Bla, bla, bla

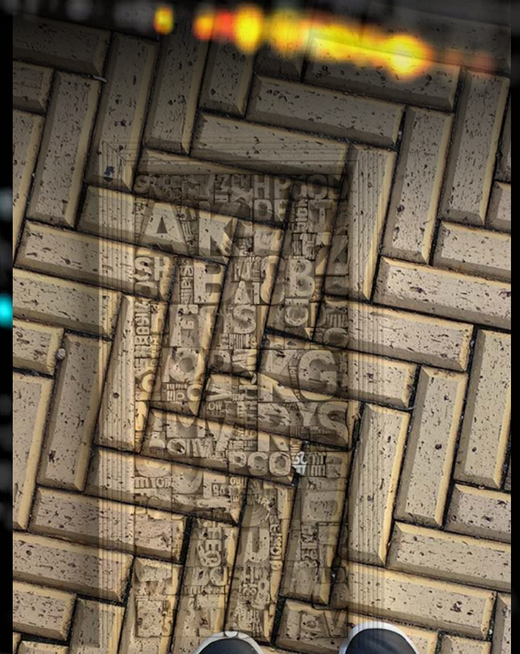
Ulises Cortés

"Ha llegado la hora", dijo la morsa, "de hablar de muchas cosas: de zapatos y barcos, y de lacre de sellar, de coles y de reyes."[1]
L. Carroll

La frase del epígrafe pertenece al poema de Lewis Carroll, "La morsa y el carpintero", de su obra *A través del espejo y lo que Alicia encontró allí*. En el poema, la morsa usa este discurso caprichoso y ridículo para distraer a un grupo de jóvenes ostras hablándoles de temas aleatorios —"*zapatos y barcos, y lacre de sellar, coles y reyes*"—, antes de que ella y el carpintero se las coman.

Este discurso tan voluble y alocado recuerda el revuelo que provocan hoy los modelos masivos del lenguaje en la prensa y en la opinión pública. Nos fascinan y desconciertan a partes iguales por su aparente elocuencia, su estridencia y su capacidad de decir casi cualquier cosa sobre casi cualquier tema con naturalidad convincente. Entre detractores y partidarios, el debate sobre su utilidad se polariza. Pero, como señala Gary Marcus, estos sistemas *no comprenden* los textos que generan: solo reproducen patrones de lenguaje, entretejen modelos de lenguaje aprendidos de nosotros, sin anclar sus palabras en una comprensión del mundo real. Son incapaces de capturar la *frónesis*, esa sabiduría práctica que guía la acción humana.

La verdadera inteligencia requiere no solo lenguaje, necesita un cuerpo y, también, una forma de conocimiento estructurado que conecte representación, contexto y propósito; la inteligencia requiere haber experimentado la vida.



Tal vez por eso, en medio de tanta información desbordada, tan gratuita como omnipresente, corremos el riesgo de convertirnos, como las ostras jóvenes, en audiencia fascinada y distraída por el torrente de paparruchas generadas por máquinas como Claude[2] o ChatGPT, sin advertir del todo las dinámicas de poder y dependencia que se esconden detrás de su aparente *magia*. Nos preparan para ser felizmente engullidos. Esta oratoria sintética tiene el peligro de pasteurizar la expresión escrita y, de alguna manera, limitar el espectro vocabular del escribiente, ya que cedemos espacio en nuestra comunicación habitual a un discurso algorítmico que se multiplica, nos entretiene y, en cierto modo, nos adormece.

Las arquitecturas de los modelos masivos han ido sufriendo una muy rápida evolución, aunque siguen apostando en gran medida en el escalado como receta hacia la inteligencia artificial general (AGI). Sin embargo, se comienza a notar un cierto estancamiento[3], a veces relacionado con la posibilidad de agotar los datos e imágenes disponibles, otra veces asociado con el ingente coste de entrenamiento y depuración del conocimiento, y, poco a poco, se vuelve la mirada hacia arquitecturas híbridas en las que hay una componente de inteligencia artificial simbólica sólida, *e.g.* las reglas explícitas, las lógicas, las ontologías, los frames o los grafos de conocimiento para representar conceptos, relaciones, variables y causalidad de forma estructurada e inteligible[4]. Por otro lado, aparecen nuevas arquitecturas, como, por ejemplo, la propuesta por Y. LeCun llamada *Joint Embedding Predictive Architecture* (JEPA)[5,6], dispuestas a convertirse en los nuevos depredadores en el dominio de los modelos masivos y aspirando a dejar a estos en el camino de la extinción.

También los usos y adopción de tales modelos masivos del lenguaje se han ido expandiendo a un ritmo vertiginoso; ya existen pocas áreas en los que no haya alguna aplicación y, lo que es peor, ninguna de ellas está certificada. En el caso de la Unión Europea, además, la existencia de la regulación no ha frenado de ninguna manera a míriadas de usuarios dispuestos emplear estos instrumentos sin prevención alguna ni reflexión crítica. El problema reside en que nuestros métodos actuales de regulación de modelos avanzados basados en la inteligencia artificial, que son enormes y complejos, están colapsando, y no tenemos a mano herramientas útiles para contenerlos.

