

EOI Santiago de Compostela

IA para la práctica y evaluación de idiomas: pasado, presente y futuro

Dr. Mariano Felice



Crédito: A. J. O'Brien/Fox Photos/
Hulton Archive/Getty Images



1° Revolución Industrial (1760-1840)

Máquina de vapor



2° Revolución Industrial (1870-1914)

Electricidad y cadena de montaje



3° Revolución Industrial (1969-2000)

Tecnología de la información



4° Revolución Industrial (2000-presente)

Inteligencia artificial y biotecnología

¿Qué es la inteligencia artificial?



Alan Turing

La inteligencia artificial es la capacidad de los ordenadores de realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, el razonamiento, la resolución de problemas y la percepción.

¿Qué es la inteligencia artificial?



Alan Turing

La inteligencia artificial es la capacidad de los ordenadores de realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, el razonamiento, la resolución de problemas y la percepción.

Aprendizaje
Automático

Procesamiento de
Lenguaje Natural

Visión Artificial

Robótica

¿Qué es el Procesamiento de Lenguaje Natural?

El Procesamiento de Lenguaje Natural es el subcampo de la Inteligencia Artificial que se encarga de interpretar, comprender y producir lenguaje humano (texto y voz).

¿Qué puede hacer el PLN por nosotros?

Detección de plagio	Análisis de vocabulario	Corrección de errores gramaticales	Pronunciación	Inteligibilidad	Diálogo
Argumentación	Calificación Automática de Textos	Evaluación de la escritura	Evaluación del habla	Cumplimiento de la tarea	Fluidez
Registro y estilo	Relevancia	Coherencia y cohesión	Mediación	Competencias comunicativas	Registro y tono
Análisis sintáctico y etiquetado	Adaptación de textos	Comprensión lectora	Comprensión auditiva	Generación automática de contenido	Reconocimiento automático del habla
Generación de textos	Estimación de dificultad de ítems	Exámenes adaptativos	Localización	Generación de feedback	Síntesis de habla

Pasado

Análisis lingüístico automático

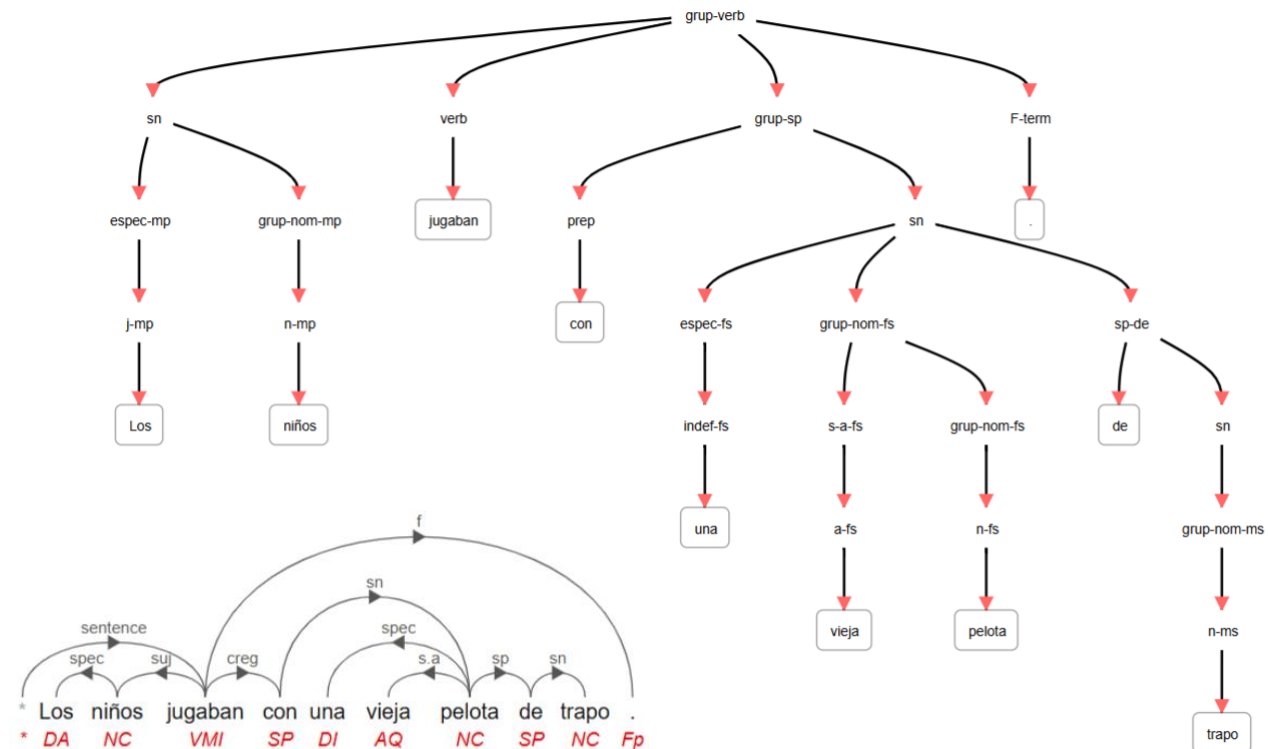
Tokenización

Los niños jugaban con una
vieja pelota de trapo .

Etiquetado morfosintáctico

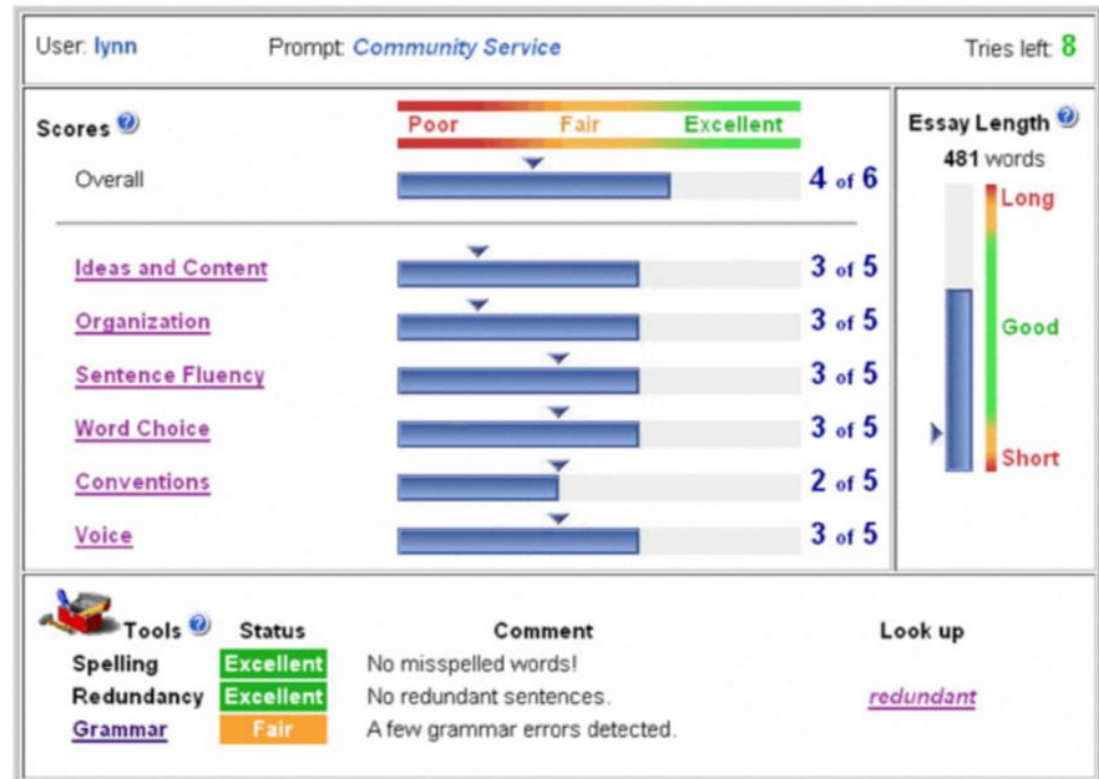
Los niños jugaban
DAOMPO NCMP000 VMII3PO
con una vieja
SP DIOFS0 AQOFS00
pelota de trapo .
NCFS000 SP NCMS000 Fp

Análisis sintáctico



Los primeros sistemas de calificación automática

- ★ Project Essay Grade (PEG) (1966)
- ★ Intelligent Essay Assessor (IEA) (Pearson, 1997)



Intelligent Essay Assessor (IEA)

Los primeros sistemas de corrección gramatical

- ★ The Writer's Workbench (AT&T/Bell Laboratories, 1982)
- ★ Microsoft Word (1993)
- ★ ALEK (2000)
- ★ GRANSKA (2000)

Los primeros sistemas de corrección gramatical

- ★ The Writer's Workbench (AT&T/Bell Laboratories, 1982)
- ★ Microsoft Word (1993)
- ★ ALEK (2000)
- ★ GRANSKA (2000)

ALEK

Tipo: Error de concordancia de número.

Patrón: /DT_a NNS/

Regla: SI NO /DT_a NNS NN/ O
SI frecuencia(original) <
frecuencia(sugerencia)

Los primeros sistemas de corrección gramatical

- ★ The Writer's Workbench (AT&T/Bell Laboratories, 1982)
- ★ Microsoft Word (1993)
- ★ ALEK (2000)
- ★ GRANSKA (2000)

ALEK

Tipo: Error de concordancia de número.

