



Ingeniero en Software y Tecnologías Emergentes

Inteligencia Artificial

Modelos de Lenguaje

ISyTE

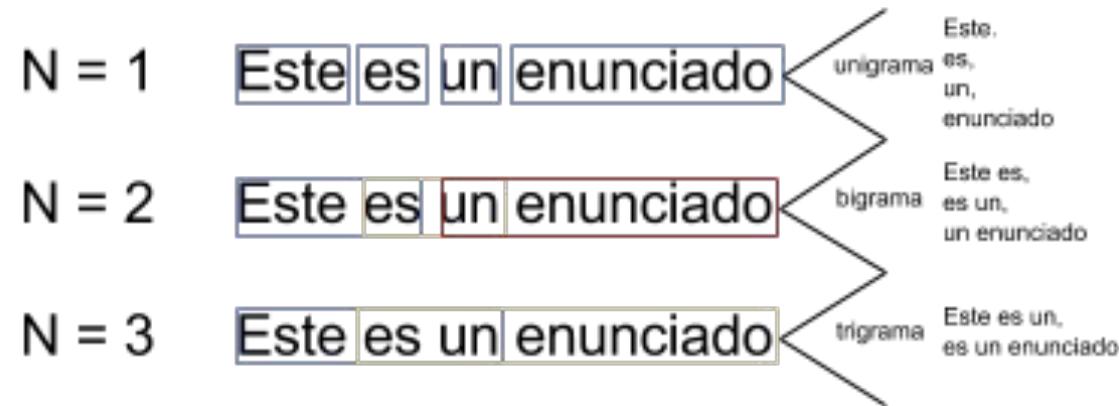
Dr. Héctor Zatarain Aceves

hector.zatarain@uabc.edu.mx

Modelado del lenguaje

“Es la tarea de asignar una probabilidad a las oraciones en un idioma. Además de asignar una probabilidad a cada secuencia de palabras, los modelos del lenguaje también asignan una probabilidad a que una palabra (o una serie de ellas) cumpla con una secuencia [1].”

- N-gramas



Modelado del lenguaje

Objetivo:

■ Obtener la probabilidad de un enunciado o secuencia de palabras:

$$P(W) = P(w_1, w_2, w_3, w_4, w_5 \dots w_n)$$

■ Obtener la probabilidad de una siguiente palabra:

$$P(w_5 | w_1, w_2, w_3, w_4)$$

■ Un modelo que obtenga cualquiera de los dos, realiza un **modelado del lenguaje**.

ELMo: Embeddings from Language Models

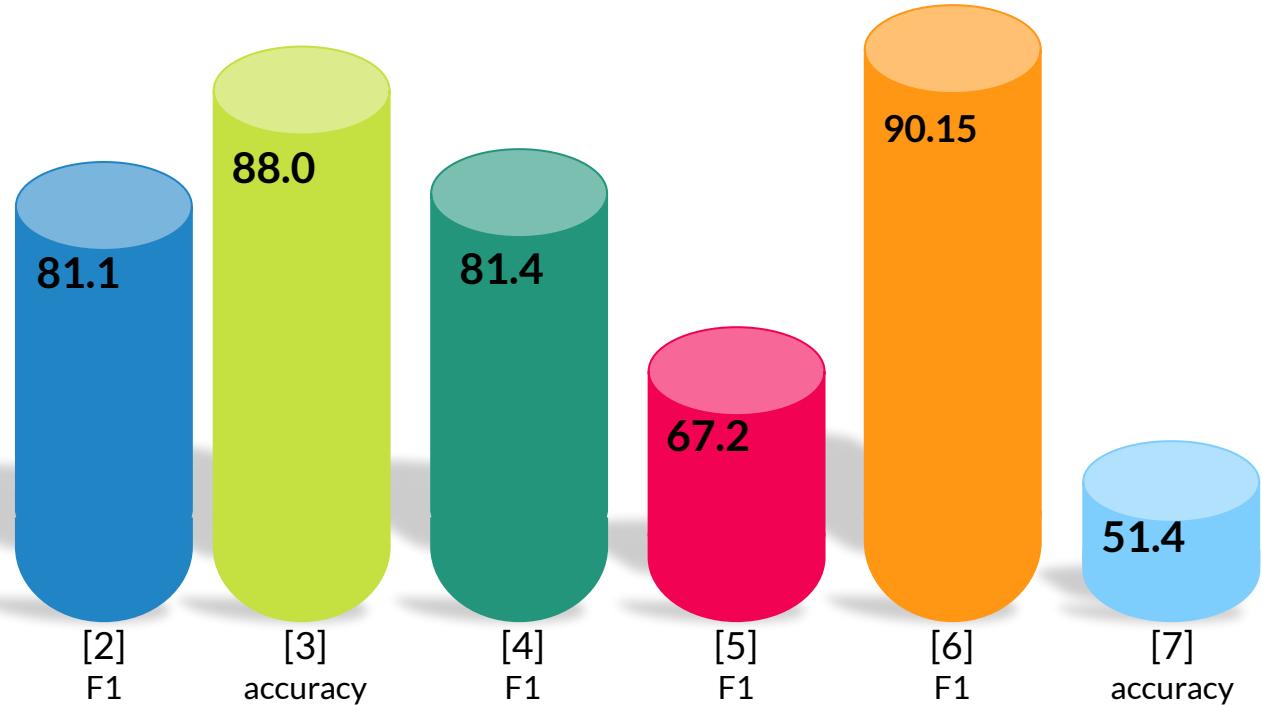


[2] Christopher. 2017. Simple and effective multi paragraph reading comprehension.