Esto, en la práctica, significa delegar decisiones sensibles en sistemas que no comprendemos y que no están certificados por ninguna autoridad y son programados por individuos a quienes no se les puede exigir ninguna responsabilidad legal. Esto equivale a aceptar un experimento a gran escala con la sociedad, donde los efectos sobre el ambiente, la democracia, la desinformación o el empleo se descubren *sobre la marcha*, y cuyos efectos son difíciles de paliar. Está claro que en estas circunstancias hay que combinar innovación con prudencia, estableciendo límites claros a los usos de *alto riesgo* y exigiendo estándares mínimos de transparencia, trazabilidad y supervisión humana efectiva antes de normalizar estos sistemas basados en las tecnologías de la IA en contextos críticos.

Esta necesidad de equilibrar innovación y prudencia cobra especial urgencia dentro del contexto de la proliferación masiva de grandes modelos de lenguaje (LLM) en escenarios bélicos y de vigilancia poblacional. En operaciones militares, estas herramientas se emplean para análisis predictivo de amenazas, generación de estrategias tácticas y procesamiento en tiempo real de inteligencia, lo que amplifica riesgos de sesgos algorítmicos y decisiones autónomas opacas[7]. De igual modo, en sistemas de seguimiento ciudadano —como reconocimiento facial o perfiles predictivos de comportamiento—, su adopción sin supervisión humana adecuada amenaza la privacidad y fomenta abusos autoritarios[8].

En operaciones militares, estas herramientas se emplean para análisis predictivo de amenazas, generación de estrategias tácticas y procesamiento en tiempo real de inteligencia, lo que amplifica riesgos de sesgos algorítmicos y decisiones autónomas opacas . De igual modo, en sistemas de seguimiento ciudadano —como reconocimiento facial o perfiles predictivos de comportamiento—, su adopción sin supervisión humana adecuada amenaza la privacidad y fomenta abusos autoritarios.

Antrophic ha establecido restricciones éticas estrictas (conocidas en el mundo tecnológico como *guardrails*) en *Claude* para prevenir su uso en escenarios bélicos sensibles, como el desarrollo de armas autónomas letales, la facilitación directa de la violencia o la vigilancia masiva doméstica. En días recientes, *Claude*[9],en su versión más moderna, *Claude Opus* 4.6, ha adquirido una relevancia especial al prohibir *Antrophic* su uso por el Departamento de Defensa norteamericano para sus operaciones bélicas. Este movimiento estuvo motivado al conocerse, a través de la prensa, que *Claude* fue usado en combinación con el Maven Smart System[10], de la empresa Palantir, para fijar las coordenadas de objetivos militares y crear los órdenes de prioridad en los ataques que luego fueron usados por los ejércitos de los Estados Unidos e Israel en sus ataques, incluidos aquellos a objetivos civiles erróneos.

La distopía de las armas letales autónomas ya está aquí, se experimenta a diario contra los gazatíes, los libaneses y los iraníes. Esta problemática ilustra la tensión entre avance tecnológico y su control legal y ético. Y más allá de esas leyes, está la práctica diaria: "En cierto modo, hay que confiar en que las fuerzas armadas hagan lo correcto", aseveró el director de tecnología del Pentágono, Emil Michael, en una entrevista con *CBS News*[11]. Así que cabe preguntarnos: ¿Podrán las leyes y la ética seguir el ritmo de los algoritmos? Las posibles respuestas son estremecedoras. El problema no está en la tecnología, sino en quien la desarrolla y, luego, la libera sin sistemas de seguridad, poniéndola en manos de quien la financia.

"Ha llegado la hora", dijo la morsa, "de hablar de muchas cosas: de zapatos y barcos, y de lacre de sellar, de coles y de reyes".

El futuro, una vez más, nos rebasa.

- 1 Carroll, Lewis. A través del espejo y lo que Alicia encontró allí. Contiene el poema «La morsa y el carpintero»
- 2 Anthropic. (2024).Claude 3.5 Sonnet (versión 1). https://www.anthropic.com/claude
- 3 The 2025 AI Index Report, https://hai.stanford.edu/ai-index
  The 2025 AI Index Report
- 4 Marcus, G., and Davis, E..(2018) Rebooting AI: Building artificial intelligence we can trust. Vintage.
- 5 Assran, M. et al Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture (2023). https://huggingface.co/papers/2301.08243
- 6 What is JEPA. https://www.turingpost.com/p/jepa
- 7 Russell, S. (2019).Human Compatible: Artificial Intelligence and the Problem of Control. Viking.
- 8 Zuboff, S. (2019).The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. PublicAffairs.
- 9 https://www.nytimes.com/2026/02/27/us/politics/anthropic-military-ai.html
- 10 https://blog.palantir.com/maven-smart-system-innovating-for-the-alliance-5ebc31709eea
- 11 https://www.cbsnews.com/news/anthropic-claude-ai-iran-war-u-s/

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.