

UN IDIOMA PROPIO

ULISES CORTÉS

Los recientes avances en modelo masivo de lenguaje (LLM, por sus siglas en inglés) y las aplicaciones construidas sobre ellos, como Bert, ChatGPT, DeepSeek, LLaMa, etcétera, han revolucionado por completo las interacciones entre los humanos y los sistemas basados en la Inteligencia Artificial debido a su capacidad para generar textos inteligibles y coherentes. Su desempeño se debe a aplicaciones construidas sobre transformadores, preentrenadas con modelos construidos a partir de enormes cantidades de datos sin procesar.

Un transformador es un modelo de aprendizaje profundo diseñado para procesar datos secuenciales, como texto, mediante el mecanismo de autoatención. Este mecanismo asigna diferentes pesos a las

partes de la entrada a fin de identificar relaciones y contexto entre los elementos de una secuencia. Estas aplicaciones pueden, entre otras tareas, responder preguntas, resumir textos y participar en conversaciones, lo que las hace adecuadas para tareas simples en una variedad de campos, incluso a veces superando a los humanos en velocidad, aunque no necesariamente en calidad.

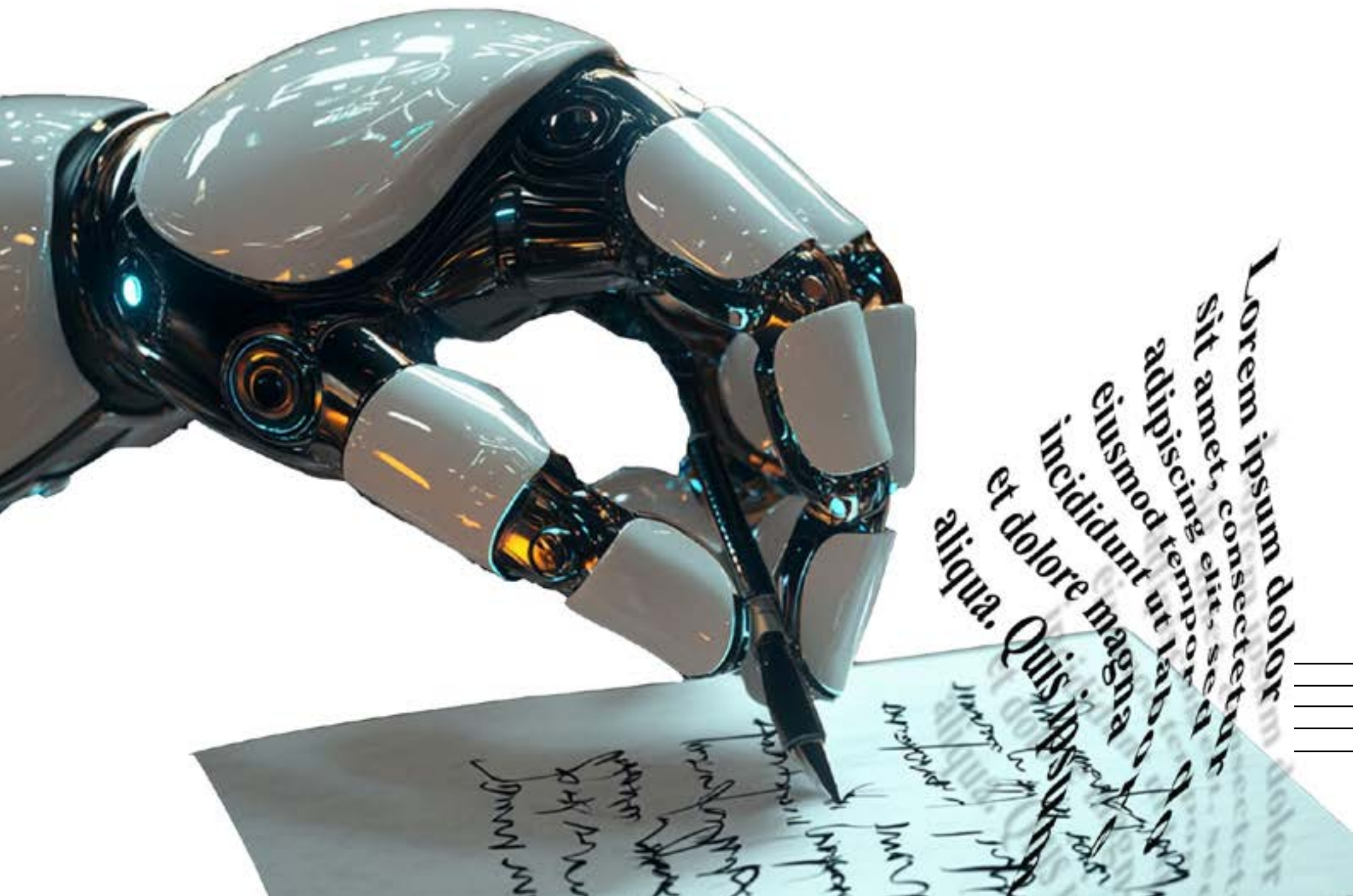
Sin embargo, a pesar de sus capacidades, estos modelos tienden a cometer errores en las respuestas, que sus desarrolladores llaman alucinaciones, lo que es un claro uso abusivo del lenguaje. Una *alucinación* ocurre cuando un LLM genera una respuesta en apariencia realista, pero que es no factual, es absurda o es inconsistente con la pregunta y el contexto proporcionado.