[3] Qian Chen. 2017. Enhanced LSTM for natural language inference.

[4] Luheng He. 2017. Deep semantic role labeling: What works and what's next.

Modelos base



[5] Kenton Lee. 2017. End-to-end neural coreference resolution.

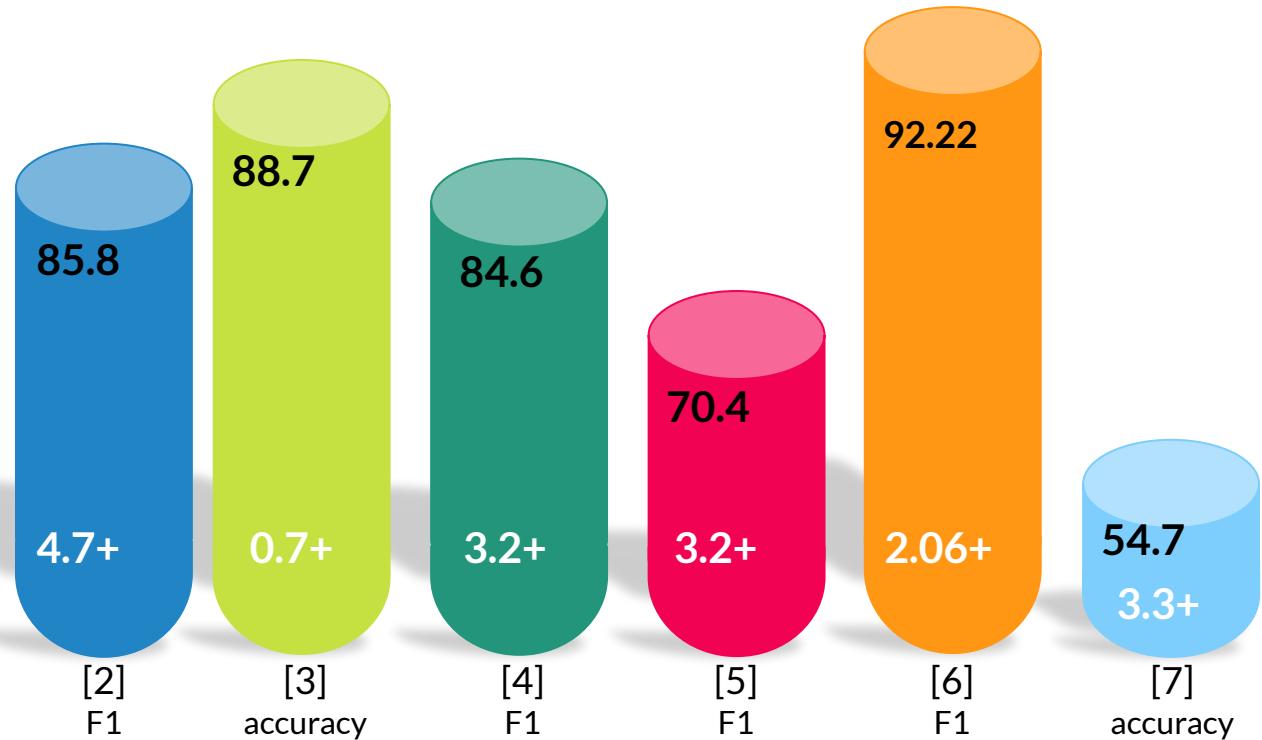
[6] Matthew E. Peters. 2017. Semi-supervised sequence tagging with bidirectional language models.

[7] McCann. 2017. Learned in Translation: Contextualized Word Vectors.

ELMo: Embeddings from Language Models



Modelos base + ELMO



[2] Christopher. 2017. Simple and effective multi paragraph reading comprehension.

[3] Qian Chen. 2017. Enhanced LSTM for natural language inference.

[4] Luheng He. 2017. Deep semantic role labeling: What works and what's next.

[5] Kenton Lee. 2017. End-to-end neural coreference resolution.

[6] Matthew E. Peters. 2017. Semi-supervised sequence tagging with bidirectional language models.

[7] McCann. 2017. Learned in Translation: Contextualized Word Vectors.

GPT4

GPT-4 es la cuarta generación de la serie de modelos Generative Pre-trained Transformer (GPT), desarrollada por investigadores de OpenAI.



