ella, es **tolerante a la imprecisión, la incertidumbre y la verdad parcial**.
- El modelo a seguir para la computación suave es **la mente humana** y su principio rector es **aprovechar la tolerancia a la imprecisión, la incertidumbre y la verdad parcial** para lograr la **trazabilidad, la robustez** y el **bajo coste** de las soluciones.

Soft Computing

¿Qué es el *Soft Computing*?

- Las ideas básicas que subyacen a la computación suave en su estado actual tienen vínculos con muchas influencias anteriores, entre ellas los **conjuntos borrosos** introducidos en los años sesenta, los trabajos de los setenta sobre el **análisis de sistemas complejos y procesos de decisión**, y los de los años ochenta sobre **teoría de posibilidades y análisis de datos blandos**.
- La inclusión de la teoría de **redes neuronales** en la computación suave llegó en un momento posterior.

Soft Computing

¿Qué es el *Soft Computing*?

- Los principales componentes de la computación suave (CS) son
 - la lógica borrosa (LB),
 - la teoría de redes neuronales (RN) y el razonamiento probabilístico (RP), con este último subsumiendo las redes de creencias,
 - los algoritmos genéticos,
 - la teoría del caos,
 - partes de la teoría del aprendizaje.
- Importante: el CS **no es una mezcla** de LB, RN y RP.
- Se trata más bien de una **asociación** en la que cada uno de los socios aporta una metodología distinta para abordar los problemas en su ámbito.
- En esta perspectiva, las principales contribuciones de LB, RN y RP son **complementarias y no competitivas**.

Soft Computing

Implicaciones del *Soft Computing*

- La complementariedad de LB, RN y RP tiene una consecuencia importante: en muchos casos **un problema puede resolverse de forma más eficaz** utilizando LB, RN y RP **en combinación y no exclusivamente**.
- Un ejemplo llamativo de una combinación particularmente efectiva es lo que se ha llegado a conocer como sistemas **neuroborrosos** (*neurofuzzy*).

Soft Computing

Implicaciones del *Soft Computing*

- Estos sistemas son cada vez más visibles como **productos de consumo e industriales**, desde aires acondicionados y lavadoras hasta fotocopiadoras y videocámaras.
- Tanto en los **productos de consumo** como en los **sistemas industriales**, el empleo de técnicas de *soft computing* conduce a sistemas que tienen un **alto cociente de inteligencia de máquina (IM)**.

Soft Computing

Implicaciones del *Soft Computing*

- La estructura conceptual de la computación suave sugiere que los **estudiantes** deben ser entrenados no solo en teoría de redes neuronales, lógica borrosa o razonamiento probabilístico, sino en **todas las metodologías asociadas**, aunque **no necesariamente en el mismo grado**.
- Hay muchas **revistas científicas** y libros con este término en su título.
- Una tendencia similar es perceptible en los títulos de las **conferencias y congresos**.

Soft Computing

Computing With Words (and Perceptions)

- In computing with words and perceptions (CWP), the objects of computation are **words, perceptions, and propositions** drawn from a **natural language**.
- The central theme of CWP is the concept of a **generalised constraint**. The **meaning** of a proposition is expressed as a generalized constraint.
- CWP is a necessary tool when the available information is **perception-based** or **not precise enough to use numbers**.

L. A. Zadeh, "Fuzzy logic = computing with words," IEEE Trans. on Fuzzy Systems, vol. 4, pp. 103-111, 1996

Soft Computing

Por lo tanto...

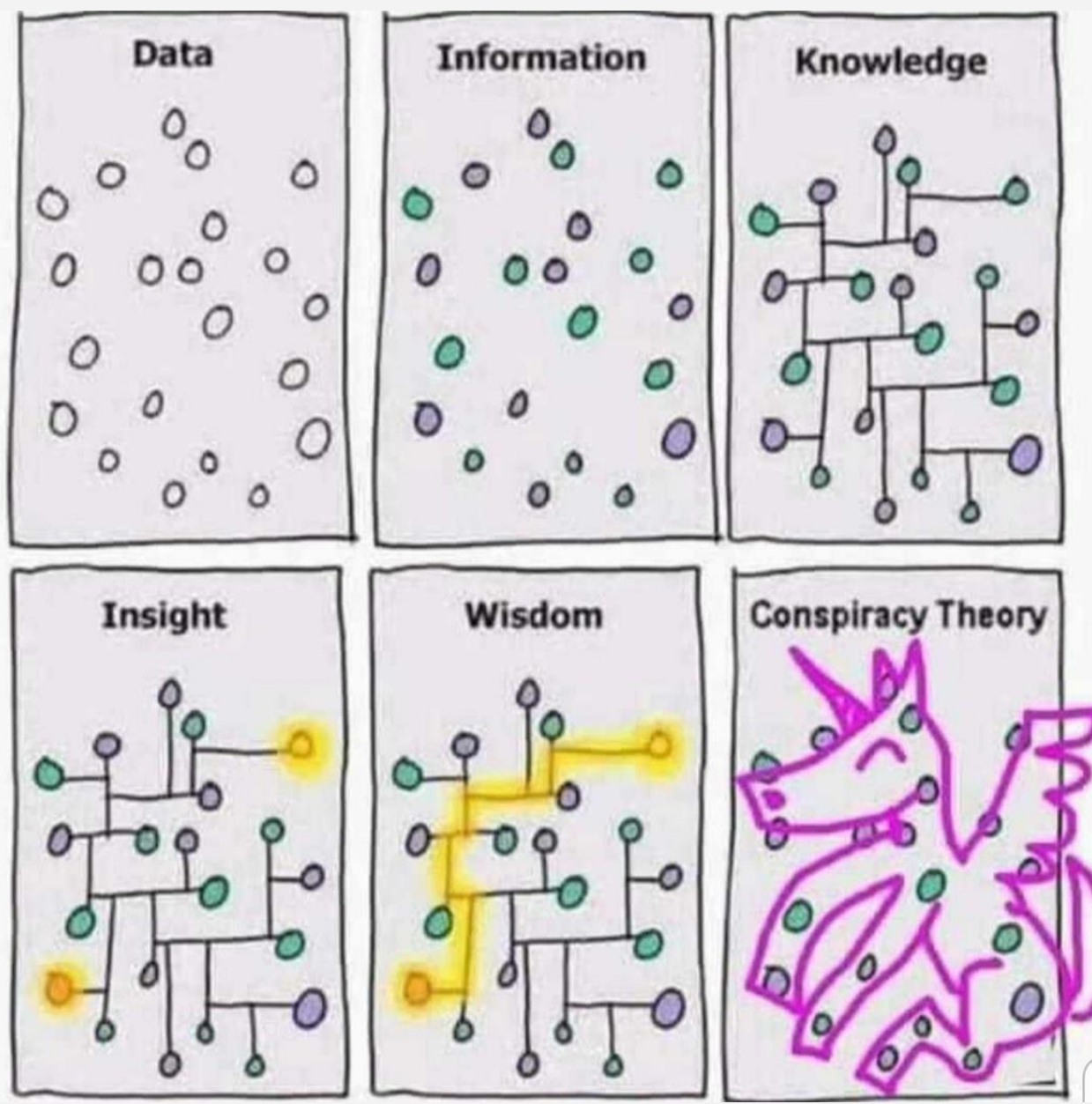
- El Soft Computing es **TRANSVERSAL** a la IA y el ML.
- Para afrontar retos complejos se deben proponer **SOLUCIONES SOFISTICADAS**, no sólo el uso de **herramientas específicas aisladas** para subproblemas muy concretos.

¡ADAPTABILIDAD!

- El concepto de '**DATA LAKE**' obliga a enfrentarnos al gran cuello de botella:

REDUCIR LA COMPLEJIDAD DEL MUNDO REAL A UN CONJUNTO DE DATOS ESTRUCTURADO, NUMÉRICO Y NORMALIZABLE (Entrada de todos los algoritmos).

3. Soft Computing, Aprovechamiento y Análisis inteligente de datos en entornos Big Data



¿Inteligencia Artificial?...

“El mono de Wall Street”*

- En 1973 Burton Gordon Malkiel escribió “**Un paseo aleatorio por Wall Street**” donde afirmaba que los mercados financieros se mueven de forma aleatoria e impredecible.
- Malkiel argumentaba que no es posible batir al mercado sin **información privilegiada**, y afirmaba vehementemente que un mono con los ojos vendados lanzando dardos sobre una lista de empresas cotizantes en la bolsa de Nueva York, generaría retornos similares a una cartera elaborada por cualquier experto analista de Wall Street.