Patrón: /DT_a NNS/

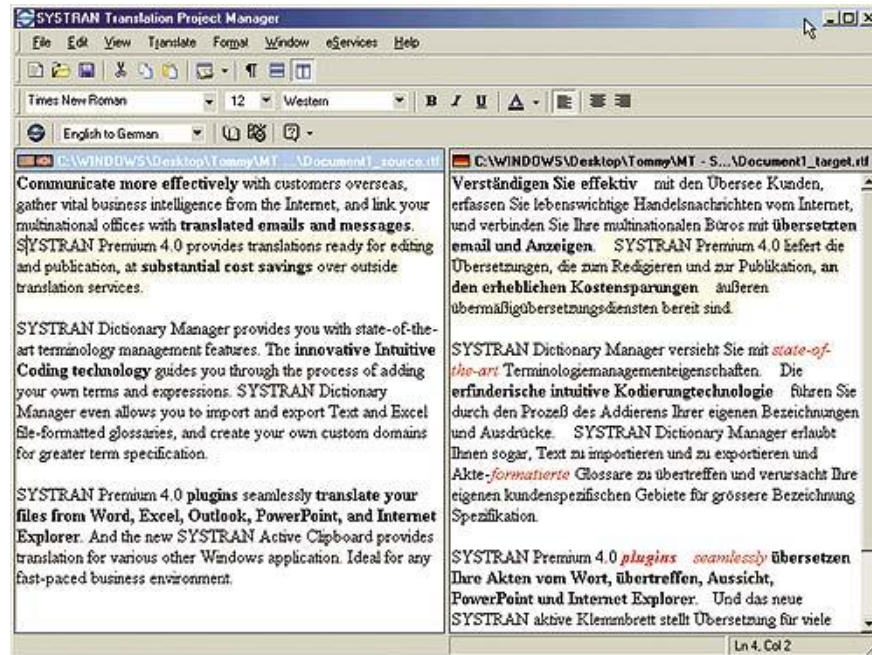
Regla: SI NO /DT_a NNS NN/ O
SI frecuencia(original) <
frecuencia(sugerencia)

They saw **a bears** in the forest.

The company hired **a systems analyst**.

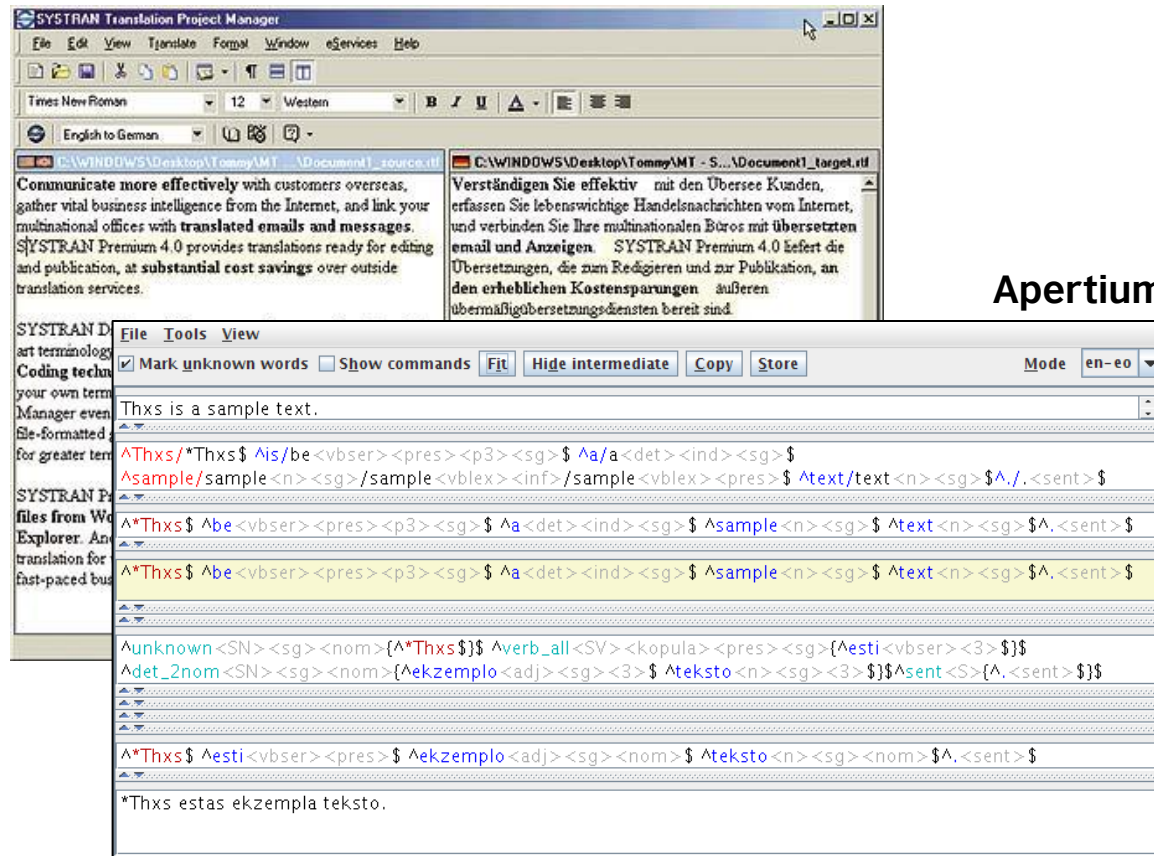
Sistemas basados en reglas (1960-2000)

SYSTRAN



Sistemas basados en reglas (1960-2000)

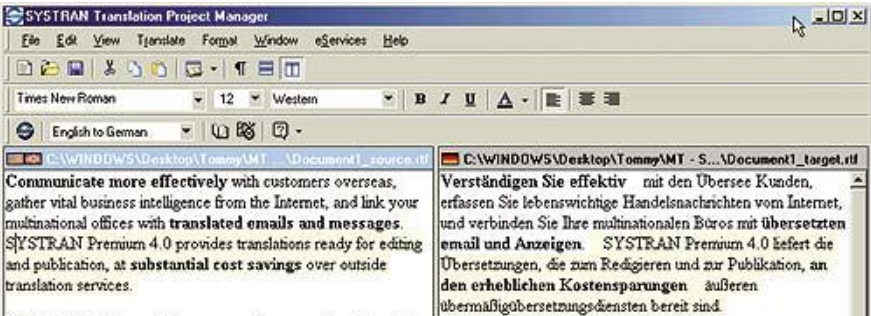
SYSTRAN



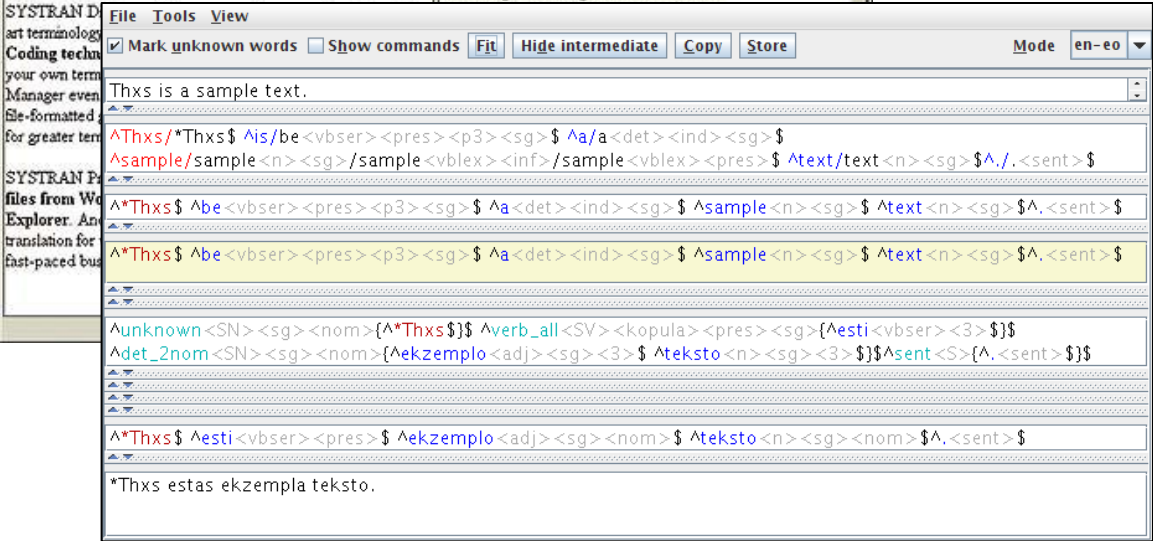
Apertium

Sistemas basados en reglas (1960-2000)

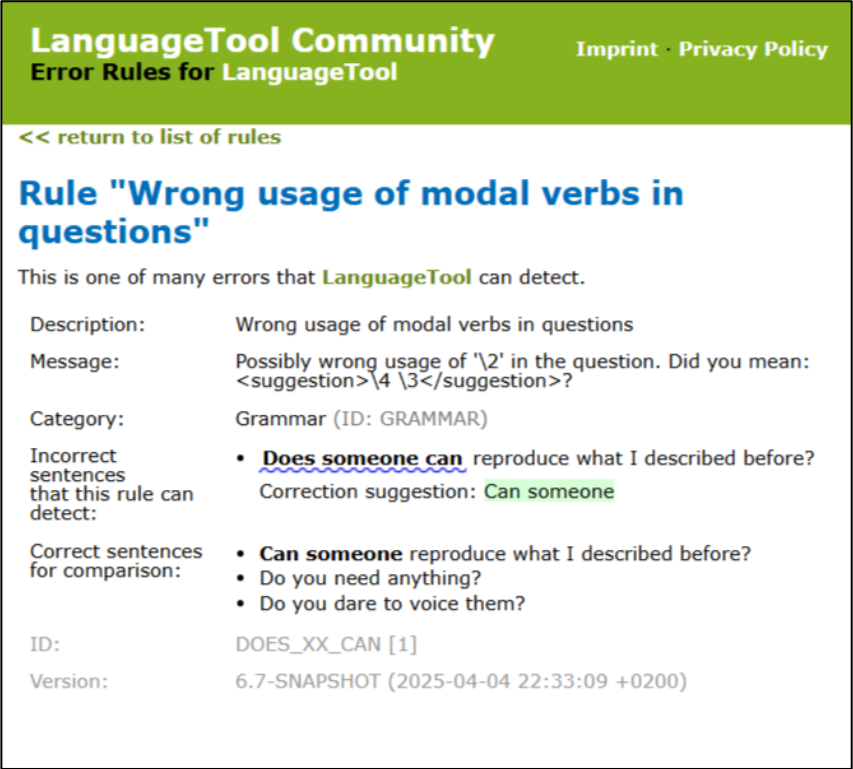
SYSTRAN



Apertium

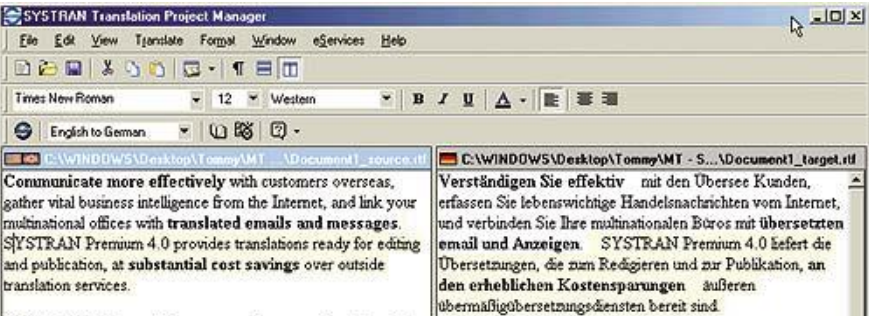


LanguageTool



Sistemas basados en reglas (1960-2000)