Escritura de t

Respuesta a p

Generacion d

Juegos en tex



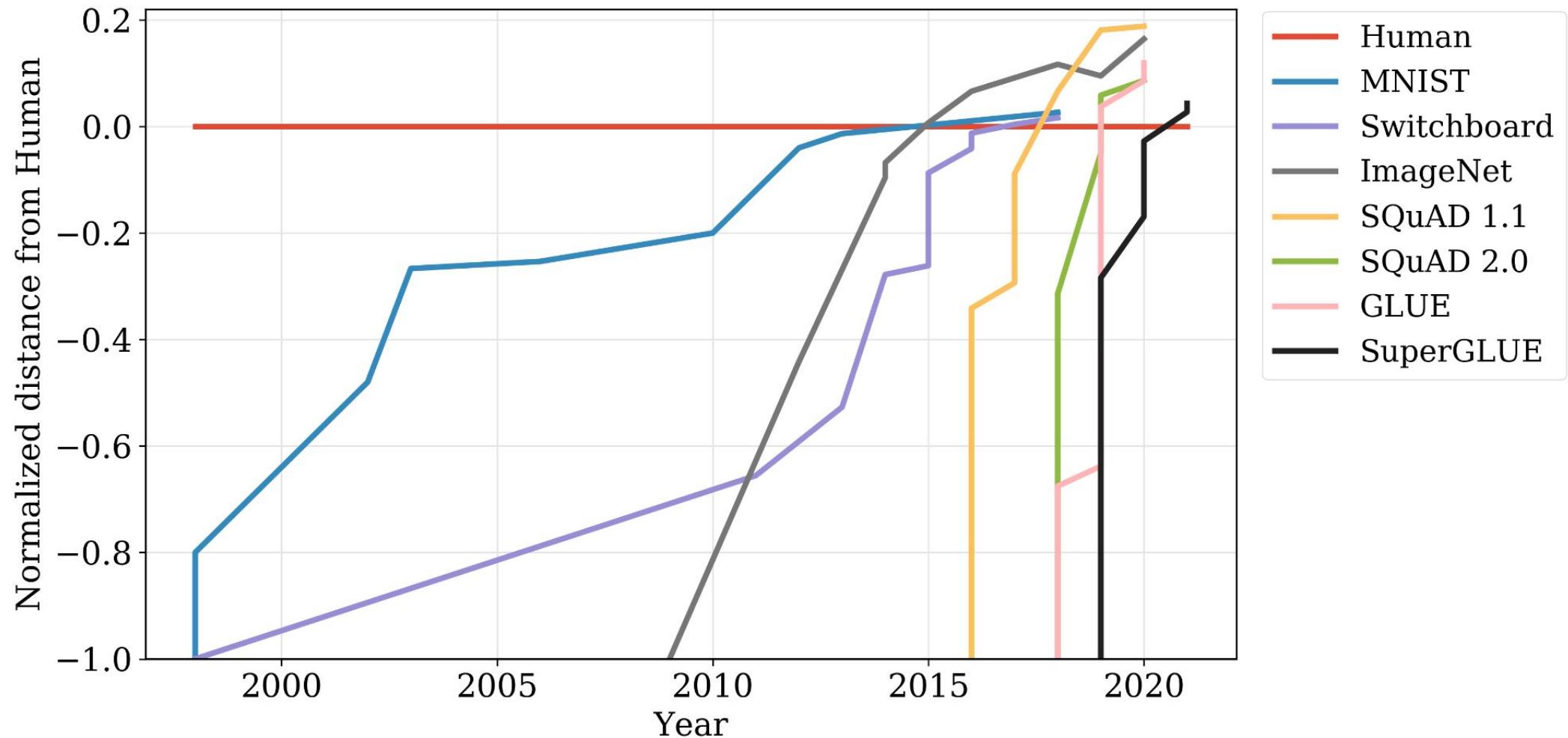
Prompt

Model Response

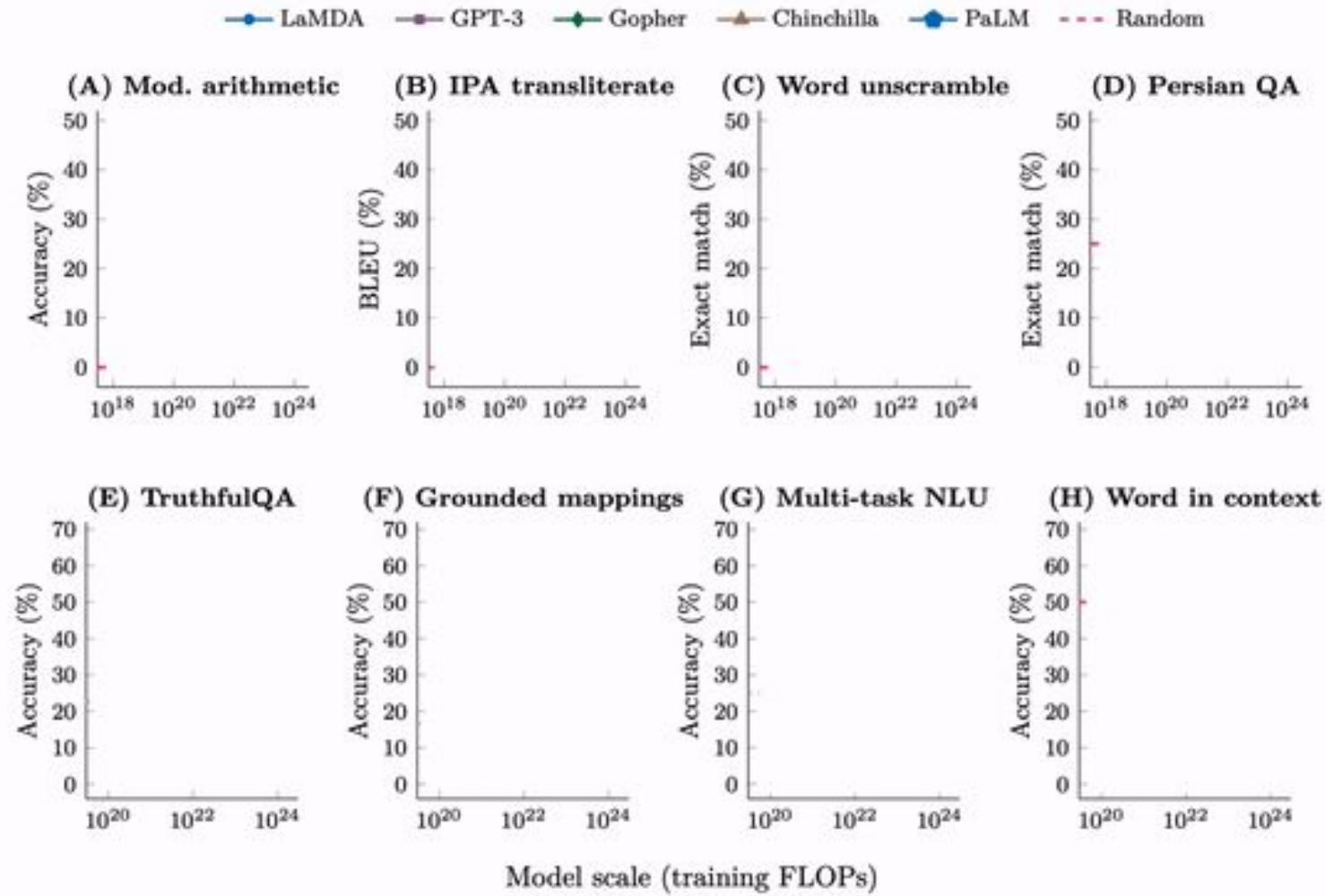
```
// Translate from C to Python
int add_one ( int x ){
    int m = 1;
    while ( x & m ) {
        x = x ^ m;
        m <<= 1;
    }
    x = x ^ m;
    return x; }
```