*<https://www.neotrading.es/la-historia-de-un-mono-unos-dardos-y-wall-street>

Burton, N. (2018). *An Analysis of Burton G. Malkiel's A Random Walk Down Wall Street*.

Ejemplos y Propuestas

El dilema del médico informatizado...

- Supongamos un médico ante un paciente con un caso **complejo, difícil** de diagnosticar (y, consecuentemente, de tratar), que decide '**ayudarse**' de un **sistema computacional supuestamente** 'inteligente' que le ayude a **tomar decisiones** al respecto.
- Dispone de varias opciones, que le han facilitado diferentes **equipos de investigación** universitarios y **empresas** del entorno hospitalario...

Ejemplos y Propuestas

El dilema del médico informatizado...

1995 (Pocos datos históricos)

Pentium 8Mb Manual

Sistema Experto Prot. borrosos
(INFEDEC...)

Sistema Experto Simple
(Reglas sencillas)

Sistema Experto Fact. certeza
(Reglas)

Red neuronal
Recurrente, perceptrón...

Árbol de decisión
(CART, ID3...)

Raz. Basado en Casos

Sistema Basado en Minería de datos (Reg. Asoc. Weka...)

Sistema bayesiano/probabilístico



Ejemplos y Propuestas

El dilema del médico informatizado...

1995 (Pocos datos históricos) Pentium 8Mb Manual	2021 (Más datos históricos) 17 32Gb Herramientas y serpientes
Sistema Experto Prot. borrosos (INFEDEC...)	Inferencia borrosa (tipo control, TSK, Mamdani...)
Sistema Experto Simple (Reglas sencillas)	
Sistema Experto Fact. certeza (Reglas)	<i>Relaciones Causales</i>
Red neuronal Recurrente, perceptrón...	<i>Deep Learning (Keras, Tensor Flow...)</i>
Árbol de decisión (CART, ID3...)	<i>Random Forest (Boost...)</i>
Raz. Basado en Casos	
Sistema Basado en Minería de datos (Reg. Asoc. Weka...)	Otras técnicas de <i>clustering</i> , clasificación, metaheurísticas...
Sistema bayesiano/probabilístico	Cuadro de Mandos, BI (<i>Power BI...</i>)

Ejemplos y Propuestas

El dilema del médico informatizado...

1995 (Pocos datos históricos) Pentium 8Mb Manual	2021 (Más datos históricos) I7 32Gb Herramientas y serpientes
Sistema Experto Prot. borrosos (INFEDEC...)	
	<i>Relaciones Causales</i>
	<i>Deep Learning (Keras, Tensor Flow...)</i>
Raz. Basado en Casos	

Ejemplos y Propuestas

El dilema del médico informatizado...

1995 (Pocos datos históricos) Pentium 8Mb Manual	2021 (Más datos históricos) I7 32Gb Herramientas y serpientes
Sistema Experto Prot. borrosos (INFEDEC...)	
	<i>Relaciones Causales</i>
	<i>Deep Learning (Keras, Tensor Flow...)</i>
Raz. Basado en Casos	

¿Con cuál nos quedaríamos?

Ejemplos y Propuestas

El dilema del médico informatizado...

1995 (Pocos datos históricos) Pentium 8Mb Manual	2020 (Más datos históricos) I7 32Gb Herramientas y serpientes
Sistema Experto Prot. borrosos (INFEDEC...)	
	<i>Relaciones Causales</i>
	<i>Deep Learning (Keras, Tensor Flow...)</i>
Raz. Basado en Casos	

¿Con cuál nos quedaríamos?
Quizá con una mezcla de
todos...

Ejemplos y Propuestas

El dilema del médico informatizado...

1995 (Pocos datos históricos) Pentium 8Mb Manual	2021 (Más datos históricos) 17 32Gb Herramientas y serpientes
Sistema Experto Prot. borrosos (INFEDEC...)	Inferencia borrosa (tipo control, TSK, Mamdani...)
Sistema Experto Simple (Reglas sencillas)	
Sistema Experto Fact. certeza (Reglas)	<i>Relaciones Causales</i>
Red neuronal Recurrente, perceptrón...	<i>Deep Learning (Keras, Tensor Flow...)</i>
Árbol de decisión (CART, ID3...)	<i>Random Forest (Boost...)</i>
Raz. Basado en Casos	
Sistema Basado en Minería de datos (Reg. Asoc. Weka...)	Otras técnicas de <i>clustering</i> , clasificación, metaheurísticas...
Sistema bayesiano/probabilístico	Cuadro de Mandos, BI (<i>Power BI...</i>)















Ejemplos y Propuestas

El dilema del médico informatizado...

1995 (Pocos datos históricos) Pentium 8Mb Manual	2021 (Más datos históricos) I7 32Gb Herramientas y serpientes
Sistema Experto Prot. borrosos (INFEDEC...) ANTIGUO y RARO	Inferencia borrosa (tipo control, TSK, Mamdani...) RAROS
Sistema Experto Simple (Reglas sencillas) VIEJUNO	
Sistema Experto Fact. certeza (Reglas) VIEJUNO	<i>Relaciones Causales</i> RAROS
Red neuronal Recurrente, perceptrón, neurofuzzy...	<i>Deep Learning (Keras, Tensor Flow...)</i>
Árbol de decisión (CART, ID3...)	<i>Random Forest (Boost...)</i>
Raz. Basado en Casos VIEJUNO	
Sistema Basado en Minería de datos (Reg. Asoc. Weka...) ANTIGUO	Otras técnicas de <i>clustering</i> , clasificación, metaheurísticas...
Sistema bayesiano/probabilístico SOBREVIVE A DURAS PENAS	Cuadro de Mandos, BI (<i>Power BI...</i>)

Ejemplos y Propuestas

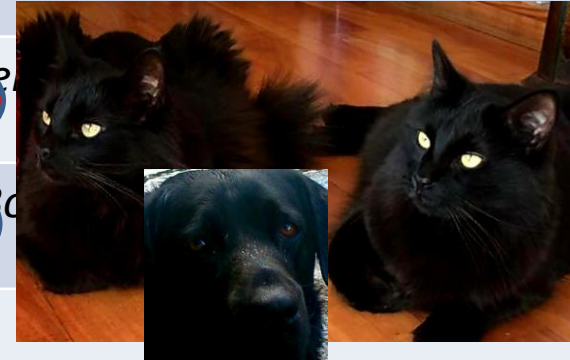
El dilema del médico informatizado...

1995 (Pocos datos históricos) Pentium 8Mb Manual	2021 (Más datos históricos) I7 32Gb Herramientas y serpientes
Sistema Experto Prot. borrosos (INFEDEC...) ANTIGUO y RARO 	Inferencia borrosa (tipo control, TSK, Mamdani...) RAROS 
Sistema Experto Simple (Reglas sencillas) VIEJUNO 	
Sistema Experto Fact. certeza (Reglas) VIEJUNO 	Relaciones Causales RAROS 
Red neuronal Recurrente, perceptrón, neurofuzzy... 	Deep Learning (Keras, Tensor Flow...) 
Árbol de decisión (CART, ID3...) 	Random Forest (Boost...) 
Raz. Basado en reglas VIEJUNO 	
Sistema Basado en Minería de datos (Reg. Asoc. Weka...) ANTIGUO 	Otras técnicas de clustering, clasificación, metaheurísticas... 
Sistema bayesiano/probabilístico SOBREVIVE A DURAS PENAS 	Cuadro de Mandos, BI (Power BI...) 

Ejemplos y Propuestas

El dilema del médico informatizado...