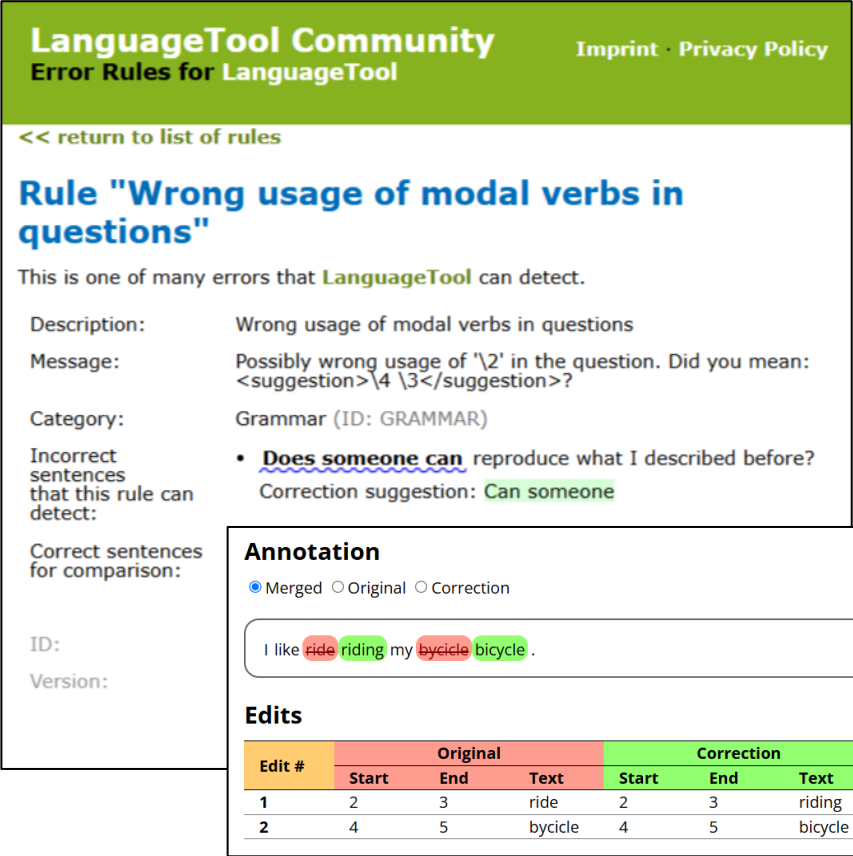
SYSTRAN



Apertium

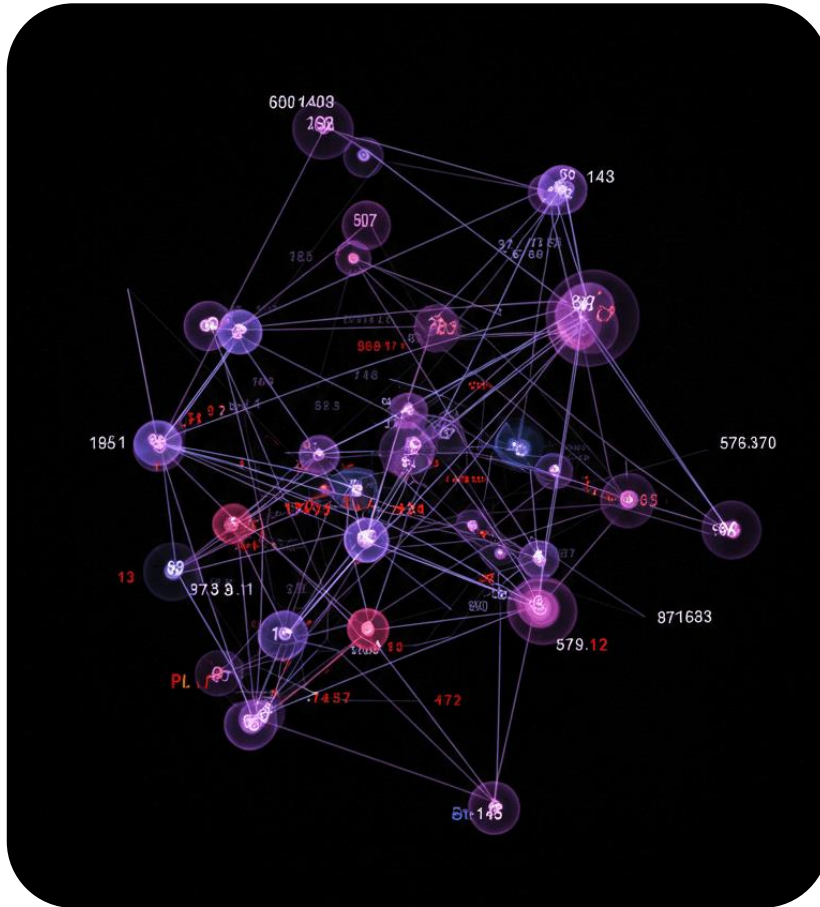


LanguageTool



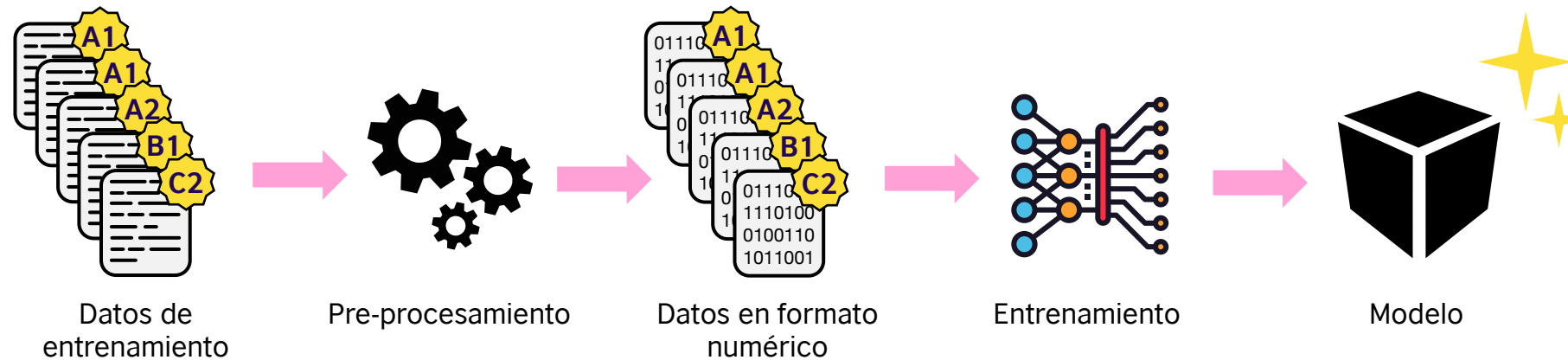
ERRANT

Aprendizaje automático (1990-presente)