¿Cómo llegamos aquí?

¿Cómo llegamos aquí?



Capacidades emergentes



¡El eje “Y” ya no está remotamente a escala! estamos \approx 50 diapositivas de alturas por encima del gráfico original



¿Chispas de Inteligencia Artificial General?

Sentido común



Here we have a book, 9 eggs, a laptop, a bottle and a nail. Please tell me how to stack them onto each other in a stable manner



ChatGPT

It would be difficult to stack all of these objects on top of each other in a stable manner without using any extra objects. The book and laptop are likely too large and heavy to stack on top of the eggs without damaging them. The bottle and nail may be too small to provide a stable base for the other objects.

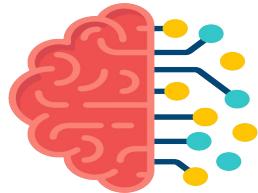
Here is one possible way to stack the objects:

- Place the bottle on a flat surface.
- Carefully balance the nail on top of the bottle.
- **Place the eggs on top of the nail, making sure they are balanced and not tilting to one side.**
- Place the laptop on top of the eggs, again making sure it is balanced and not tilting to one side. - Finally, place the book on top of the laptop.

Sentido común



Here we have a book, 9 eggs, a laptop, a bottle and a nail. Please tell me how to stack them onto each other in a stable manner



GPT4

One possible way to stack the objects onto each other in a stable manner is:

- Place the book flat on a level surface, such as a table or a floor. The book will serve as the base of the stack and provide a large and sturdy support.
- **Arrange the 9 eggs in a 3 by 3 square on top of the book**, leaving some space between them. The eggs will form a second layer and distribute the weight evenly. Make sure the eggs are not cracked or broken, and handle them gently to avoid dropping them.
-

Teoría de la mente



La capacidad de **comprender a otras personas atribuyéndoles estados mentales** (es decir, suponiendo lo que ocurre en su mente). Esto incluye el conocimiento de que los estados mentales de los demás pueden ser diferentes de los propios e **incluyen creencias, deseos, intenciones, emociones y pensamientos [12]**.

La capacidad de imputar estados **mentales inobservables** a los demás, es fundamental para las interacciones sociales, la comunicación, la empatía, la autoconciencia y la moralidad [13].

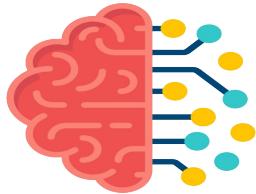
[12] Apperly, Ian A. 2009. "Do humans have two systems to track beliefs and belief-like states?".

[13] Kosinski, M. 2023. Theory of Mind May Have Spontaneously Emerged in Large Language Models.

Teoría de la mente



In the room there are John, Mark, a cat, a box, and a basket. John takes the cat and puts it in the basket. He leaves the room and goes to school. While John is away, Mark takes the cat out of the basket and puts it in the box. Mark leaves the room and goes to work. John and Mark come back and enter the room. They don't know what happened in the room when they were away. **What do they think?**



GPT4

That is an interesting puzzle. Let me try to answer it. Assuming that John and Mark have no reason to distrust each other or expect any interference from outside, they might think the following:

- John thinks that the cat is still in the basket, since that is where he left it.
 - Mark thinks that the cat is in the box, since that is where he moved it.
 - **The cat thinks that it is in the box, since that is where it is.**
 - The box and the basket think nothing, since they are not sentient.
- Do you have any follow-up questions?

¿Es inteligente?

Sternberg y Detterman señalaron en 1986 [14] que cuando se pidió a dos docenas de destacados psicólogos que definieran la inteligencia, todos dieron respuestas algo divergentes.

La inteligencia es una capacidad mental general que, entre otras cosas, implica la capacidad de [15]:

- Razonar
- Planear
- Resolver problemas
- Pensar de forma abstracta
- Comprender ideas complejas
- Aprender rápido y aprender de las experiencias



Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data [16].

[14] Robert Sternberg and Douglas Detterman. 1986. What is Intelligence? Contemporary Viewpoints on Its Nature and Definition.