1995 (Pocos datos históricos) Pentium 8Mb Manual	2021 (Más datos históricos) I7 32Gb Herramientas y serpientes
Sistema Experto Prot. borrosos (INFEDEC...) ANTIGUO y RARO S C	Inferencia borrosa (tipo control, TSK, Mamdani...) RAROS S C
Sistema Experto Simple (Reglas sencillas) VIEJUNO S C	
Sistema Experto Fact. certeza (Reglas) VIEJUNO S C	Relaciones Causales RAROS S C
Red neuronal Recurrente, perceptrón, neurofuzzy... S C	Deep Learning RAROS S C
Árbol de decisión (CART, ID3...) S C	Random Forest RAROS S C
Raz. Basado en reglas VIEJUNO S C	
Sistema Basado en Minería de datos (Reg. Asoc. Weka...) ANTIGUO S C	Otras técnicas de clustering, clasificación, metaheurísticas... S C
Sistema bayesiano/probabilístico SOBREVIVE A DURAS PENAS S C	Cuadro de Mandos, BI (Power BI...) S C



Ejemplos y Propuestas

Problema del médico informatizado...

1995 (Pentium 80)	2021 (Más datos históricos) 17 32Gb Herramientas y serpientes
Sistema Experto (INFEDEC...) ANTIGUO	Inferencia borrosa (tipo control, TSK, Mamdani...) RAROS
Sistema Experto Simple (Reglas sencillas) VIEJUNO	
Sistema Experto Fact. certeza (Reglas) VIEJUNO	...les RAROS
Red neuronal Recurrente, perceptrón, neurofuzzy...	
Árbol de decisión (CART, ID3...)	Random...
Raz. Basado en los VIEJUNO	
Sistema Basado en Minería de datos (Reg. Asoc. Weka...) ANTIGUO	Otras técnicas de clusterización, metaheurísticas...
Sistema bayesiano/probabilístico SOBREVIVE A DURAS PENAS	Cuadro de Mandos, BI (Power BI...)

REDUCIR LA COMPLEJIDAD DEL MUNDO REAL A UN CONJUNTO DE DATOS ESTRUCTURADO, NUMÉRICO Y NORMALIZABLE (Entrada de todos los algoritmos).



Ejemplos y Propuestas

Problema del médico informatizado

1995 (Antiguo)	2021 (Más datos históricos, más herramientas)
Pentium 80	17 32Gb Herramientas
Sistema Experto (INFEDEC...) ANTIGUO	Inferencia Control, TSK, Mam...
Sistema Experto Simple (Reglas sencillas) VIEJUNO	
Sistema Experto Fact. certeza (Reglas) VIEJUNO	...les RAROS
Red neuronal Recurrente, per...	
Árbol de decisión (CART)	Random...
Raz... VIEJUNO	
Sistema basado Minería de datos (Reg. Asoc. Weka...) ANTIGUO	Otras técnicas de cluster... fificación,
Sistema bayesiano/probabilístico SOBREVIVE A DURAS PENAS	Cuadro de Ma...s, BI (Power BI...)

REDUCIR LA COMPLEJIDAD
CONJUNTO DE DATOS ES
NORMALIZABLE (Entrada)

¡Y ADEMÁS EL MÉDICO ES CUBANO!

EL MUNDO REAL A UN
TURADO, NUMÉRICO Y
en todos los algoritmos).



Ejemplos y Propuestas

Problema del médico informatizado

¿REALMENTE
HACEMOS IA?

1995 (Pentium 80486)

Sistema Experto (INFEDEC...)

Sistema (Regla)

Sistema (Regla)

Red neuronal Recurrente

Árbol de decisión (CART)

Razonamiento VIEJUNO

Sistema basado en Minería de datos (Reg. Asoc. Weka...) ANTIGUO

Sistema bayesiano/probabilístico SOBREVIVE A DURAS PENAS

2000 (Más datos históricos)

Otras técnicas de clusterización, metaheurísticas...

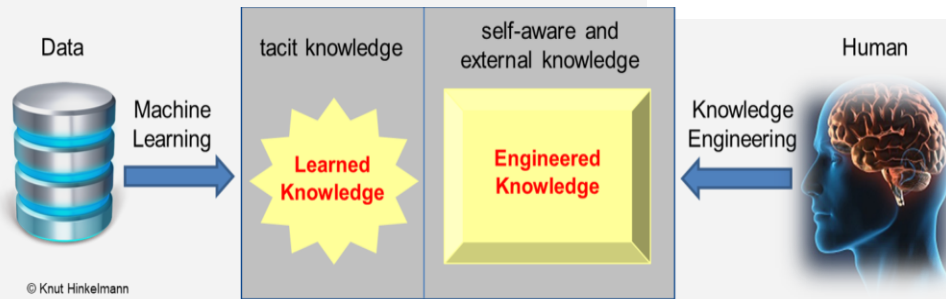
Cuadro de Mandos, BI (Power BI...)

¡ERRO!

¡Y ADIÓS!

¡NO COY NOS).





4. Ingeniería de Conocimiento y Machine Learning

Origen*

- Allen **Newell** y Herbert **Simon**: Sistemas de Producción.
- Richard **Waldinger**: SRI Stanford.



Ingeniería de Conocimiento

Origen*

Artificial intelligence Artificial intelligence--Medical applications Artificial intelligence--Research Automatic theorem proving Buchanan, Bruce G
California--Stanford Choice (Psychology) Clinical medicine--Decision making--Data processing Communication Computational linguistics
Computer-assisted instruction Computer programming Computers Conflict (Psychology) Diagnosis Diagnosis--
Computer programs Diagnosis--Data processing Educational technology Electronic data processing Electronic data processing--Distributed processing English
language English language--Data processing Expert systems
(Computer science) Expert systems (Computer science)--Evaluation Feigenbaum, Edward A Games of strategy (Mathematics)
Heuristic programming Knowledge acquisition (Expert systems) Knowledge representation (Information theory)
Linguistics LISP (Computer program language) Machine theory Mathematics--Research Medical
informatics Medicine--Data processing MYCIN (Computer system) Natural language
processing (Computer science) Prisoner's dilemma game Problem solving
Programming languages (Electronic computers) Rational choice theory Recursion theory
Recursive functions Research--Data processing Social interaction Stanford University--Computer Science Department Stanford
University--Heuristic Programming Project Structural analysis (Engineering) Structural analysis (Engineering)--Data
processing Structural engineering--Computer programs

*Stanford University Heuristic Programming Project

Knowledge Engineering: The Applied Side of Artificial Intelligence

Edward A. Feigenbaum
Computer Science Department
Stanford University
Stanford, CA USA 94305



1.0 Introduction: Symbolic Computation and Inference

This paper will discuss the applied artificial intelligence work that is sometimes called "knowledge engineering". The work is based on computer programs that do symbolic manipulations and symbolic inference, not calculation. The programs I will discuss do essentially no numerical calculation. They discover qualitative lines-of-reasoning leading to solutions to problems stated symbolically.

1.1 Knowledge

Since in this paper I often use the term "knowledge", let me say what I mean by it. The knowledge of an area of expertise—of a field of practice—is generally of two types: a) Facts of the domain—the widely shared knowledge that is written in textbooks, and in journals of a field; that constitutes the kind of material that a professor would lecture about in a class. b) Equally as important to the practice of a field is the heuristic knowledge—knowledge which constitutes the rules of expertise, the rules of good practice, the judgmental rules of the field, the rules of plausible reasoning. These rules collectively constitute what the mathematician, George Polya, has called the "art of good guessing". In contrast to the facts of

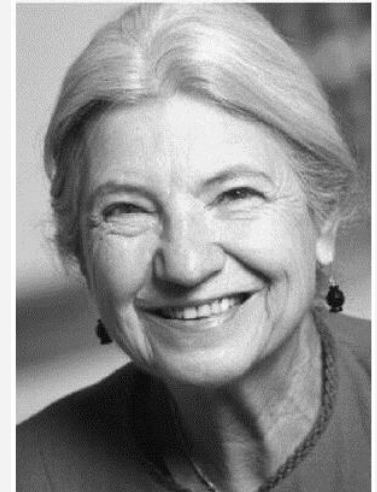
Origen*

- **CITRIS and data and knowledge engineering: What is old and what is new?**

Ruzena Bajcsy, Rick McGeerb

Data & Knowledge Engineering

Volume 50, Issue 3, September 2004, Pages 261-276



***CITRIS** and the Banatao Institute is a University of California Berkeley research center focused on creating IT solutions that generate social and economic benefits for everyone.