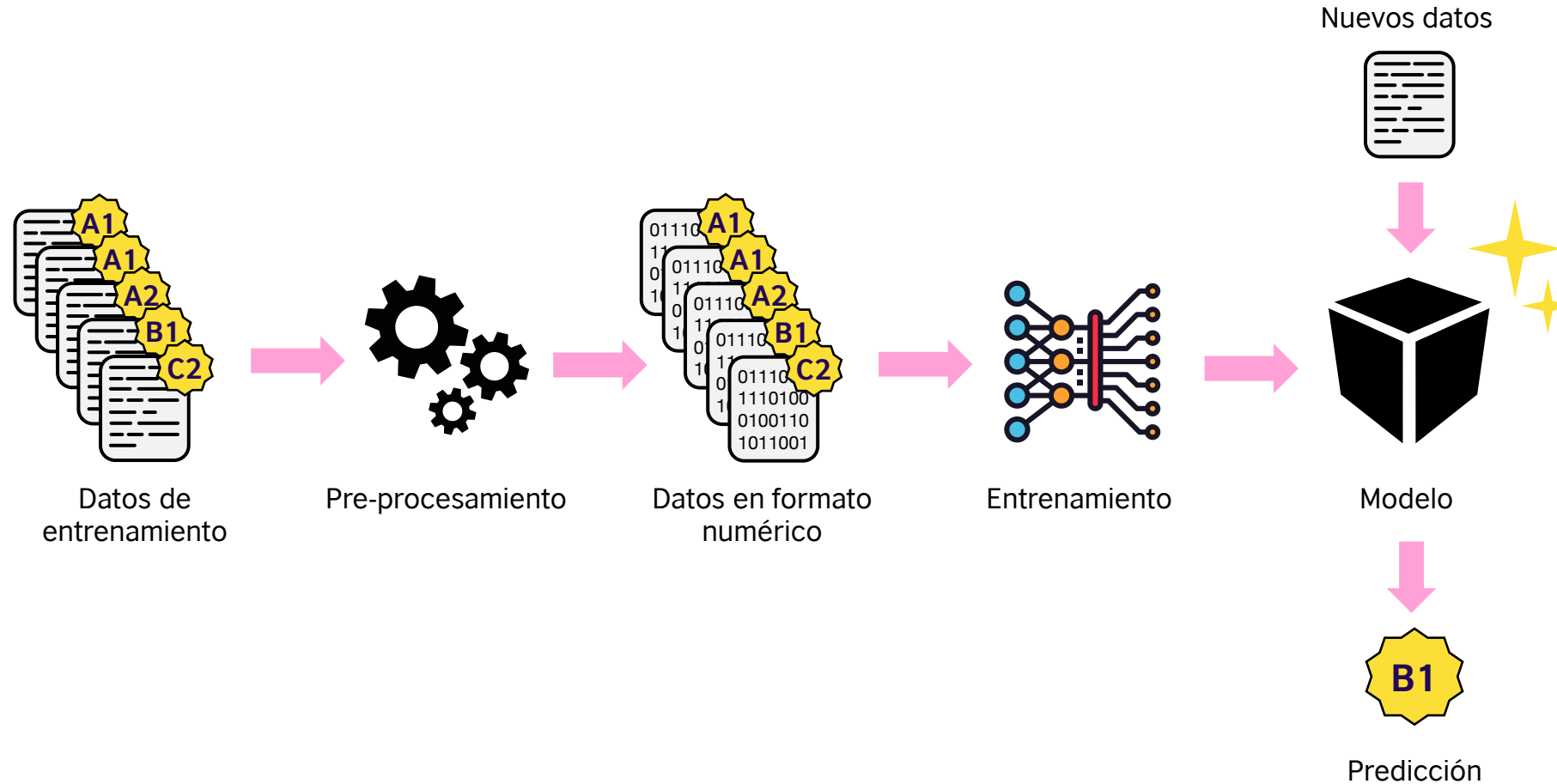


- ★ Disponibilidad de grandes volúmenes de datos (corpora).
- ★ Variedad de algoritmos.
- ★ La máquina aprende asociaciones entre los datos de entrada y salida para poder realizar predicciones.
- ★ Es la tecnología detrás de la IA que conocemos hoy.

¿Cómo funciona el aprendizaje automático?



¿Cómo funciona el aprendizaje automático?



Atributos (features)

A2

Contenido

- Solapamiento con la consigna
- Similitud con buenas composiciones

Lenguaje

- Frecuencia de vocabulario
- Relación tipo-token
- Tasas de error
- Probabilidad promedio de las oraciones

No me acuerdo de muchos sueños cuando me levanto. Además no recuerdo casi nada de los bonitos sino de los horribles. Cuando tengo el bonito lo quería seguir viendo y por eso vuelvo a dormirme pensando en ese sueño. Algunas veces logro y otras no. En sueños yo veo a algunas personas famosas que me gustan y hablo con ellos como si fuéramos amigos desde hace tiempo. Y siempre yo pienso cuando me levanto: ¡Si fuera real!

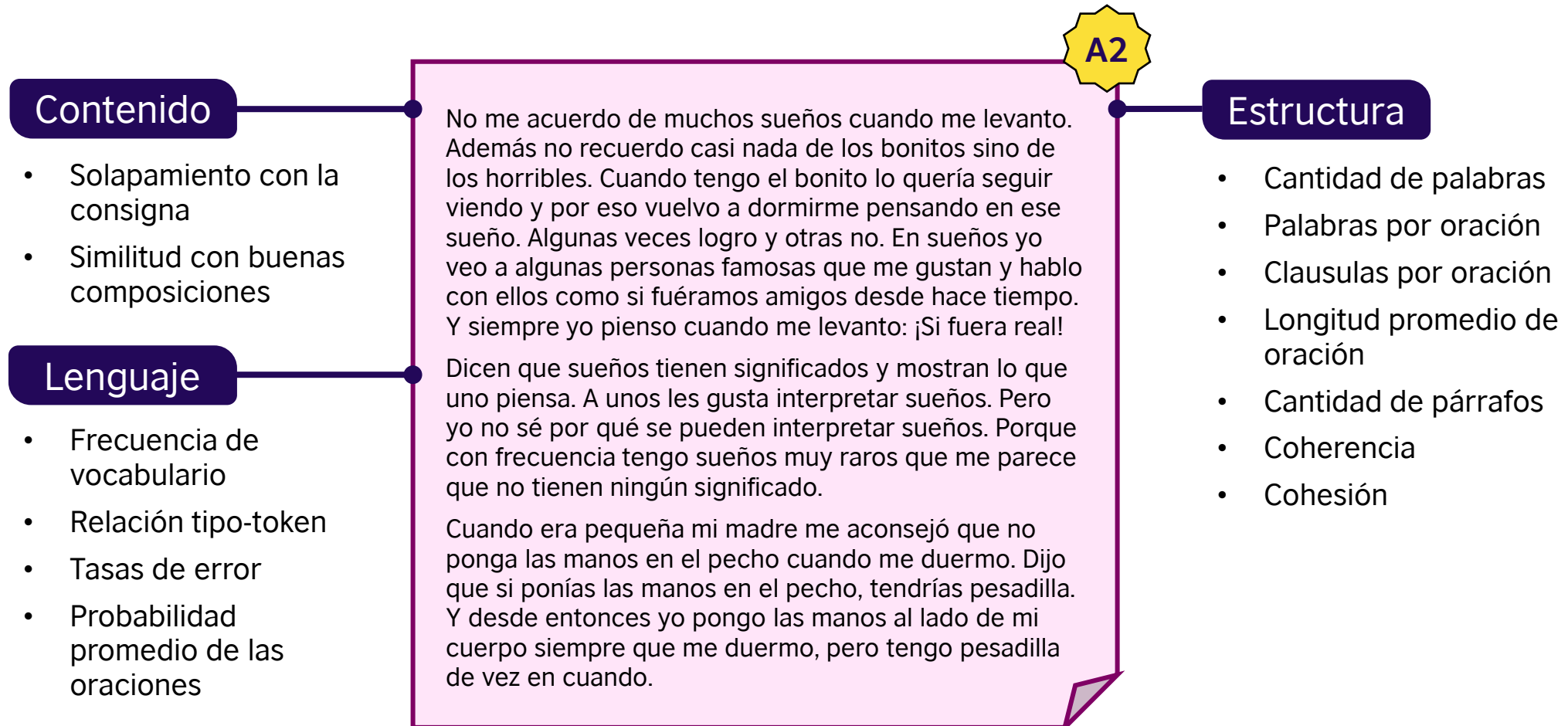
Dicen que sueños tienen significados y muestran lo que uno piensa. A unos les gusta interpretar sueños. Pero yo no sé por qué se pueden interpretar sueños. Porque con frecuencia tengo sueños muy raros que me parece que no tienen ningún significado.

Cuando era pequeña mi madre me aconsejó que no ponga las manos en el pecho cuando me duermo. Dijo que si ponías las manos en el pecho, tendrías pesadilla. Y desde entonces yo pongo las manos al lado de mi cuerpo siempre que me duermo, pero tengo pesadilla de vez en cuando.

Estructura

- Cantidad de palabras
- Palabras por oración
- Clausulas por oración
- Longitud promedio de oración
- Cantidad de párrafos
- Coherencia
- Cohesión

Atributos (features)



Representación numérica (vector)

[0.5381, 0.9907, 0.9364, 0.9145, 0.411, 0.1033, 0.7923, 0.5032, 0.6202, 0.331, 0.5972, 0.5306, 0.217, 0.8941, 0.735, 0.3623, 0.2395, 0.1826, 0.6314, 0.915, 2]

Modelos de lenguaje (n-gramas)

- ★ Secuencias de n palabras (generalmente de 2, 3, 4, 5)
- ★ Predicen sólo de izquierda a derecha, utilizando las $n-1$ palabras anteriores.

Modelos de lenguaje (n-gramas)

- ★ Secuencias de n palabras (generalmente de 2, 3, 4, 5)
- ★ Predicen sólo de izquierda a derecha, utilizando las $n-1$ palabras anteriores.

Contra viento y

Modelos de lenguaje (n-gramas)

- ★ Secuencias de n palabras (generalmente de 2, 3, 4, 5)
- ★ Predicen sólo de izquierda a derecha, utilizando las $n-1$ palabras anteriores.

Contra viento y marea.

Modelos de lenguaje (n-gramas)

- ★ Secuencias de n palabras (generalmente de 2, 3, 4, 5)
- ★ Predicen sólo de izquierda a derecha, utilizando las $n-1$ palabras anteriores.

Contra viento y marea.

Al que madruga, Dios lo

Modelos de lenguaje (n-gramas)

- ★ Secuencias de n palabras (generalmente de 2, 3, 4, 5)
- ★ Predicen sólo de izquierda a derecha, utilizando las $n-1$ palabras anteriores.

Contra viento y marea.

Al que madruga, Dios lo bendiga.

Modelos de lenguaje (n-gramas)

- ★ Secuencias de n palabras (generalmente de 2, 3, 4, 5)
- ★ Predicen sólo de izquierda a derecha, utilizando las n-1 palabras anteriores.

Contra viento y marea.

Al que madruga, Dios lo bendiga.

Al que madruga, Dios lo

Modelos de lenguaje (n-gramas)

- ★ Secuencias de n palabras (generalmente de 2, 3, 4, 5)
- ★ Predicen sólo de izquierda a derecha, utilizando las n-1 palabras anteriores.

Contra viento y marea.

Al que madruga, Dios lo bendiga.

Al que madruga, Dios lo ayuda.

Modelos de lenguaje (n-gramas)

- ★ Secuencias de n palabras (generalmente de 2, 3, 4, 5)
- ★ Predicen sólo de izquierda a derecha, utilizando las n-1 palabras anteriores.

Contra viento y marea.

Al que madruga, Dios lo bendiga.