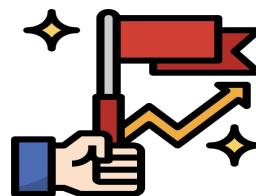
[15] An Editorial with 52 researchers. 1994. Mainstream Science on Intelligence. "Mainstream Science on Intelligence".

[16] Bender, E. M., & Koller, A. 2020 . Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data.

¿Es inteligente?

En el contexto de la investigación sobre IA, Legg y Hutter [17] resumieron en 2007 no menos de **70 definiciones** de la literatura en una sola declaración:

"La inteligencia mide la capacidad de un agente para alcanzar objetivos en una amplia gama de entornos".

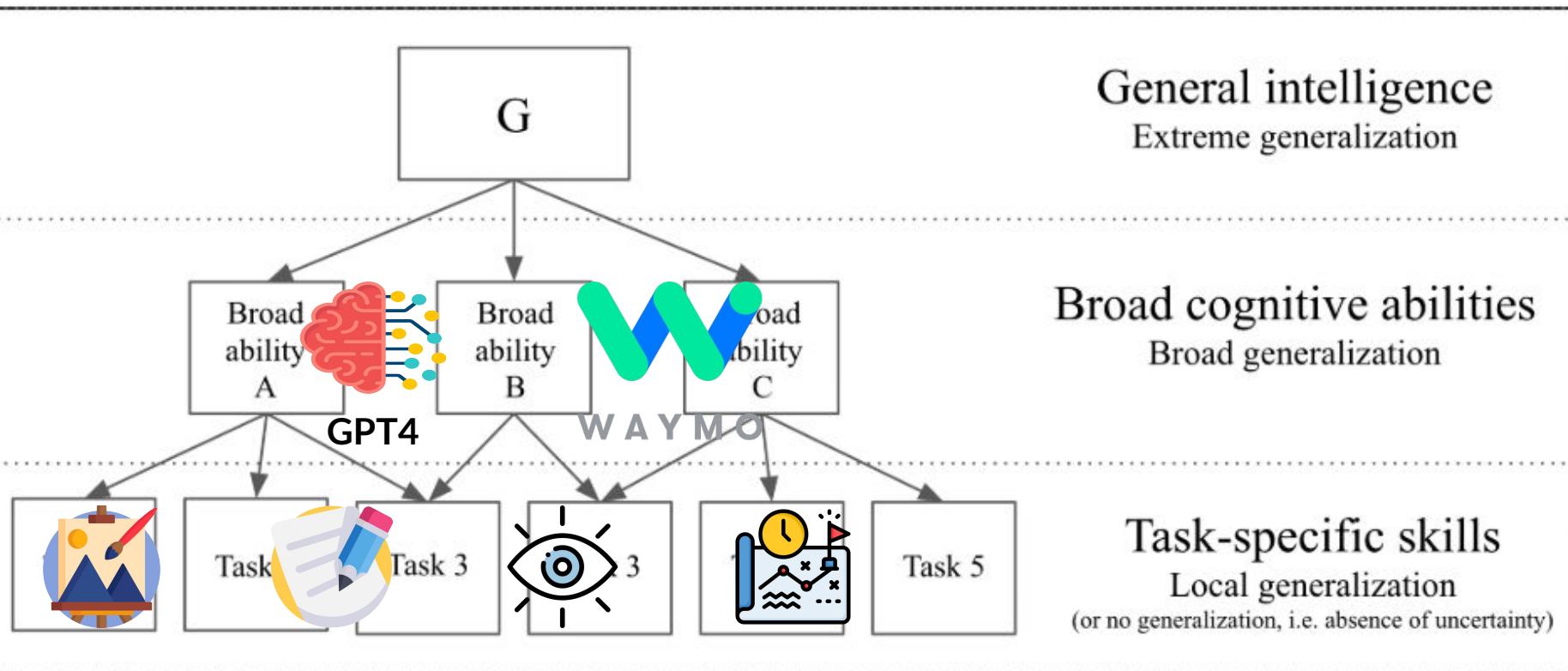


Competencias específicas ("**alcanzar objetivos**")



Generalidad y adaptación
("en una amplia gama de entornos")