Definición*

- La **Ingeniería del Conocimiento (IC)** se refiere a todos los aspectos técnicos, científicos y sociales que intervienen en la construcción, el mantenimiento y la utilización de los **sistemas basados en el conocimiento**.
- Campo **multidisciplinar** que incorpora conceptos y métodos de varios ámbitos de la informática, como la **inteligencia artificial**, las **bases de datos**, los **sistemas expertos**, los sistemas de **apoyo a la toma de decisiones** y los **sistemas de información**.
- Desde el punto de vista del desarrollo de software, utiliza principios y metodologías de la **ingeniería de software**.

*KEOD'21, IKE'21...

Definición*

- Está relacionada con la **lógica** (tanto en el ámbito de las matemáticas como en el de la filosofía) y también está fuertemente relacionada con las **ciencias cognitivas y sociales**.
- El conocimiento es producido por los **seres humanos** y estructurado de acuerdo con nuestra **comprensión** de cómo funciona el **razonamiento** y la **lógica humana**.
- Importancia de las **ontologías**: construcción de marcos conceptuales compartidos.

*KEOD'21, IKE'21...

Ingeniería de Conocimiento y Machine Learning*

- Muchas soluciones de IA actuales se basan en enfoques de **aprendizaje automático**, en muchos casos con gran éxito.
- El aprendizaje automático ayuda a resolver **tareas complejas** basándose en **datos del mundo real** y no en la pura intuición. Es el más adecuado para construir sistemas de IA cuando el conocimiento **no es conocido** o es **tácito**.
- Aunque el aprendizaje automático ya es capaz de afrontar las tareas de aprendizaje con **gran cantidad de datos**, todavía existen algunos retos.

***AAAI-MAKE 2021**: Combining Machine Learning and Knowledge Engineering
AAAI Spring Symposium on March 22-24, 2021, Stanford University, Palo Alto, California, USA.

Ingeniería de Conocimiento y Machine Learning*

- Muchas tareas requieren **grandes cantidades de datos de entrenamiento**, especialmente aquellas en las que los eventos a predecir son **poco frecuentes**.
- A menudo, los resultados de las máquinas sirven simplemente como **base para las decisiones**, que finalmente son **tomadas por los humanos**.
- Además, muchos casos empresariales y escenarios de la vida real exigen **conocimientos previos y explicaciones de los resultados** y el comportamiento.

***AAAI-MAKE 2021**: Combining Machine Learning and Knowledge Engineering
AAAI Spring Symposium on March 22-24, 2021, Stanford University, Palo Alto, California, USA.

Ingeniería de Conocimiento y Machine Learning*

- En medicina, por ejemplo, los médicos probablemente **no considerarán o anularán** las sugerencias si no hay una **explicación adecuada**.
- En el ámbito de la **conducción autónoma**, la seguridad y el control son fundamentales y exigen **enfoques simbólicos** que puedan **complementar** adecuadamente el aprendizaje automático.
- Los **agentes conversacionales** requieren conocimientos del **dominio e información contextual** para proporcionar respuestas satisfactorias.

***AAAI-MAKE 2021**: Combining Machine Learning and Knowledge Engineering
AAAI Spring Symposium on March 22-24, 2021, Stanford University, Palo Alto, California, USA.

Ingeniería de Conocimiento y Machine Learning*

- Además, las áreas de aplicación, como la banca, los seguros y las ciencias de la vida, están muy **reguladas** y exigen el **cumplimiento de leyes y normativas**.
- Este conocimiento específico de la aplicación debe **representarse y cumplirse estrictamente**, según el escenario de la aplicación, lo que constituye el ámbito de la **ingeniería del conocimiento**.

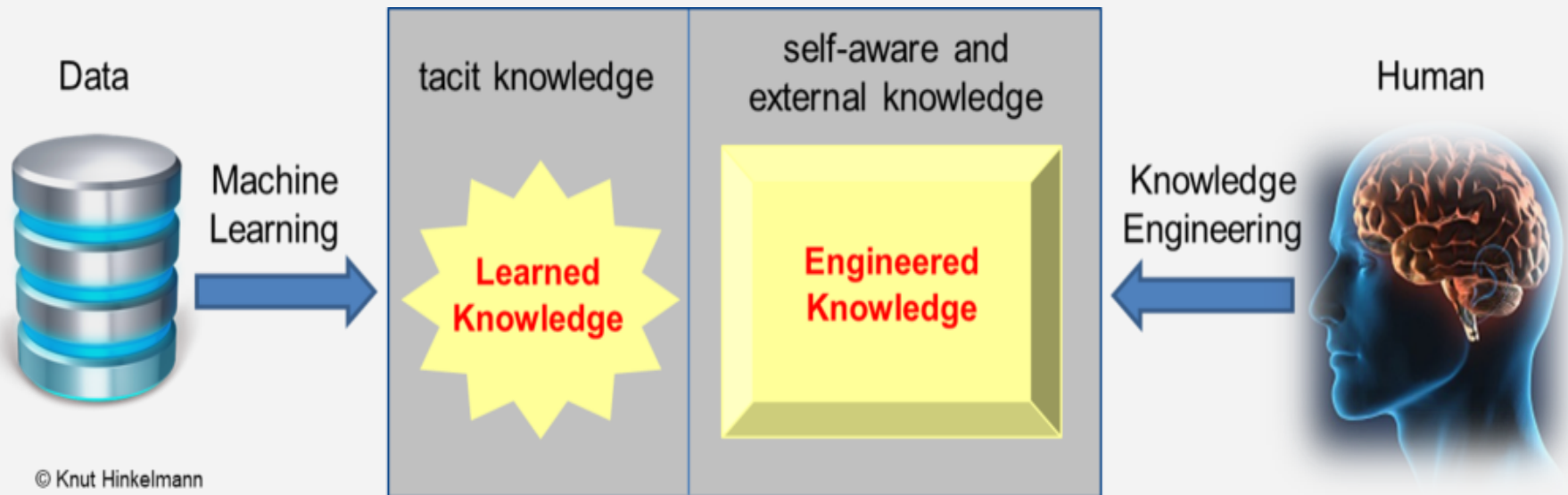
***AAAI-MAKE 2021**: Combining Machine Learning and Knowledge Engineering
AAAI Spring Symposium on March 22-24, 2021, Stanford University, Palo Alto, California, USA.

Ingeniería de Conocimiento y Machine Learning*

- La IC y los **sistemas basados en el conocimiento**, que hacen **explícito** y **accesible** el conocimiento de los expertos, suelen basarse en la **lógica** y pueden **explicar** sus conclusiones.
- Estos sistemas suelen requerir un **mayor esfuerzo inicial** durante su desarrollo que los sistemas que utilizan enfoques de aprendizaje automático.
- Sin embargo, los enfoques de **aprendizaje automático simbólico** y de **aprendizaje de ontologías** son prometedores para **reducir el esfuerzo** de la ingeniería del conocimiento.

***AAAI-MAKE 2021**: Combining Machine Learning and Knowledge Engineering
AAAI Spring Symposium on March 22-24, 2021, Stanford University, Palo Alto, California, USA.

Ingeniería de Conocimiento y Machine Learning*



***AAAI-MAKE 2021**: Combining Machine Learning and Knowledge Engineering
AAAI Spring Symposium on March 22-24, 2021, Stanford University, Palo Alto,
California, USA.

Ingeniería de Conocimiento y Machine Learning*

- Debido a sus puntos fuertes y débiles **complementarios**, existe una **demanda constante** en las empresas y entidades de **integrar** la ingeniería del conocimiento y el aprendizaje automático para **escenarios complejos**.
- Centrarse en un único aspecto no permite **aprovechar** todo el **potencial de la IA**.

***AAAI-MAKE 2021**: Combining Machine Learning and Knowledge Engineering
AAAI Spring Symposium on March 22-24, 2021, Stanford University, Palo Alto, California, USA.

Ingeniería de Conocimiento y Machine Learning*

- A pesar de los avances en los sistemas de **planificación y programación automatizados**, siguen necesitando alimentarse de **descripciones del dominio y del problema** cuidadosamente diseñadas y ajustadas para dominios y problemas **concretos**.
- La ingeniería del conocimiento para la planificación y la programación de la IA se ocupa de la **adquisición**, el **diseño**, la **validación** y el **mantenimiento** de los modelos de dominio, así como de la **selección** y la **optimización** de la maquinaria adecuada para trabajar con ellos.