Al que madruga, Dios lo ayuda.

No sé para qué quiere un coche si nunca aprendió a

Modelos de lenguaje (n-gramas)

- ★ Secuencias de n palabras (generalmente de 2, 3, 4, 5)
- ★ Predicen sólo de izquierda a derecha, utilizando las n-1 palabras anteriores.

Contra viento y marea.

Al que madruga, Dios lo bendiga.

Al que madruga, Dios lo ayuda.

No sé para qué quiere un coche si nunca aprendió a leer.

Modelos de lenguaje (n-gramas)

- ★ Secuencias de n palabras (generalmente de 2, 3, 4, 5)
- ★ Predicen sólo de izquierda a derecha, utilizando las n-1 palabras anteriores.

Contra viento y marea.

Al que madruga, Dios lo bendiga.

Al que madruga, Dios lo ayuda.

No sé para qué quiere un coche si nunca aprendió a leer.

No sé para qué quiere un coche si nunca aprendió a

Modelos de lenguaje (n-gramas)

- ★ Secuencias de n palabras (generalmente de 2, 3, 4, 5)
- ★ Predicen sólo de izquierda a derecha, utilizando las n-1 palabras anteriores.

Contra viento y marea.

Al que madruga, Dios lo bendiga.

Al que madruga, Dios lo ayuda.

No sé para qué quiere un coche si nunca aprendió a leer.

No sé para qué quiere un coche si nunca aprendió a conducir.

Modelos de lenguaje (n-gramas)

- ★ Secuencias de n palabras (generalmente de 2, 3, 4, 5)
- ★ Predicen sólo de izquierda a derecha, utilizando las n-1 palabras anteriores.

Contra viento y marea.

Al que madruga, Dios lo bendiga.

Al que madruga, Dios lo ayuda.

No sé para qué quiere un coche si nunca aprendió a leer.

No sé para qué quiere un coche si nunca aprendió a conducir.

- ★ Permiten asignar probabilidades a secuencias de palabras.

El motor es viejo y habría que revisarlo.

El motores viejo y abría que revisarlo.

Modelos de lenguaje (n-gramas)

- ★ Secuencias de n palabras (generalmente de 2, 3, 4, 5)
- ★ Predicen sólo de izquierda a derecha, utilizando las n-1 palabras anteriores.

Contra viento y marea.

Al que madruga, Dios lo bendiga.

Al que madruga, Dios lo ayuda.

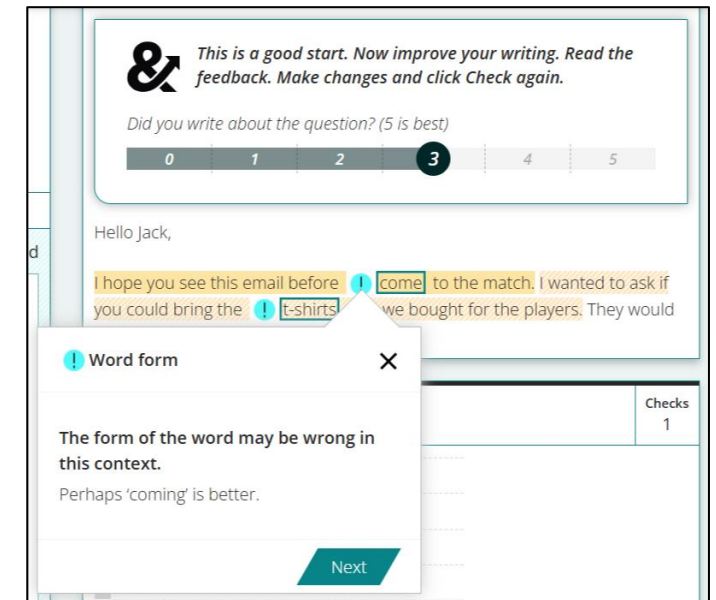
No sé para qué quiere un coche si nunca aprendió a leer.

No sé para qué quiere un coche si nunca aprendió a conducir.

- ★ Permiten asignar probabilidades a secuencias de palabras.

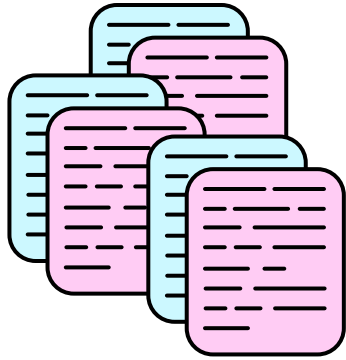
El motor es viejo y habría que revisarlo.

El motores viejo y abría que revisarlo.

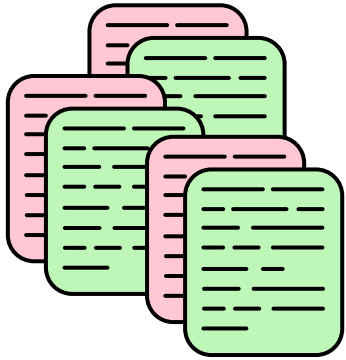


Write&Improve

Traducción automática estadística



Άφησα το βιβλίο μου στο σπίτι .
I left my book at home .



Esta cerca de el centre de la ciudad .
Está cerca del centro de la ciudad .

Resumen

- ★ Diversidad de algoritmos.
- ★ Redes neuronales poco eficientes.
- ★ Comprensión semántica superficial.

sociología – sociedad

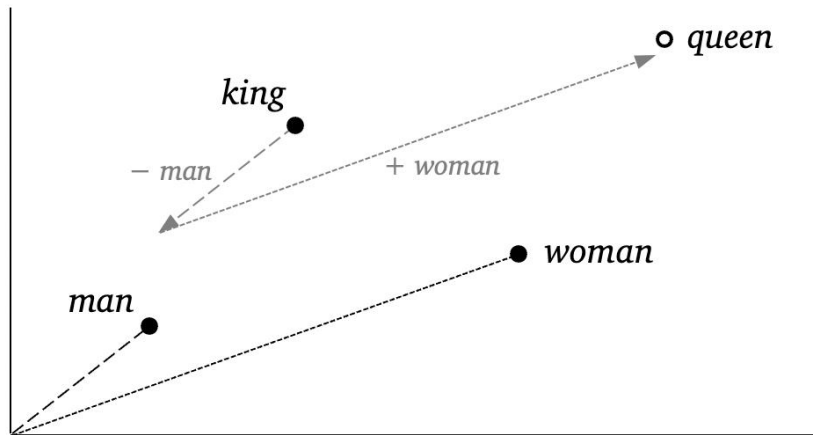
coche – bus – vehículo

“Los efectos de las redes sociales en la salud mental” ≠
“Instagram y Twitter en la mira de los psicólogos”

Presente

Representaciones vectoriales (word embeddings)

★ Word2Vec (2013)



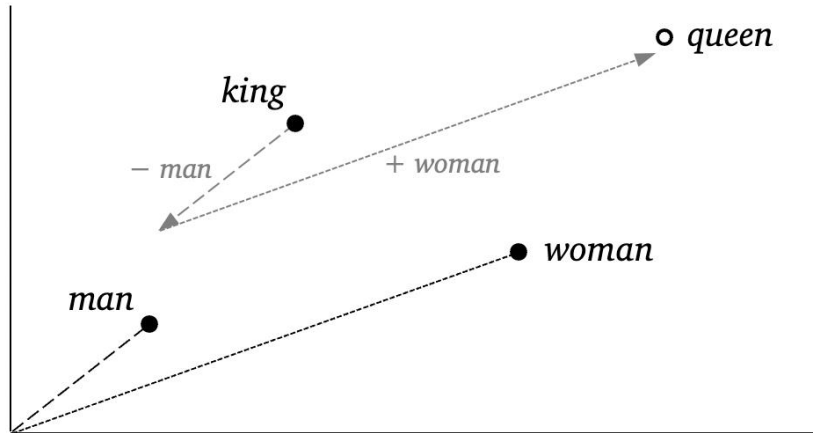
Fuente: <https://towardsdatascience.com/whats-behind-word2vec-95e3326a833a/>

★ Los vectores capturan las relaciones entre las palabras.

$$\text{king} - \text{man} + \text{woman} = \text{queen}$$

Representaciones vectoriales (word embeddings)

★ Word2Vec (2013)



Fuente: <https://towardsdatascience.com/whats-behind-word2vec-95e3326a833a/>

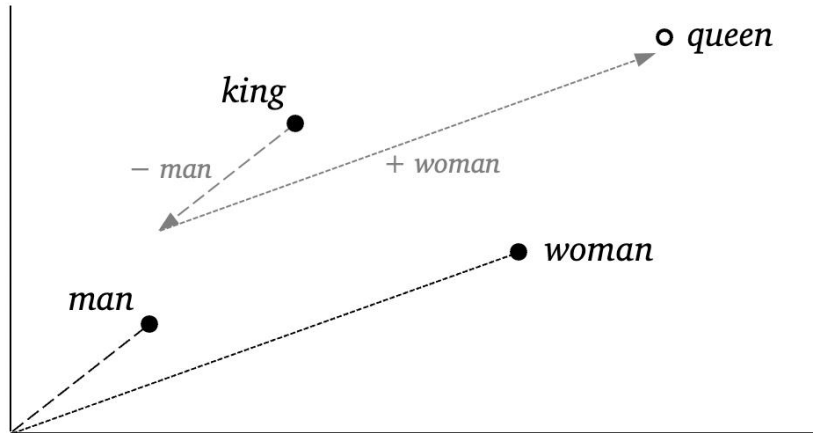
★ Los vectores capturan las relaciones entre las palabras.