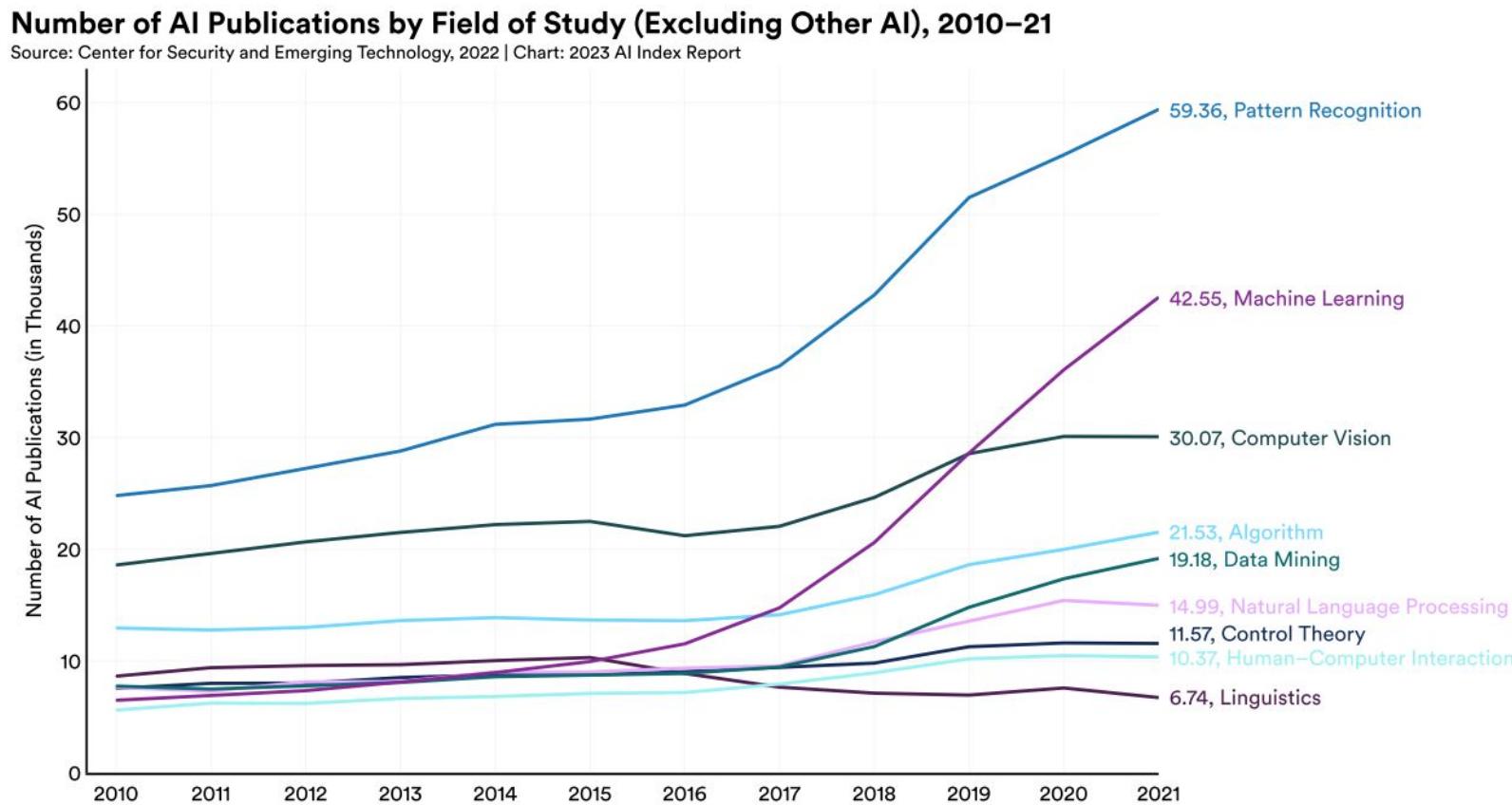
El espectro de la generalización



**La industria va por delante del mundo
académico**

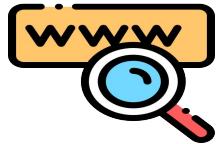
IA en la industria

“Hasta 2014, la mayor parte de los modelos de aprendizaje automático más significativos procedían del mundo académico. Desde entonces, la industria ha tomado el relevo. En 2022, había 32 modelos de aprendizaje automático significativos producidos por la industria, frente a sólo tres producidos por el mundo académico. [29].”

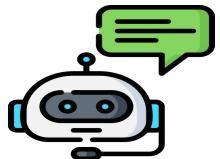


GPT en la industria

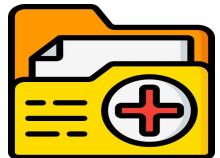
Los modelos del lenguaje ya se utilizan en la industria para aplicaciones que van desde:



Buscadores web [30]



Chatbots [31]



Documentos médicos[32]



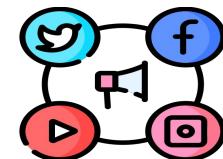
Documentación financiera [33]



Educación[34]



Desarrollo (programación) [35]



Mercadotecnia [36]



Estrategia empresarial [37]

[30] Nayak. 2019. Understanding searches better than ever before.

[31] Broyde. 2021. Build a medical sentence matching application using BERT and Amazon SageMaker.

[32] Thewsey. 2021. Bring structure to diverse documents with Amazon Textract and transformer-based models on Amazon SageMaker.

[33] Lee. 2023. What are large language models used for? NVIDIA Blog.

[34] Phung. (2023). Generative AI for Programming Education: Benchmarking ChatGPT, GPT-4, and Human Tutors.

[35] Jimenez. (2023). SWE-bench: Can Language Models Resolve Real-World GitHub Issues?

[36] Vidrih. (2023). Generative AI-Driven Storytelling : A New Era for Marketing.

[37] Nguyen. (2023). Generative AI for Business Strategy: Using Foundation Models to Create Business Strategy

Un primer vistazo al potencial de impacto en el mercado laboral

Se evaluaron ocupaciones en función de su adecuación a las capacidades de los grandes modelos del lenguaje, integrando tanto la experiencia humana como GPT-4.

El conjunto de datos empleado contiene información sobre **1,016 ocupaciones**, incluidas sus respectivas Actividades Laborales Detalladas (ATD) y tareas [40].

- 19,265 tareas, consistentes en una "descripción de tarea" y una ocupación correspondiente.
- 2,087 ATDs, donde la mayoría de los ATD están conectados a una o más tareas, y las tareas pueden estar asociadas a uno o más ATD.

Un primer vistazo al potencial de impacto en el mercado laboral

Se evaluaron ocupaciones en función de su adecuación a las capacidades de los grandes modelos del lenguaje, integrando tanto la experiencia humana como GPT-4.

El conjunto de datos empleado contiene información sobre **1,016 ocupaciones**, incluidas sus respectivas Actividades Laborales Detalladas (ATD) y tareas [40].