*The International Competition on Knowledge Engineering for Planning and Scheduling, (KEPS'19, UC Berkeley)

Ingeniería de Conocimiento y Machine Learning*

- Estos procesos repercuten directamente en el **éxito** de las aplicaciones de planificación y programación del **mundo real**.
- La importancia de las técnicas de ingeniería del conocimiento queda claramente demostrada por la **diferencia de rendimiento** entre los planificadores **independientes del dominio** y los planificadores que explotan el conocimiento **dependiente del dominio**.

*The International Competition on Knowledge Engineering for Planning and Scheduling, (KEPS'19, UC Berkeley)

5. Análisis Inteligente de Datos guiado por IC

Un camino para combinar IC y ML...*

- Las tendencias actuales de la sociedad contrastan con la **percepción común de que la gente está cada vez más educada**. De hecho, los hechos demuestran que vivimos en una época en la que se observa:
 - Una “innumerabilidad” rampante (es decir, **ignorancia en temas relacionados con las matemáticas**, incluyendo conceptos muy básicos) en la población media, independientemente de la edad, el género y la clase social. Como ejemplo, véase este caso: "¿Por qué debemos pagar la misma cantidad por un tercio de libra de carne que por un cuarto de libra de carne en McDonald's? Nos están cobrando de más".

***Keynote de Doug Lenat**, CEO of Cycorp, Inc. **en el AAAI-MAKE 2021**, Stanford University, Palo Alto, California, USA.



Un camino para combinar IC y ML...*

- Las tendencias actuales de la sociedad contrastan con la **percepción común de que la gente está cada vez más educada**. De hecho, los hechos demuestran que vivimos en una época en la que se observa:
 - Una disminución general de la conexión con la realidad y el conocimiento, que termina en una ignorancia autojustificada de los hechos (¡porque puedes confiar en Google!).
 - La incredulidad ampliamente aceptada de que existe (y puede existir) una realidad objetiva real.
 - La suposición de que las respuestas a las preguntas están siempre disponibles...

*Inspirado en la **Keynote de Doug Lenat**, CEO of Cycorp, Inc. **en el AAAI-MAKE 2021**, Stanford University, Palo Alto, California, USA.

Un camino para combinar IC y ML...*

- Hoy en día, la gente confía en las máquinas para obtener conocimientos, mientras que los sistemas de inferencia actuales **no tienen la lógica necesaria** para reunir conocimientos de muchas fuentes diferentes y sacar conclusiones o respuestas.
- Por ejemplo, ¿es la catedral de Ciudad Real más alta que la torre del Micalet? La respuesta no se puede encontrar de forma completamente automática en Google, porque ninguna página web existente habla de ello explícitamente, aunque **las piezas individuales de la respuesta** están disponibles...
- (Q A Systems...)

*Inspirado en la **Keynote de Doug Lenat**, CEO of Cycorp, Inc. en el **AAAI-MAKE 2021**, Stanford University, Palo Alto, California, USA.

Un camino para combinar IC y ML...*

- Esto se debe a las dificultades del aprendizaje automático:
 - Situaciones en las que los expertos no están de acuerdo con los resultados
 - Riesgo de sesgo de confirmación
 - Respuestas absurdas o conclusiones intermedias
 - Poca cantidad de datos disponibles
 - Cuando la respuesta implica el "sentido común"
 - Cuando los mejores expertos son mucho mejores que la persona media

*Inspirado en la **Keynote de Doug Lenat**, CEO of Cycorp, Inc. **en el AAAI-MAKE 2021**, Stanford University, Palo Alto, California, USA.

Un camino para combinar IC y ML...*

- Hay cosas que el aprendizaje automático puede hacer sobre **textos generados por humanos** (desde las redes sociales al Quijote) para entender preguntas y respuestas, utilizando diferentes lógicas limitadas: desde la **lógica proposicional** hasta la **lógica completa de primer orden**.
- Pero la cantidad de preguntas que un humano puede formular y que podemos responder es **extremadamente limitada** (aún hoy en día) con respecto a las preguntas que cualquier humano que lea un texto o vea una representación argumental o una película puede responder. Y, sobre todo, hay una pregunta que la IA apenas puede responder: ¿**por qué?**

*Inspirado en la **Keynote de Doug Lenat**, CEO of Cycorp, Inc. en el **AAAI-MAKE 2021**, Stanford University, Palo Alto, California, USA.

Un camino para combinar IC y ML...*

- La mayor parte de las veces, la lógica de primer orden no es lo **suficientemente rápida a nivel computacional** (por no hablar de la lógica de segundo orden o de orden superior).
- La idea propuesta por Lenat consiste en dividir los **dos ingredientes importantes del razonamiento**, para hacer frente al **compromiso entre expresividad y eficiencia**:
 - aspectos epistemológicos
 - aspectos heurísticos

*Inspirado en la **Keynote de Doug Lenat**, CEO of Cycorp, Inc. **en el AAAI-MAKE 2021**, Stanford University, Palo Alto, California, USA.

Un camino para combinar IC y ML...*

- Al separar los dos, se pueden utilizar lenguajes extremadamente **eficientes** para la parte **heurística**, mientras se reservan lenguajes **muy potentes** (y menos eficientes) para la parte **epistemológica**.
- Tenemos **dos lenguajes en el juego**: el lenguaje de nivel heurístico (HL) y el lenguaje de nivel epistemológico (EL).

*Inspirado en la **Keynote de Doug Lenat**, CEO of Cycorp, Inc. **en el AAAI-MAKE 2021**, Stanford University, Palo Alto, California, USA.



Un camino para combinar IC y ML...*

- La otra aportación interesante al razonamiento es el **contexto**: el conocimiento **puede ser diferente en distintos contextos**.
- El sistema es capaz de utilizar la **ABDUCCIÓN** (o razonamiento abductivo) en contraste con la **DEDUCCIÓN** para proponer posibilidades probables en las respuestas, como por ejemplo, por qué ocurre algo en una historia.
- RAZONAMIENTO APROXIMADO: **Lógica Borrosa**.

*Inspirado en la **Keynote de Doug Lenat**, CEO of Cycorp, Inc. **en el AAAI-MAKE 2021**, Stanford University, Palo Alto, California, USA.



Un camino para combinar IC y ML...*

- Por lo tanto, el camino hacia la **combinación del aprendizaje automático y las bases de conocimiento** consiste en explotar su sinergia latente. En la práctica:
 - Usar las bases de conocimiento como **generadoras de** una gran cantidad de **datos de entrenamiento** para los sistemas de aprendizaje automático.
 - Utilizar el aprendizaje automático para **generar hipótesis** y, a continuación, utilizar las bases de conocimiento para **comprobar si hay errores o contradicciones**.

*Inspirado en la **Keynote de Doug Lenat**, CEO of Cycorp, Inc. **en el AAAI-MAKE 2021**, Stanford University, Palo Alto, California, USA.



Un camino para combinar IC y ML...*

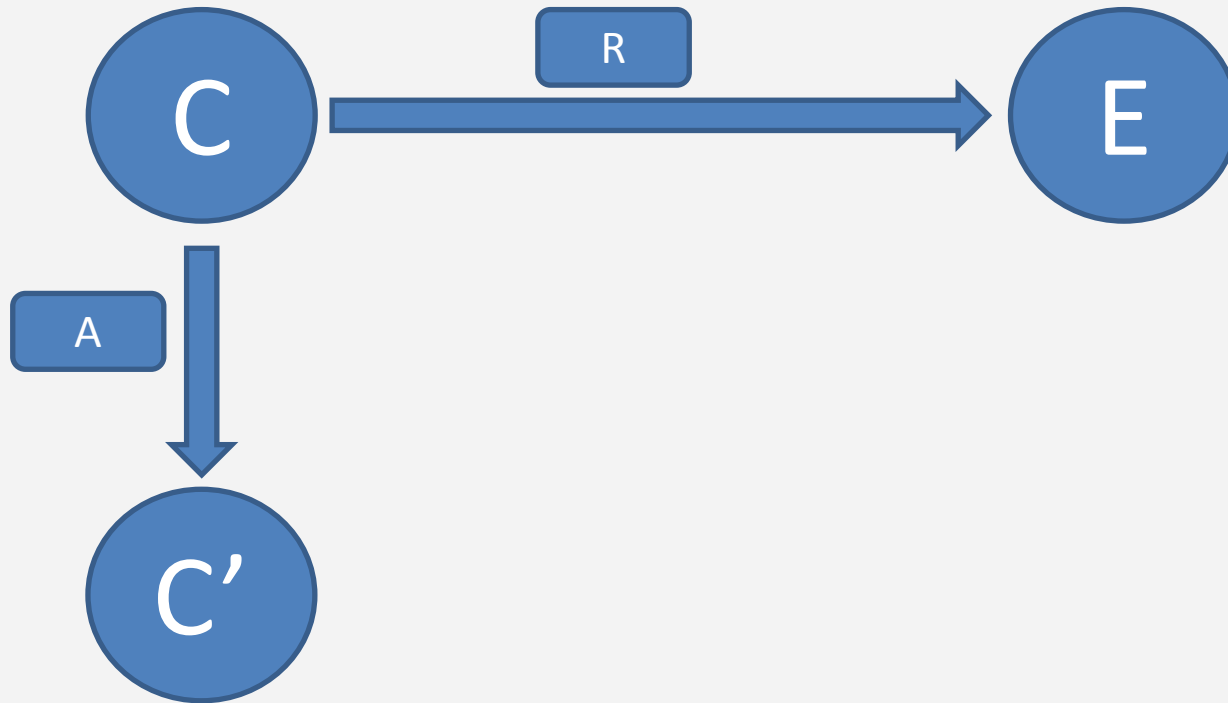
- Uno de los principales efectos a largo plazo es la posibilidad de trasladar el poder de análisis **de la correlación a la causalidad**.
- Esto puede aportar un enfoque **totalmente nuevo** a todo el campo.