$\text{king} - \text{man} + \text{woman} = \text{queen}$

$\text{France} - \text{Paris} + \text{Tokyo} =$

Representaciones vectoriales (word embeddings)

★ Word2Vec (2013)



Fuente: <https://towardsdatascience.com/whats-behind-word2vec-95e3326a833a/>

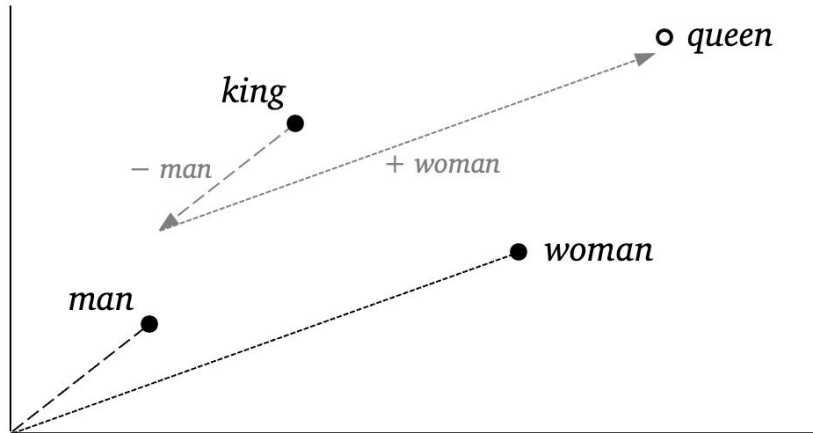
★ Los vectores capturan las relaciones entre las palabras.

king – man + woman = queen

France – Paris + Tokyo = Japan

Representaciones vectoriales (word embeddings)

★ Word2Vec (2013)



Fuente: <https://towardsdatascience.com/whats-behind-word2vec-95e3326a833a/>

★ Los vectores capturan las relaciones entre las palabras.

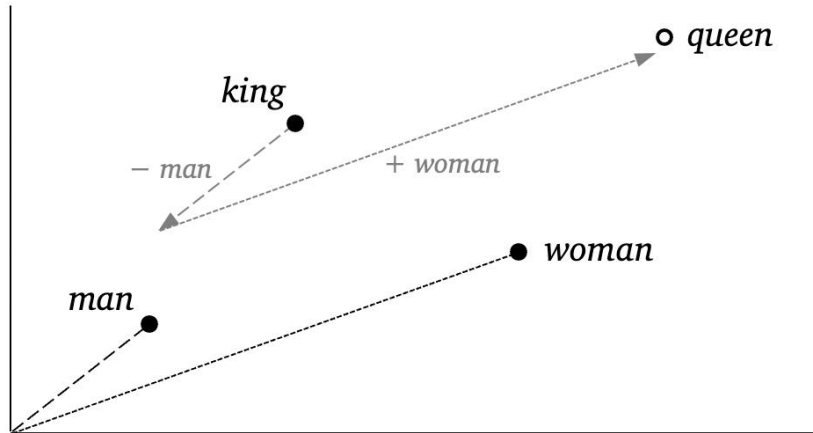
king - man + woman = queen

France - Paris + Tokyo = Japan

bigger - big + cold =

Representaciones vectoriales (word embeddings)

★ Word2Vec (2013)



Fuente: <https://towardsdatascience.com/whats-behind-word2vec-95e3326a833a/>

★ Los vectores capturan las relaciones entre las palabras.

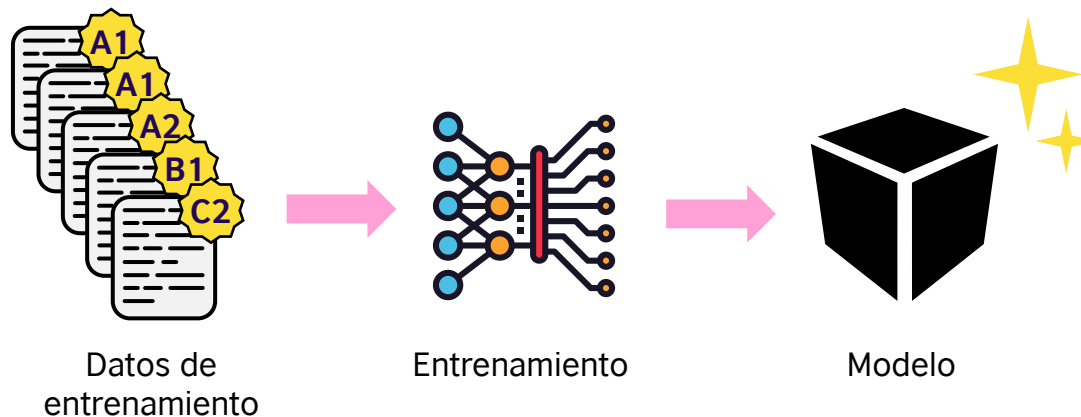
king - man + woman = queen

France - Paris + Tokyo = Japan

bigger - big + cold = colder

Aprendizaje profundo (deep learning)

- ★ Grandes redes neuronales.
- ★ Las máquinas identifican *features* por su cuenta.
- ★ Las máquinas comienzan a “comprender” el contenido (semántica).
- ★ Mayor capacidad de abstracción y performance.
- ★ Enormes avances en tareas como la traducción, el procesamiento de imágenes, la simplificación de textos, etc.



Transformers (2017)

- ★ Mecanismo de atención, que identifica qué elementos de la entrada son importantes.

The animal didn't cross the street because it was too tired .

- ★ BERT (2018)
 - ★ Distingue entre significados de una misma palabra (vectores contextuales)
 - ★ Bidireccional

No sé para qué quiere un coche si nunca aprendió a conducir.
- ★ Es el motor detrás de la IA generativa.

- ★ Usos en la enseñanza de idiomas:

- ★ Ayuda en la creación y resolución de ejercicios (ej. cloze)
- ★ Andamiaje/apoyo a la escritura.
- ★ Detección de respuestas relevantes.

La IA generativa



- ★ Irrupción de ChatGPT (2022).
- ★ Explosión y “democratización” de la IA.
- ★ Generación de contenido (texto, audio, imágenes y video).
- ★ Alucinaciones e imprecisiones.
- ★ Han forzado a redefinir las evaluaciones.

Los Grandes Modelos de Lenguaje (LLMs)

- ★ Predicción de palabras + capacidad de seguir instrucciones.
- ★ Entrenados mediante “Aprendizaje por Refuerzo con Retroalimentación Humana” (RLHF).
- ★ Multilingües.
- ★ Requieren destreza en la redacción de instrucciones (*prompt engineering*).
- ★ Usos en la práctica y evaluación de idiomas:
 - Generación de contenido (textos, ejercicios)
 - Explicaciones
 - Traducción
 - Resúmenes
 - Conversaciones
 - Evaluación y comentarios



Los Grandes Modelos de Lenguaje (LLMs)

- ★ Son el motor detrás de la mayoría de las aplicaciones de generación de contenido

Create a text on a certain topic

Creates a text on a certain topic. Can be used during a lesson or as a homework task after a reading, listening or vocabulary lesson.

Input a topic or vocabulary, or both, to create a tailored text for reading!

Topic

Artificial Intelligence

or / and

Target Vocabulary

Large Language Model x AI x ChatGPT

Human x Robot x Evolution x Singu

Genre

Report

Length range

100-250

Level

B2

Do the magic!

From which source would you like to generate questions?

Topic Document Text Link Audio/video

Give the wizard a document, and it will create questions about its content!

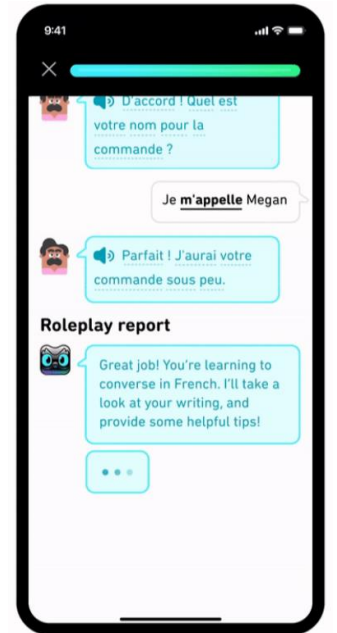
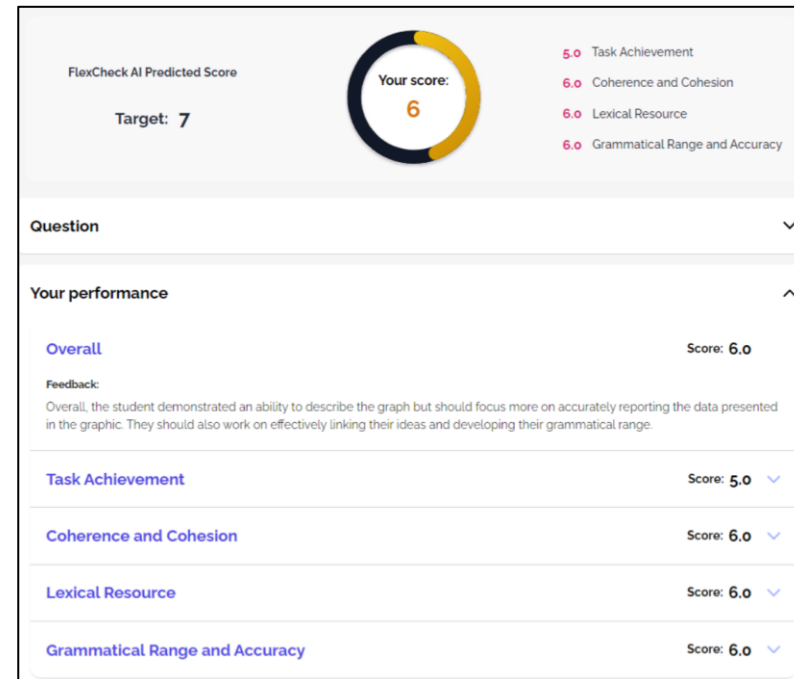
Click here to upload a file

PDF, PowerPoint, OpenDocuments, Pages, Text

Max 200 MB

Continue

- ★ ...y las aplicaciones de práctica y evaluación de idiomas.