Estos errores (o alucinaciones) de los LLMs pueden llevar a la difusión de información falsa, lo que, en algunos casos, ocurre de manera no intencionada debido las prisas o a la falta de razonamiento crítico por parte de los usuarios al interpretar y validar las respuestas generadas.

No obstante, un problema aún más profundo reside en la falta de auditoría y certificación exhaustiva y oficial de estos modelos comerciales. Esta carencia incrementa de forma substancial el riesgo de que los LLM perpetúen y amplifiquen sesgos de género, racismo y la polarización de los discursos, lo que puede tener consecuencias perjudiciales en aplicaciones sensibles como la toma de decisiones y, en general, erosionar la confianza pública en los sistemas basados en la inteligencia artificial.

La creación de un modelo masivo de lenguaje (LLM) para el español, auspiciado por el gobierno mexicano, es necesario por varias razones. En primer lugar, garantiza la soberanía tecnológica de México, reduciendo la dependencia de modelos extranjeros y permitiendo el desarrollo de herramientas adaptadas a nuestro contexto cultural y lingüístico, algo que es fundamental para la seguridad nacional y el desarrollo estratégico del país.

La creación de un modelo masivo de lenguaje (LLM) para el español, auspiciado por el gobierno mexicano, es necesario por varias razones. En primer lugar, garantiza la soberanía tecnológica de México, reduciendo la dependencia de modelos extranjeros y permitiendo el desarrollo de herramientas adaptadas a nuestro contexto cultural y lingüístico, algo que es fundamental para la seguridad nacional y el desarrollo estratégico del país.





Four horizontal lines for text input.

Este proyecto salvaguardaría la riqueza y diversidad del español, asegurando no solo la precisión en las traducciones, sino también preservando los matices, matices y expresiones propias de la cultura mexicana en el ámbito tecnológico. Además, establecería estándares para el uso del idioma en tecnologías avanzadas, manteniendo su capacidad expresiva y adaptabilidad a nuevos contextos.

Un proyecto de esta magnitud no solo representa una inversión en tecnología, sino en la identidad y el futuro de México, colocándolo como líder en IA en Latinoamérica y garantizando que el desarrollo tecnológico se alinee con los valores legales, culturales y de género y con las necesidades específicas para el desarrollo integral del país.



ULISES CORTÉS

Catedrático de Inteligencia Artificial de la Universitat Politècnica de Catalunya. Coordinador Científico del grupo High-Performance Artificial Intelligence del Barcelona Supercomputing Center. Miembro del Observatori d'Ètica en Intel·ligència Artificial de Catalunya y del Comitè d'Ètica de la Universitat Politècnica de Catalunya. Es miembro del comité ejecutivo de EurAI. Participante como experto de México en el grupo de trabajo Data Governance de la Alianza Global para la Inteligencia Artificial (GPAI). Doctor Honoris Causa por la Universitat de Girona.

Además, un LLM mexicano ayudaría a preservar la riqueza del español de México, con sus 68 lenguas indígenas reconocidas y sus variantes regionales únicas. Como dice la experta en lingüística, Concepción Company Company, "la lengua es un vehículo esencial para la identidad cultural". Desde la vertiente económica, ayudaría a impulsar sectores clave como la educación, el comercio y las industrias creativas, fortaleciendo la competitividad de México en la economía digital global, en particular, en el espacio hispanoparlante. Desde el punto de vista tecnológico facilitaría el acceso a herramientas basadas en la IA para empresas y ciudadanos, reduciendo la brecha tecnológica y promoviendo la inclusión digital.

La investigación y el desarrollo tecnológico llevados a cabo dentro de México, representan una oportunidad invaluable para impulsar la innovación y el crecimiento económico del país. Al invertir en la creación de modelos masivos del lenguaje propios, se fomenta la generación de empleo de alta tecnología, se atrae inversión extranjera y se fortalece la posición de México como un actor relevante en la economía digital global. El desarrollo de un LLM, patrocinado por el gobierno, para el español es un paso necesario hacia la soberanía tecnológica, que permitirá a México competir en mejores condiciones en el mercado internacional de las aplicaciones basadas en la IA.