*Inspirado en la **Keynote de Doug Lenat**, CEO of Cycorp, Inc. **en el AAAI-MAKE 2021**, Stanford University, Palo Alto, California, USA.

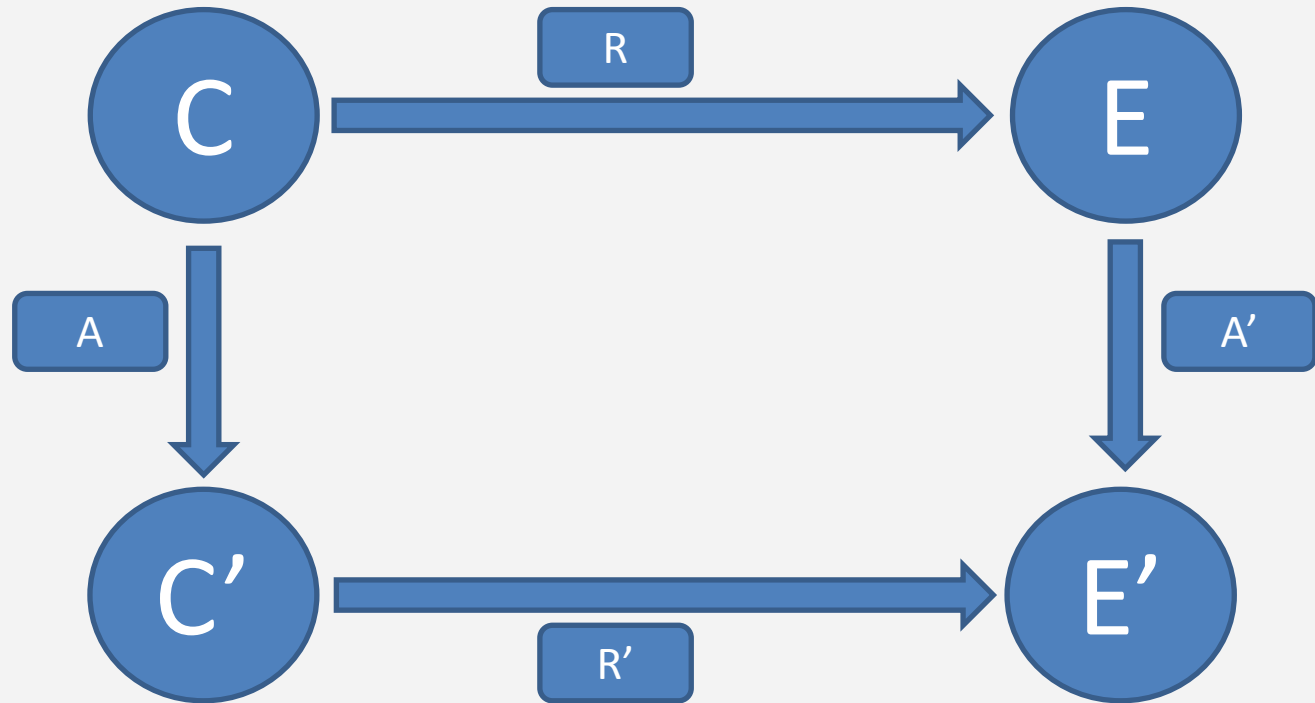
De la correlación a la causalidad...



De la correlación a la causalidad...



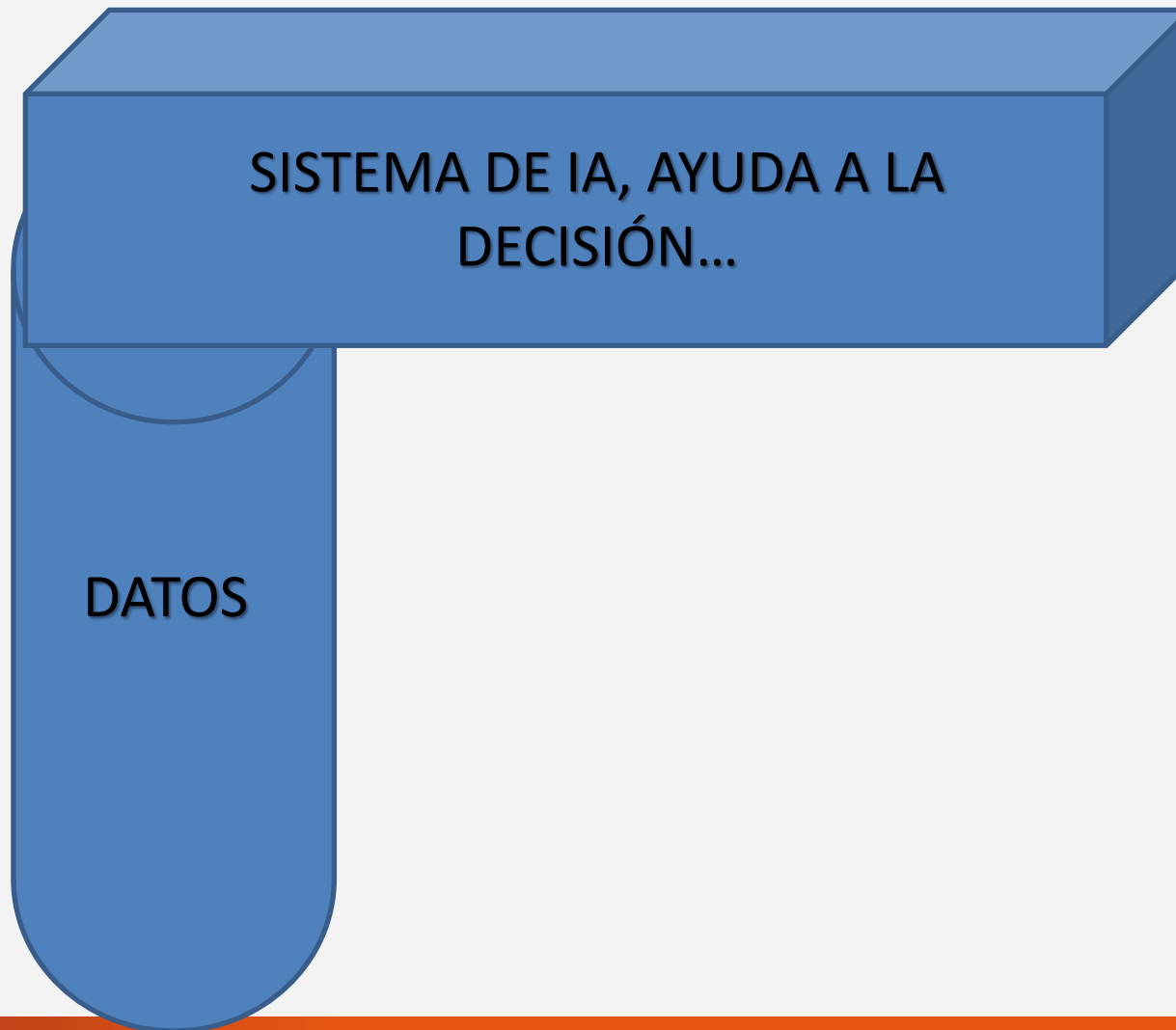
De la correlación a la causalidad...