Reconocimiento automático del habla (ASR) + Síntesis de voz (TTS)

- ★ Reconocimiento “listo para usar”.
- ★ Voces hiperrealistas, diversos acentos y clonación de voz.
- ★ Permiten implementar soluciones de accesibilidad.
- ★ Usos en la práctica y evaluación de idiomas:
 - Práctica de la pronunciación.
 - Conversación.
 - Comprensión auditiva.



Whisper

Deepgram



ElevenLabs

Generación de imágenes y video

- ★ Generan imágenes y videos (cortos) a partir de instrucciones (prompt).
- ★ Costoso y de cómputo intensivo.
- ★ Resultados difíciles de controlar.
- ★ Usos en la práctica y evaluación de idiomas:
 - Creación de material audiovisual personalizado.
 - Integración en tareas multimodales.
 - Apoyo a la comprensión auditiva y desarrollo del habla.



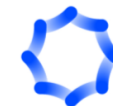
Sora



LUMA AI



ImageFX



synthesia

¿Quiénes están usando IA en sus productos?

 **Criterion.**

 **turnitin™**

 **INGER**

Fonetti
The Read Aloud App

Write&Improve
with Cambridge

 **grammarly**

 **speechace**

MyACCESS!
SCHOOL EDITION
BECAUSE WRITING MATTERS

+Babbel


duolingo


Amira
LEARNING


LinguaMetrics™

 **Cognii**

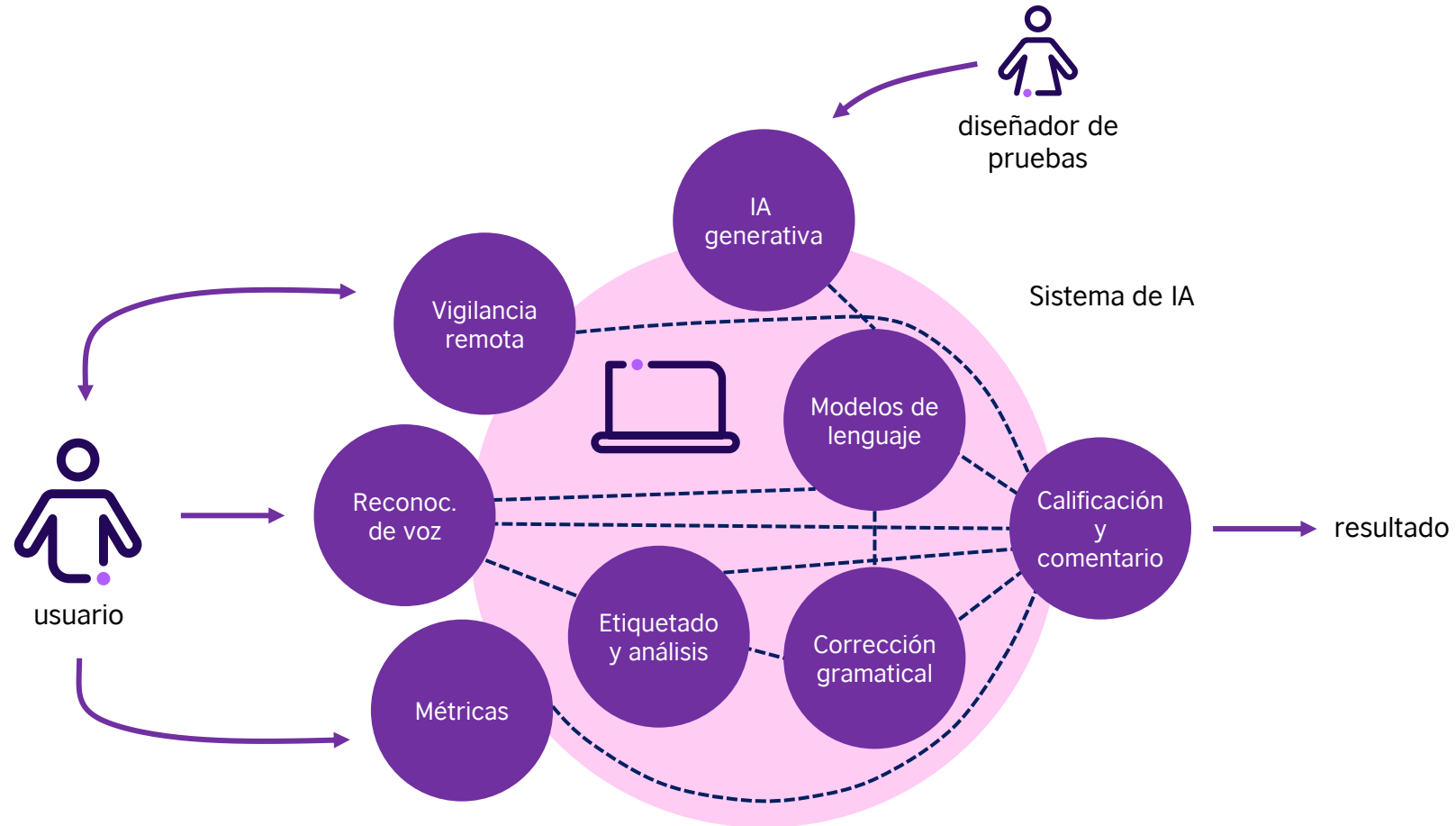
 **BLUE CANOE**

 **MetaMetrics®**

La IA en las pruebas estandarizadas

Organización	Prueba	Contenido	Adaptabilidad	Calificación
ETS	TOEFL iBT Home Edition			S+W
Pearson	PTE & Versant			S+W
Duolingo	DET	✓	✓	S+W
Cambridge	Linguaskill		✓	S+W
British Council	Primary English Test		✓	S+W

Integración de la IA en pruebas



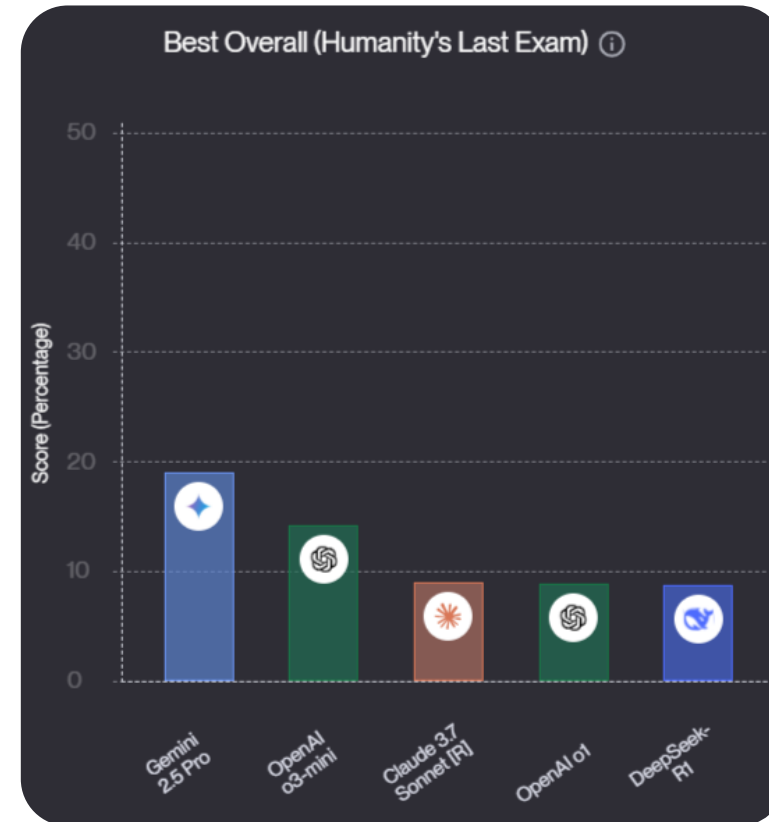
¿Qué modelo es el mejor?

- ★ La performance depende de la tarea.
- ★ Los resultados dependen de las instrucciones utilizadas.
- ★ Comparaciones rápidamente obsoletas.
- ★ Analizar resultados sobre corpus estándar.
- ★ Consultar sitios especializados:

<https://llm-stats.com/>

<https://artificialanalysis.ai/>

<https://www.vellum.ai/llm-leaderboard>



Futuro



Multilingüe

Multimodal

Abierto



Accesible

Adaptativo

Personalizado

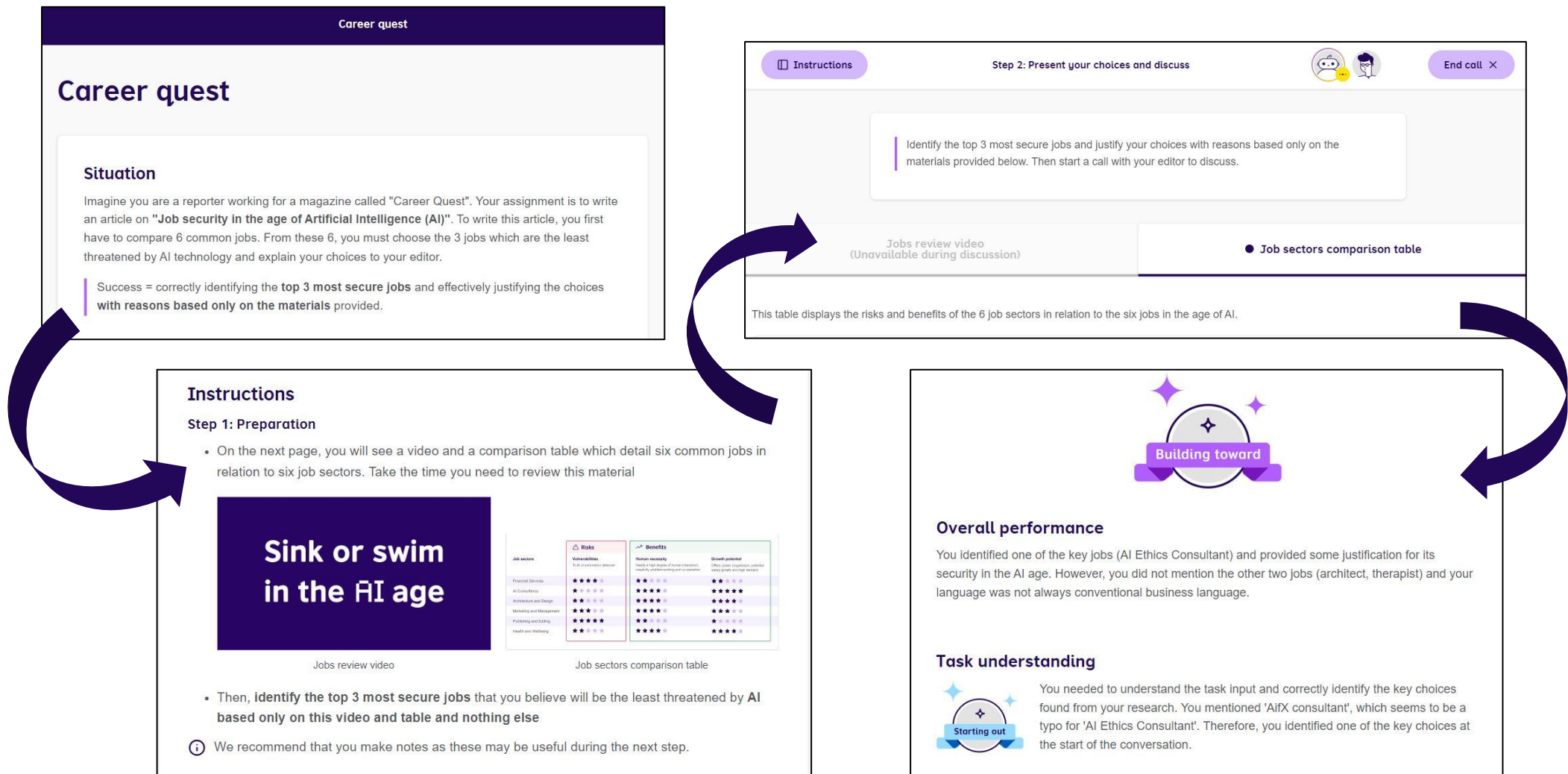


Realista

Inmediato

Flexible

Un ejemplo del British Council



Consideraciones éticas

El uso responsable de la IA en el British Council

Priorizar a los aprendices

Proteger a los humanos

Adoptar un enfoque equilibrado

Ser crítico y estratégico

Ética desde el diseño

Respeto por la Autonomía del Ser Humano

Privacidad y Gestión de Datos

Equidad

Bienestar Individual, Social y Ambiental

Transparencia

Responsabilidad y Supervisión

Consideraciones éticas

ist

• 11 mo. ago

Sep 12, 2023 11:18 AM, 1499 Views

☆☆☆☆☆

HUGE SCAM

Stick to the reliable ones such as IELTS or TOEFL. These scammers voided my test saying they weren't able to verify if all rules were followed. Didn't provide any further reason. I was able to dispute the charge and get my money back though. It's funny how this system is just accepted as a valid way to just grab money from people. Huge Scam, just stay away.

ist

• 9 mo. ago

Ok-Reward:

How reliable are scores for the speaking section?

Speaking

So, guys, I have been really anxious about the speaking section. I already sound somewhat confused in my own mother tongue due to autism + adhd + a minor speech problem so the whole idea of having to come up with a decent, coherent spoken answer in 15-30 seconds is scary to me. I have been practicing with and while I have aced the other sections, I have fared badly in the speaking one: I got 3/4 in all the subsections, and no matter what I do, I am not able to get 4/4. I am worried because I need a 25 in speaking.

8

7

Share

Ark

• Jun 08, 2023

☆☆☆☆☆

Joke

This test is a joke. They want you to How sensible and financing in our time write full text without their eyes, or look find any sensible

da_foamy_pancake

• 9mo ago

• s AI tends to score people lower on speaking than what a human would rate your speaking. I also got 3/4 in every speaking subsection and ended up getting 26/30 on the actual test. My advice is to keep practicing and relax, you got this!

7

Share

watchsmart

• 9mo ago

Top 10% Commenter

Yeah. The problem is that the AI doesn't seem to have decimals. It can give you 3/4 (22.5 out of 30) or it can give you 4/4 (30 out of 30). Nothing in-between.

3

Share

2 more replies

This is not test just a speed race where you not allowed to think twice. For example in my laptop I have very bad keyboard, I could use external one, but if I even touch a touch pad or push button on wrong keyboard they ban straight away.

When we stopped to be human being and we cannot correct ourselves? where is a time for stress? I do not have problem with language, but I am human being and stress make huge impact on me.

☆☆☆☆☆

1 Comment

butte

• 3 mo. ago

speaking AI is absoluteley uselles

Speaking

I just did an independent talking test and got scored 3/4 by the AI (same as what even though I just talked in my native language about something not at all related in french, gets you the same grade as giving concise english answers????

14

34

Share

tentigen

• 3mo ago

doesn't really assess what you say, but how you say it (smo human scorer will determine whether or not you actually answer the que we think of. Haha

7

Share

Far_Organization_610

• 3mo ago

Is the one used on ?

2

Share

5 more replies

Far_Organization_610

• 3mo ago

Edited 3mo ago

Haha ok I'm gonna try this now. But yes, I agree the AI is kinda bad, I com three sections but on speaking it seems almost random, weird.

Edit: ??? you were right, I got a 3/4??

4

Share

butte

OP • 3mo ago

We pay 280\$ for this shit????

2

Share

3 more replies

butte

OP • 3mo ago

Also the transcript is not at all accurate with what english person..

3

Share

butte

• 6 mo. ago

practice AI doesn't understand what I'm saying..

Speaking

Hey I'm currently studying for the and I deem my speaking skill as my best quality. Even though that's the case the AI that's supposed to grade my speaking results always gives me a bad grade, because it can't seem to comprehend what the hell I'm talking about. My mic is pretty good so this shouldn't be the reason for it and I actually have a decent accent. I just checked the transcript and the AI just understand every fifth word wrong. For example: "enter" becomes "and through", "over time" becomes "of the time", "incentivize" becomes "inside essential". Can someone confirm that it's not the AI that grades this during the actual test, because I've no clue how to improve that.

Thanks a lot for the answer

1

8

Share

Infamous_Focus9060

• 6mo ago

Edited 6mo ago

Same issue here! My transcripts are bizarre in some points. I've never faced issues with pronunciation, got 109/120 in my score a few years ago and I'm going crazy with these 3/4 scores on regularly built sentences. It goes from complex words to changing "a" to "I". I even changed my microphone setting but it didn't help as much as I thought it would. Tried also doing these activities on my cell phone, but no improvement.

And yes, I've sent the audios to a seasoned teacher, and he pointed out that there wasn't any cohesion issue, as the results show. I'm sort of in panic - if this is how I am going to be evaluated on test day, then it's wasted money and it's better to do the IELTS.

2

Share

Infamous_Focus9060

• 2mo ago

Just for the record: got an 116 with 30L/30R/28S/28W. Don't trust the AI grading on platform, you're gonna do fine

1

Share

@t

• Aug 02, 2023 03:24 PM, 1602 Views

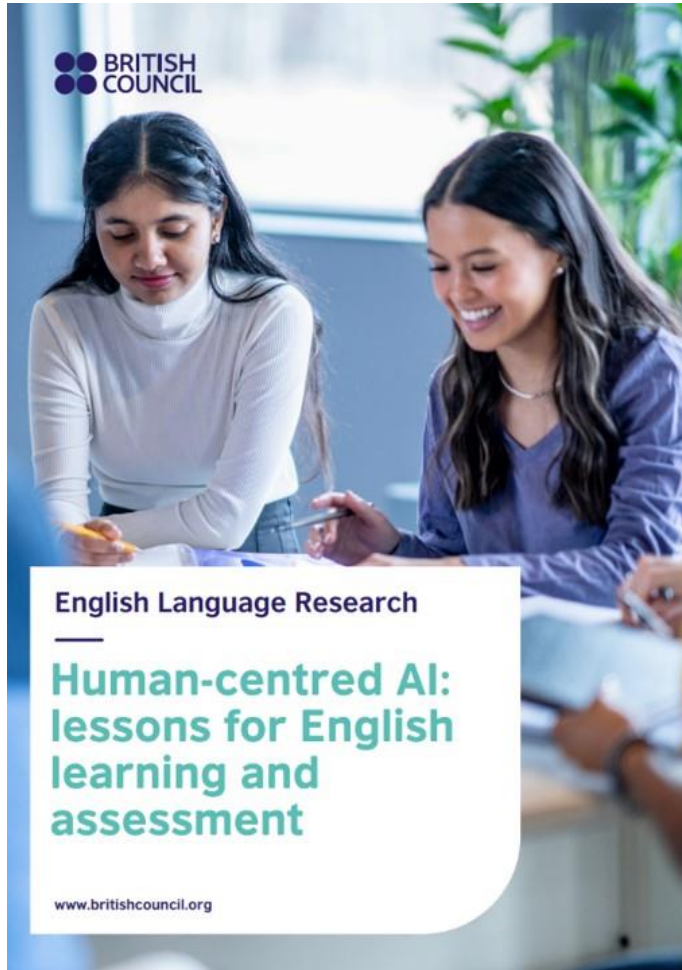
☆☆☆☆☆

A Challenging Experience: My Review of

Overall, my experience with's English testing service has been disappointing, both in terms of time and money. The strict rules imposed during the exam felt dehumanizing, reducing test-takers to mere robotic beings. It is regrettable that diligently adhering to these rules did not guarantee a passing result. I believe that this testing process goes beyond ensuring integrity and instead adds unnecessary stress and frustration for test-takers.

☆☆☆☆☆

1 Comment



Dr Mariano Felice



Richard Spiby



Professor Barry
O'Sullivan OBE



Dr Adam Edmett

¡Gracias por vuestra ~~a tensión!~~
atención

Mariano.Felice@britishcouncil.org

En resumen...

