



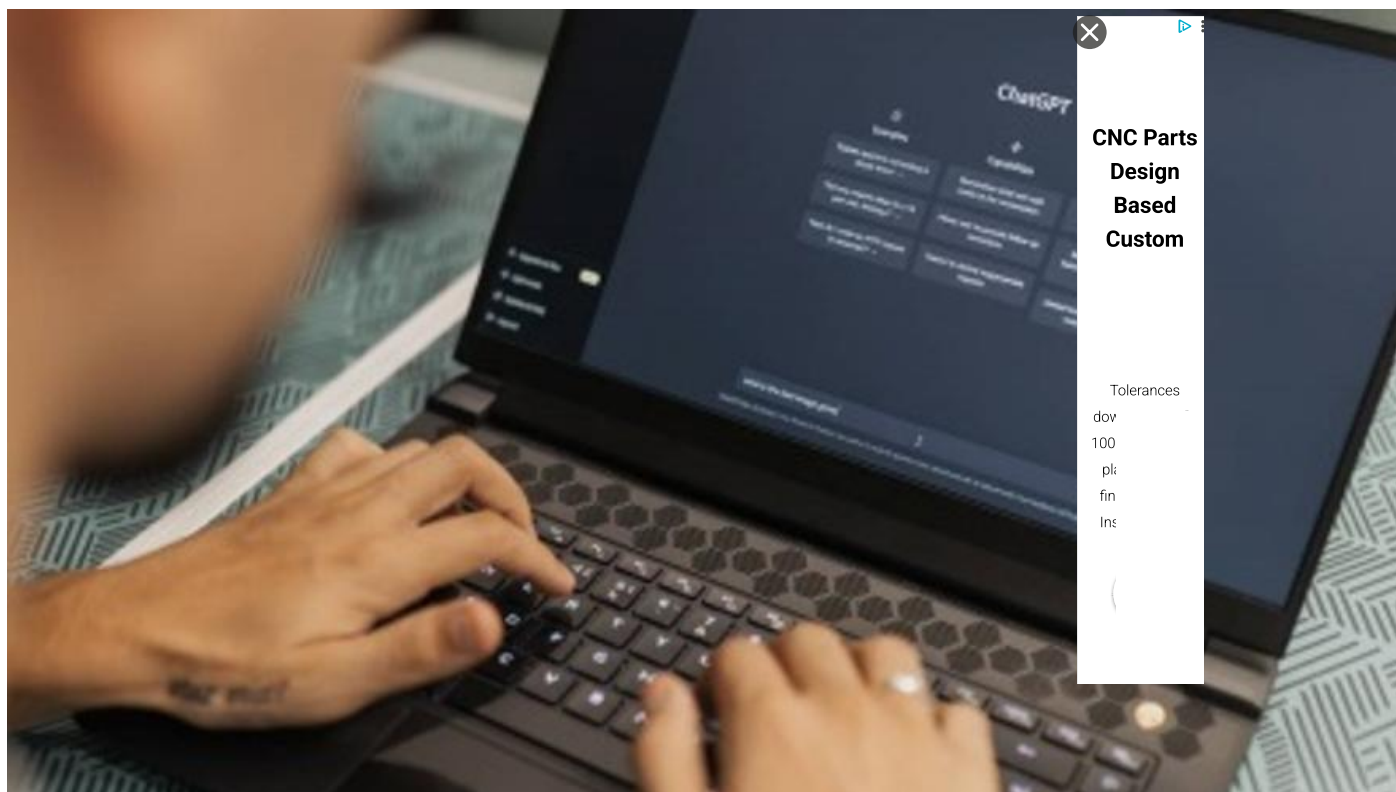
La inteligencia artificial vasca Latxa supera al ChatGPT

El centro HiTZ de la UPV/EHU asegura que las mejoras implantadas son “prometedoras”



Ane López / NTM

Bilbao | 30-04-24 | 06:49



El Centro Vasco de Tecnología del Lenguaje de la UPV/EHU, HiTZ, ha conseguido que Latxa supere a ChatGPT. / URJC

“Cuando salió ChatGPT era inimaginable que se pudiera conseguir



del Lenguaje de la [UPV/EHU](#), HiTZ, lo ha conseguido y Latxa “ya supera al ChatGPT en competencia lingüística”. De esa modo, está por delante de ChatGPT original lanzado hace un año, y por primera vez supera la última versión (GPT-4) en competencia lingüística”.

Los modelos grandes de lenguaje son un tipo de modelo de inteligencia artificial que emplea técnicas de aprendizaje automático para comprender y generar lenguaje humano, basándose en el conocimiento obtenido de conjuntos de datos masivos. Y gracias a el grupo de investigación de la UPV/EHU, el euskera cuenta ya con su gran modelo de lenguaje: Latxa, “estrenado” en enero de este mismo año, y “mejorado” en pocos meses.

PUBLICIDAD

Según explican los investigadores en una nota, “está basado en los modelos LLaMA de Meta y reúne modelos de entre 7 y 70 mil millones de parámetros. Los LLMs de hoy en día tienen un impresionante rendimiento en idiomas con muchos recursos”; por ejemplo, ChatGPT.