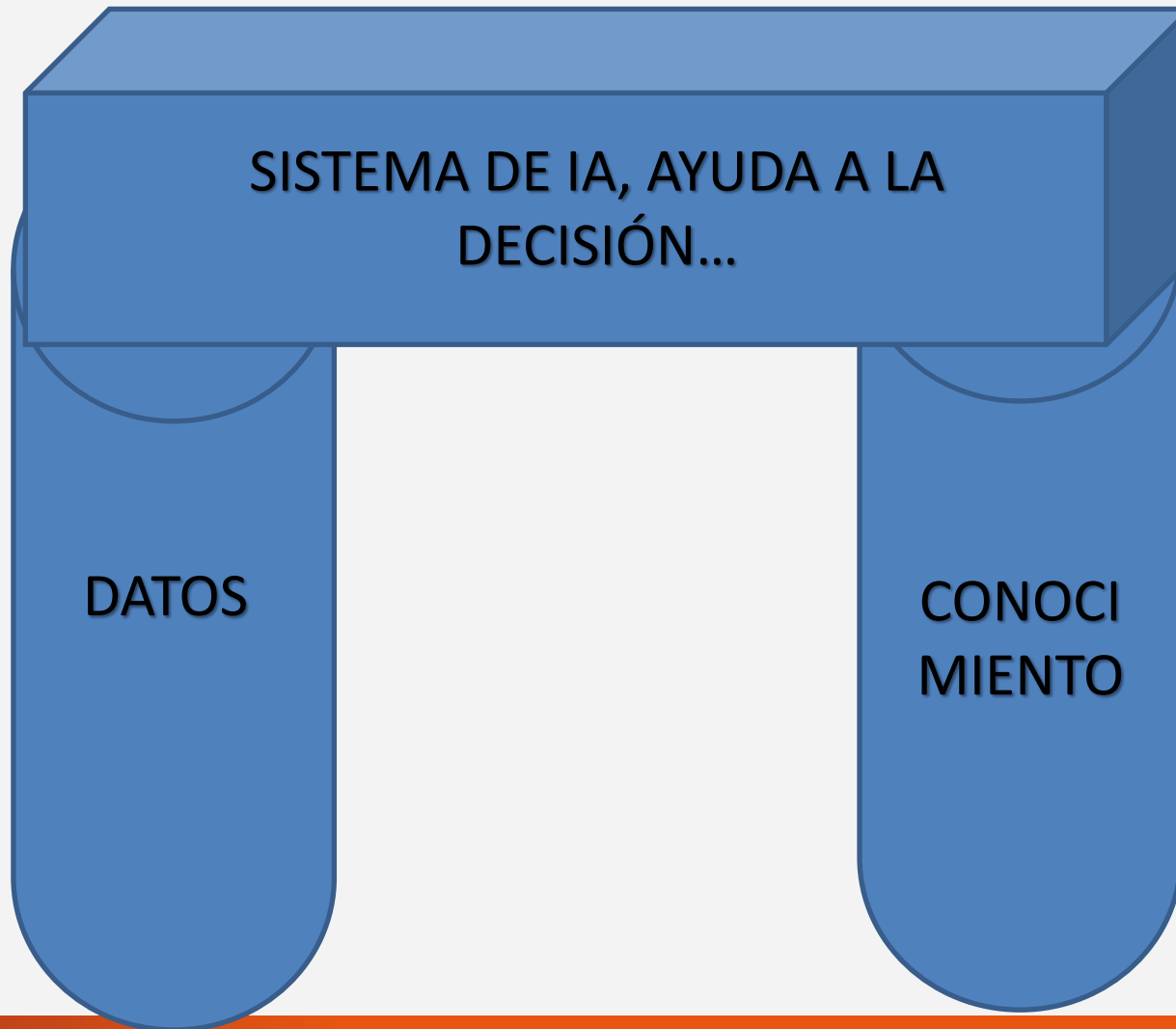
6. Resumen y reflexiones



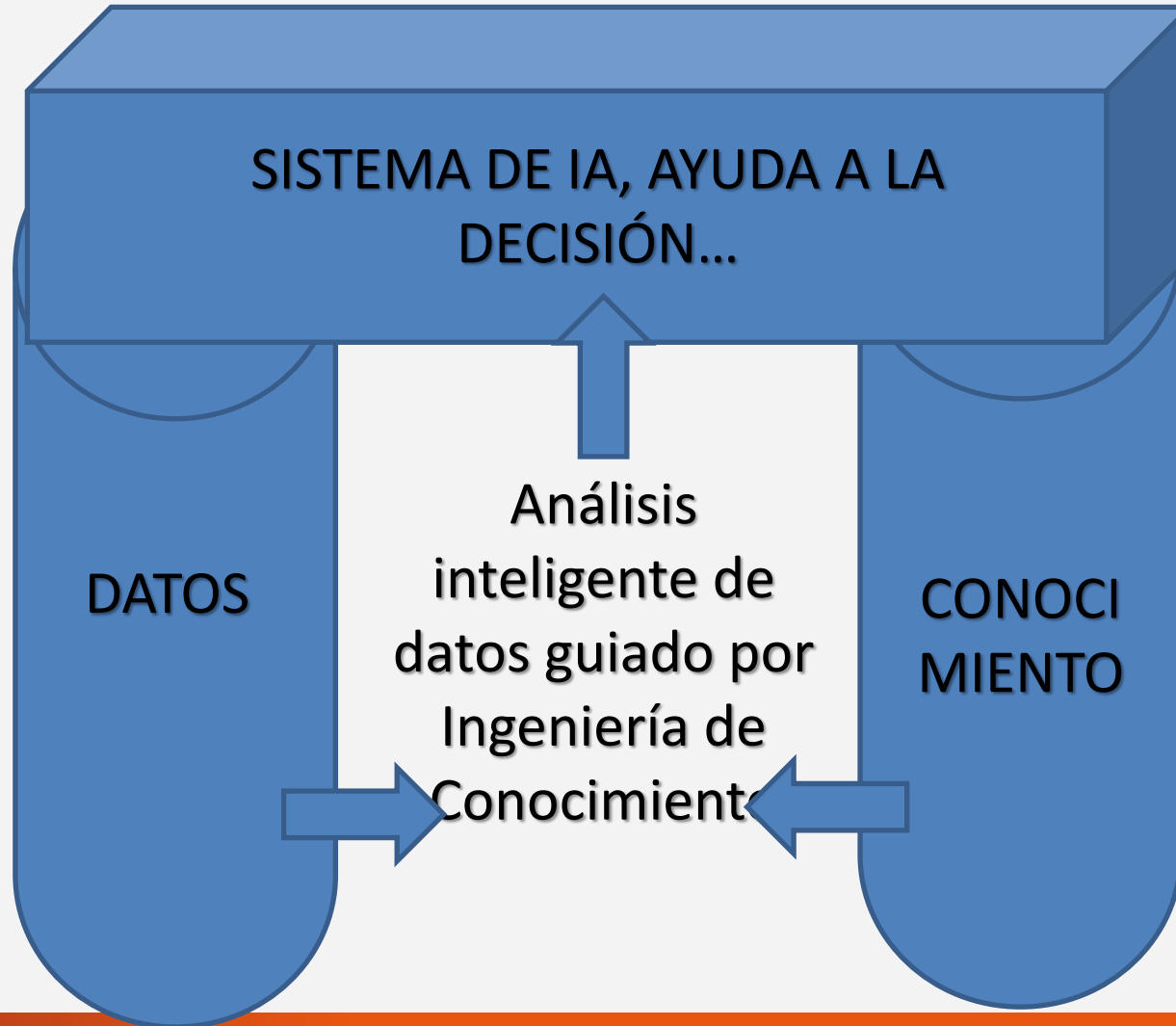
SISTEMA DE IA, AYUDA A LA
DECISIÓN...