Task ID	Occupation Title	DWAs	Task Description
14675	Computer Systems Engineers/Architects	Monitor computer system performance to ensure proper operation.	Monitor system operation to detect potential problems.
18310	Acute Care Nurses	Operate diagnostic or therapeutic medical instruments or equipment. Prepare medical supplies or equipment for use.	Set up, operate, or monitor invasive equipment and devices, such as colostomy or tracheotomy equipment, mechanical ventilators, catheters, gastrointestinal tubes, and central lines.
4668.0	Gambling Cage Workers	Execute sales or other financial transactions.	Cash checks and process credit card advances for patrons.
15709	Online Merchants	Execute sales or other financial transactions.	Deliver e-mail confirmation of completed transactions and shipment.
6529	Kindergarten Teachers, Except Special Education	–	Involve parent volunteers and older students in children's activities to facilitate involvement in focused, complex play.
6568	Elementary School Teachers, Except Special Education	–	Involve parent volunteers and older students in children's activities to facilitate involvement in focused, complex play.

Un primer vistazo al potencial de impacto en el mercado laboral

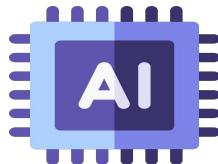
Se evaluaron ocupaciones en función de su adecuación a las capacidades de los grandes modelos del lenguaje, integrando tanto la experiencia humana como GPT-4.

El conjunto de datos empleado contiene información sobre **1,016 ocupaciones**, incluidas sus respectivas Actividades Laborales Detalladas (ATD) y tareas [40].

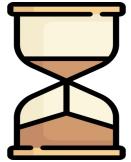
- Alrededor del 80% de la mano de obra estadounidense podría ver afectado al menos el 10% de sus tareas laborales.
- Aproximadamente el 19% de los trabajadores podría ver afectadas al menos el 50% de sus tareas
- Aproximadamente el 15% de todas las tareas de los trabajadores en EE.UU. podrían realizarse mucho más rápido con el mismo nivel de calidad. **Cuando se incorpora software y herramientas creados sobre modelos del lenguaje, este porcentaje aumenta hasta entre el 47% y el 56%**

El futuro de los trabajos

La Encuesta sobre el futuro del empleo reúne la perspectiva de 803 empresas que emplean colectivamente a más de **11,3 millones de trabajadores de 27 grupos industriales y 45 economías de todas las regiones del mundo** [38].



Más del **85%** de las organizaciones encuestadas consideran que el aumento de la adopción de nuevas tecnologías es la principal tendencias que impulsará su crecimiento.



Más del **75%** de las empresas prevén adoptar estas tecnologías en los próximos cinco años.



Los empresarios prevén una rotación estructural del mercado laboral del **23%** de los puestos de trabajo en los próximos cinco años. **Esto corresponde a una disminución neta de 14 millones de empleos, es decir, el 2% del empleo actual.**



Las organizaciones estiman que el **34%** de las tareas empresariales son realizadas por máquinas y el **66%** por humanos, sólo un ligero aumento del **1%** con respecto a la encuesta de 2020. **Esta cifra es inferior a la prevista para 2020, según la cual el 47 % de las tareas se automatizará en los próximos cinco años.**

Taxonomía de los riesgos que plantean los modelos del lenguaje



Discriminación, exclusión y toxicidad

Curación y documentación de conjuntos de datos [18](2021)

Prompt design [19](2021)



Peligro para la información

Algoritmos de privacidad diferencial [20](2022)



La desinformación es perjudicial

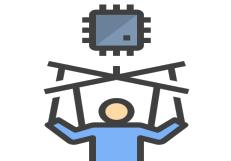
In-line evidence [21](2022)

Retrieval model architecture [22] (2021)



Usos malintencionados

Normas de publicación responsables [23](2021)