Pero en el caso del [euskera](#) y otros idiomas con pocos recursos, “su rendimiento deja mucho que desear”. Este hecho “aumenta la brecha tecnológica entre idiomas” con muchos y pocos recursos, al menos en cuanto a herramientas digitales. Es por eso que HiTZ ha desarrollado Latxa con el objetivo de “superar esos límites y fomentar el desarrollo de investigaciones, innovaciones y productos que funcionan con euskera” basado en LLMs.



conseguido resultados tan buenos con un corpus relativamente tan pequeño”, porque “abre la puerta a más mejoras según investiguemos en nuevas técnicas”.

De hecho, de cara al futuro sostienen que “es un resultado muy prometedor, no solamente para el euskera sino para el resto de lenguas con pocos recursos digitales”. La nueva versión de Latxa ha sido entrenada sobre el mayor corpus público en euskera, que también se distribuye junto con los modelos. Este corpus extiende el ya existente EusCrawl con otros corpus contruidos en su mayoría sobre contenido público en Internet.

PUBLICIDAD

En total son más de 4 millones de documentos y 1.200 millones de palabras, doblando en tamaño a los corpus existentes hasta el momento. Con el fin de evaluar la calidad de los modelos, se han construido varios bancos de prueba sobre competencia lingüística, comprensión lectora, cultura general y exámenes profesionales. Etxaniz, señala que los resultados obtenidos con esta herramienta “abren la puerta a más mejoras según investiguemos en nuevas técnicas”.

Por su parte, Naiara Pérez asegura que “el euskera se encuentra en la posición 50 entre los idiomas del mundo según la cantidad de texto en Internet, y hay decenas de otras lenguas que tienen cantidades similares de texto, con lo que las técnicas aplicadas al euskera pueden también

[Kiosko](#)

Para expertos

Por el momento, Latxa está orientado al trabajo de los expertos, no para interactuar directamente con los usuarios, según los responsables de la herramienta. Sin embargo, se encuentran ya trabajando en esta opción:

“Estamos trabajando en modelos que sean capaces de seguir las instrucciones del usuario, pero aún no está claro si es posible construir modelos de diálogo en euskera que tengan una calidad similar al español o al inglés. Este es precisamente el propósito de nuestra investigación. Estamos trabajando en varias direcciones, incluyendo la investigación en técnicas que permitan transferir las capacidades conversacionales ya existentes para otros idiomas al euskera”, apunta Eneko Agirre, director de HiTZ.



Mejoras

Para medir la mejora de Latxa, se sometió al sistema a varias pruebas de rendimiento. “Probamos los modelos Llama del inglés, así como GPT-3.5 Turbo (equivalente al ChatGPT lanzado en noviembre de 2022) y GPT-4 Turbo (el mejor modelo que tiene OpenAI) y se ve claramente que el mejor modelo Latxa supera a los Llama y a GPT-3.5 Turbo en todos los casos de prueba. El mejor modelo también supera GPT-4 Turbo en competencia lingüística, por primera vez para un modelo abierto de idiomas con pocos recursos digitales, pero no en el resto de bancos de prueba. Por último, es de subrayar que a medida que Latxa va creciendo de tamaño, los resultados también mejoran”.

4 millones

Su nueva versión incluye más de 4 millones de documentos y 1.200 millones de palabras, doblando en tamaño a los corpus existentes hasta el momento. Así, sostienen que “es destacable obtener resultados tan buenos con un corpus tan pequeño. De hecho, esto nos permite realizar más mejoras a medida que investigamos nuevas tecnologías. Se trata de un resultado muy prometedor, no solo para el euskera sino también para otras lenguas con escasos recursos digitales”.

TEMAS Euskera - Inteligencia artificial - UPV-EHU

Te puede interesar



[Kiosko](#)



Este ritual matutino de 7 segundos que desatasca el equivalente a semanas de heces atascadas

Patrocinado por Floravia

Esta aspiradora inalámbrica muy potente se vende por un precio insignificante

Patrocinado por AspiraPro

Nuevo EBRO s800
Luxury, seguridad a lo grande

Patrocinado por EBRO

Este conejito de juguete parece y se comporta como un conejo de verdad

Patrocinado por BunnyPal

Revolución solar: Lo que cambia con la nueva regulación en...

Patrocinado por Ayuda...



[Kiosko](#)



Almeida sanciona a Pablo Iglesias por exceso de aforo en la Taberna Garibaldi
Deia

Prohíben a David Broncano la entrada en Zarautz
Deia

Los motivos del conflicto entre Lalachus y Karlos Arguiñano
Deia



[Kiosko](#)





MÁS EN SOCIEDAD

SOCIEDAD

El director de Gogora afirma que una paz sostenible exige "evitar la simplificación del relato del pasado"



SOCIEDAD

Un hombre acepta 147 años de prisión por pagar a madres para que violasen a sus hijas mientras grababa



SOCIEDAD

Rocío, asesinada en Huelva, es la décima víctima mortal por violencia machista en 2025



SOCIEDAD

La justicia archiva de forma provisional el caso en el que se investiga a Nacho Cano por 'Malinche'



© Editorial Iparraguirre, S.A
Todos los derechos reservados

[Aviso legal](#) [Política privacidad](#) [Condiciones de uso](#) [Cookies](#) [Código ético](#) [Preferencias de privacidad](#)



[Kiosko](#)