Resumen y Reflexiones



Resumen y Reflexiones



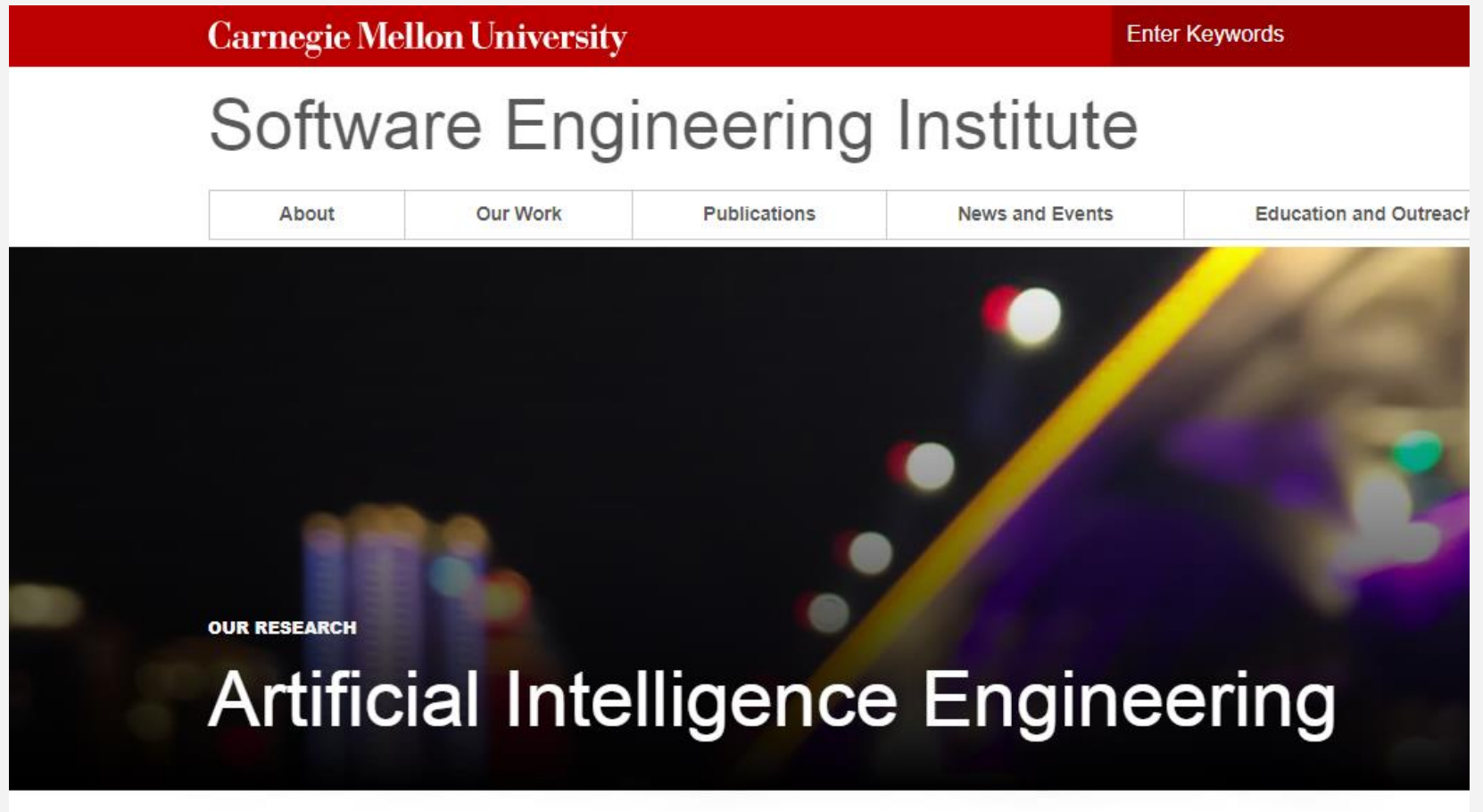
Análisis guiado por IC

- Representación del conocimiento y razonamiento
- IA híbrida
- IA explicable
- IA conversacional
- Sistemas basados en reglas
- Sistemas de recomendación
- Sistemas de interpretación de escenas
- Ontologías y web semántica
- Ciencia de los datos
- Sistemas expertos
- Integración e interoperabilidad
- Adquisición del conocimiento
- Representación del conocimiento

Análisis guiado por IC

- Formulación de dominios y descripciones de problemas
- Métodos y herramientas para la adquisición de conocimientos de dominio
- Adquisición y refinamiento del conocimiento de control
- Lenguajes formales para la descripción de dominios
- Reutilización del conocimiento del dominio
- Formatos para la especificación de heurísticas, parámetros y conocimientos de control para los solucionadores
- Importación de conocimiento del dominio a partir de ontologías generales
- Reformulación automatizada de problemas
- Procesos automatizados de extracción de conocimiento
- Métodos de visualización de modelos de dominio y espacios de búsqueda

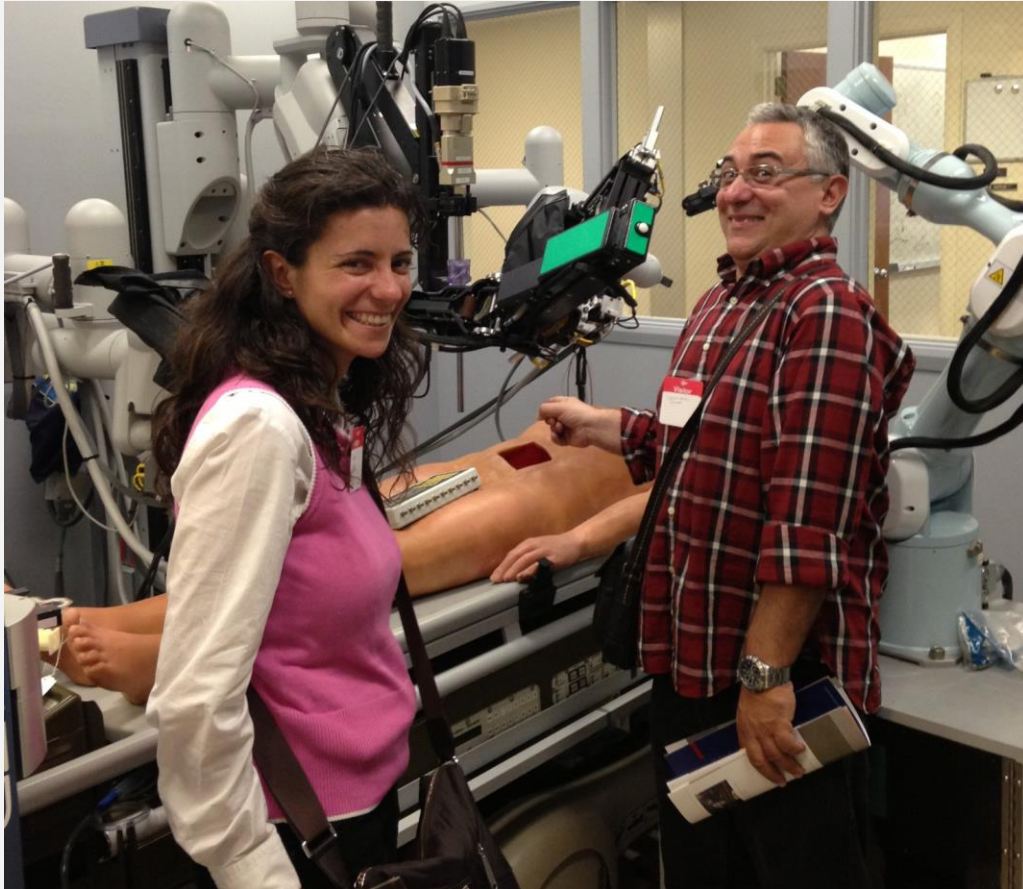
Análisis guiado por IC



Análisis guiado por IC

- Formación en las asignaturas de grado y posgrado:
 - Sistemas basados en conocimiento
 - Aprendizaje automático
 - Ingeniería de software
 - Lógica
 - Razonamiento aproximado
 - Programación declarativa
 - ...
 - [Mundo Hacker 2021 - Virtual Edition - El mayor evento de ciberseguridad en España \(mundohackerday.com\)](https://mundohackerday.com)
 - [Informe OBS - Inteligencia artificial, inteligencia computacional y análisis inteligente de datos](#)







GRACIAS

viu | **Universidad**
Internacional
de Valencia