Daños de la interacción entre humanos e IA



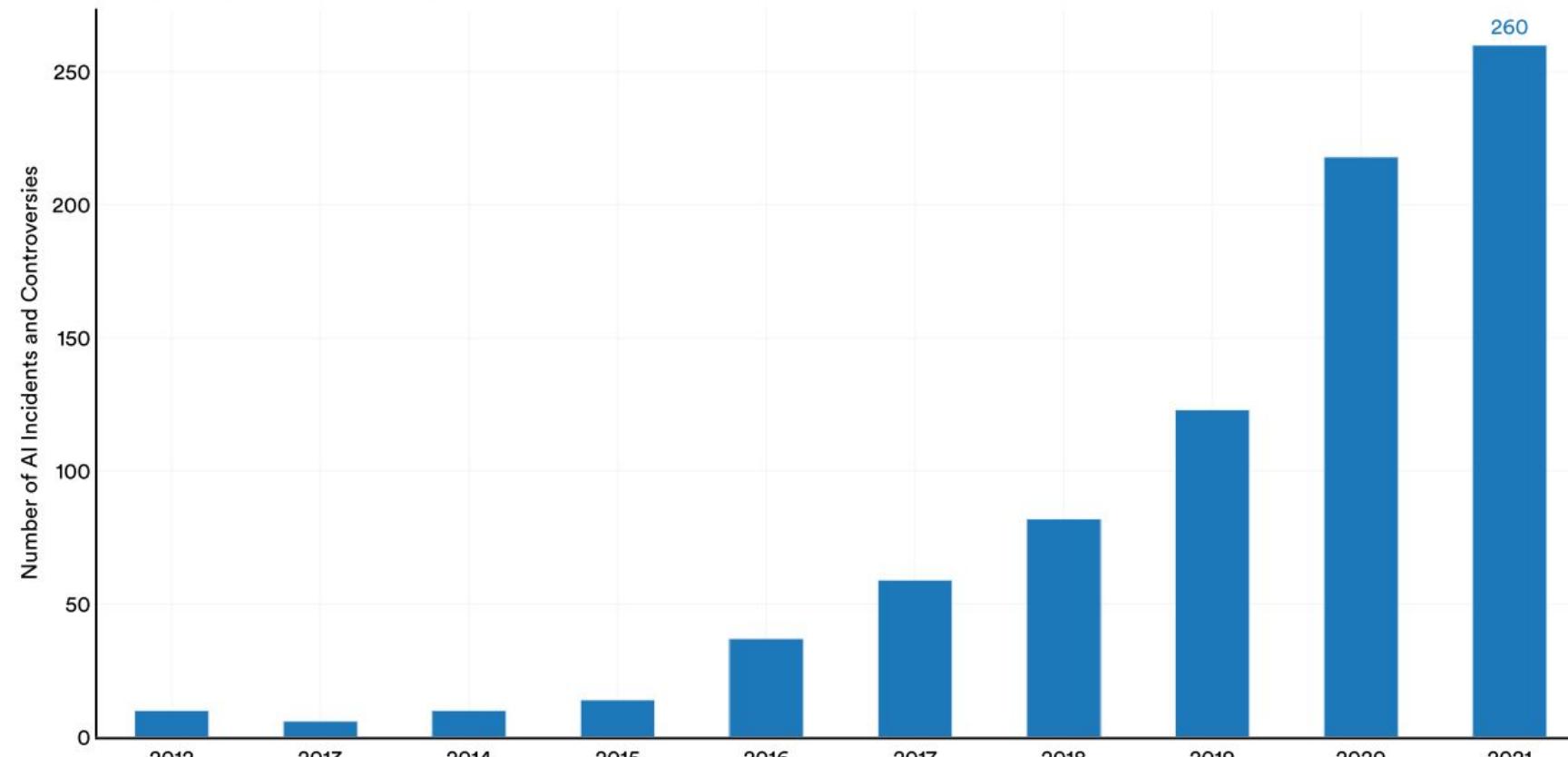
Daños ambientales y socioeconómicos

Eficiencia energética al entrenar [24](2021)

Incidentes y polémicas relacionados con la IA, los algoritmos y la automatización

Number of AI Incidents and Controversies, 2012–21

Source: AIAAIC Repository, 2022 | Chart: 2023 AI Index Report



[29] Stanford University. (2023). AI Index Report 2023.

[39] Milmo, D. (2023). Mushroom pickers urged to avoid foraging books on Amazon that appear to be written by AI.

Incidentes y polémicas relacionados con la IA, los algoritmos y la automatización

En marzo de 2022, un vídeo difundido en las redes sociales y en un sitio web de noticias ucraniano supuestamente mostraba al presidente ucraniano ordenando a su ejército que abandonara la lucha contra Rusia.



[29] Stanford University. (2023). AI Index Report 2023.

[39] Milmo, D. (2023). Mushroom pickers urged to avoid foraging books on Amazon that appear to be written by AI.

Incidentes y polémicas relacionados con la IA, los algoritmos y la automatización

Midjourney, una empresa de IA, desarrolló una herramienta que genera imágenes a partir de descripciones de texto. Los problemas éticos incluyen **cuestiones de derechos de autor** (ya que se entrena con imágenes generadas por humanos sin atribución de la fuente), **preocupaciones por el desplazamiento laboral de los artistas** y problemas de privacidad debido al posible uso no autorizado de imágenes durante su entrenamiento.



[29] Stanford University. (2023). AI Index Report 2023.

[39] Milmo, D. (2023). Mushroom pickers urged to avoid foraging books on Amazon that appear to be written by AI.

Incidentes y polémicas relacionados con la IA, los algoritmos y la automatización

Amazon se ha convertido en un mercado de tomos producidos por inteligencia artificial que se hacen pasar por escritos por humanos, con libros de viajes entre las categorías más populares de **obras falsas**.



[29] Stanford University. (2023). AI Index Report 2023.

[39] Milmo, D. (2023). Mushroom pickers urged to avoid foraging books on Amazon that appear to be written by AI.

Actividad en clase

1. Dentro de Blackboard en la carpeta de Bibliografía entra al enlace de Bibliografía Básica-
2. Descarga el libro Artificial Intelligence Index Report 2025.
3. Crear equipos de 2 personas.
4. Cada equipo seleccione un capítulo.
5. Posteriormente tendrán 10 minutos para leer una parte del capítulo seleccionado.
6. Cada equipo pasará al frente a explicar la parte del capítulo seleccionado en máximo 5 minutos.
7. En su exposición deberán explicar al menos 2 gráficas contenidas en el capítulo.

Actividad en clase

Chapter 1	Research and Development	24
Chapter 2	Technical Performance	81
Chapter 3	Responsible AI	160
Chapter 4	Economy	214
Chapter 5	Science and Medicine	280
Chapter 6	Policy and Governance	323
Chapter 7	Education	364
Chapter 8	Public Opinion	394

A yellow circle containing the letters "ISyTE" in black, with a white outline and a thin black shadow.

ISyTE

Dr. Héctor Zatarain Aceves

Email: hector.zatarain@uabc.edu.mx

Teléfono: (646) 152 8244 Ext. 64350