



JOSÉ CASADO.
Ingeniero de Telecomunicación.

Inteligencia Artificial Modelos de lenguaje versus modelos del mundo

Durante los últimos años, la IA se ha presentado como una tecnología destinada a comprender el mundo, asistir a las personas y tomar decisiones cada vez más sofisticadas. Sin embargo, bajo esa narrativa optimista se esconde una realidad más incómoda: los sistemas dominantes no entienden el mundo que describen y, al mismo tiempo, **se están utilizando principalmente para extraer información sobre quienes los usan**. El problema no es coyuntural ni se resolverá automáticamente con más datos o más potencia de cálculo, porque responde a limitaciones estructurales y a incentivos económicos muy concretos.

Yann LeCun es un científico de computación francoestadounidense ampliamente reconocido como uno de los 'padrinos' de la inteligencia artificial». En 2018, recibió el premio A. M. Turing

de la ACM por sus contribuciones fundamentales al *deep learning*). LeCun afirmó que los modelos de lenguaje actuales LLMs operan casi exclusivamente sobre texto. No perciben, no expe-

rimentan ni interactúan directamente con la realidad física o social.

Todo lo que 'saben' les llega filtrado por descripciones humanas: libros, artículos, conversaciones, publicaciones en redes o documentos. Ese material no es el mundo, sino una representación parcial y profundamente mediada de él. El lenguaje arrastra errores, sesgos culturales, malentendidos, propaganda y ficciones. Funciona como una sombra

Los sistemas dominantes no entienden el mundo que describen y, al mismo tiempo, se están utilizando principalmente para extraer información sobre quienes los usan

proyectada de la realidad, no como la realidad misma. Por muy grande que sea el volumen de texto procesado, el resultado sigue siendo una acumulación de representaciones indirectas, no de experiencia.

Esta limitación tiene consecuencias claras. Los sistemas basados en predicción estadística del lenguaje pueden generar discursos coherentes y convincentes, pero carecen de un anclaje sólido en la causalidad, las restricciones físicas o las consecuencias reales de las acciones. De ahí que produzcan respuestas plausibles pero falsas, o que fallen de forma inesperada en situaciones que requieren comprender cómo funciona el mundo más allá de cómo se habla de él. No se trata de errores anecdóticos, sino de un efecto directo de su arquitectura: optimizar la probabilidad de la siguiente palabra no equivale a razonar sobre la realidad.

Durante años se ha asumido que este problema se corregiría con escala: más datos, modelos más grandes, más parámetros. Pero aumentar el tamaño de un sistema que solo ve sombras no lo acerca al mundo real; simplemente le permite describir esas sombras con mayor fluidez. La mejora cuantitativa no soluciona un déficit cualitativo. Por eso empieza a ganar peso un enfoque distinto, basado en sistemas que construyen representaciones internas del entorno, aprenden de la interacción, tienen memoria persistente, y pueden simular escenarios antes de actuar. En estos modelos, el lenguaje no desaparece, pero deja de ser el núcleo del sistema para convertirse en una interfaz que traduce entre humanos y mecanismos que sí están anclados en dinámicas reales.

Este contraste entre apariencia y comprensión técnica se refleja también en cómo se despliega la IA en productos

de consumo. Muchas aplicaciones se presentan como asistentes fiables o agentes autónomos, pero en la práctica resultan erráticas para tareas simples. Se normaliza así una experiencia imperfecta bajo la promesa de que el sistema ‘aprende’, cuando en realidad se está pidiendo al usuario que tolere el fallo como estado permanente. La paradoja es que, mientras la utilidad es limitada, la capacidad de capturar información personal es extraordinariamente eficaz.

Aquí aparece la otra cara del problema. Los sistemas actuales destacan menos por ayudar que por observar. Para funcionar, reclaman acceso a conversaciones, documentos, correos, hábitos, rutinas y contexto vital.

Ese acceso no es un efecto secundario: es el activo central. El valor económico no reside tanto en ejecutar bien una tarea como en perfilar al usuario con precisión creciente. La IA se convierte así en una herramienta para inferir quién eres, cómo decides, qué te motiva y en qué momento eres más vulnerable.

Este paso del registro de comportamiento a la inferencia profunda marca una diferencia cualitativa. No se trata solo de saber qué haces, sino de anticipar qué harás y cómo influirte. Cuando un sistema aprende qué estímulos funcionan contigo, puede ajustar mensajes, recomendaciones o decisiones de forma personalizada y opaca.

La frontera entre asistencia y manipulación se vuelve difusa. Por no hablar de potenciales escenarios de modelos que incorporen publicidad. Ya no hace falta convencer a una persona con un argumento general; basta con pulsar la tecla emocional adecuada en el momento oportuno.

La tendencia se intensifica con la incorporación de agentes capaces de actuar

en nombre del usuario. Al atravesar servicios críticos (banca, sanidad, administración, trabajo...) estos sistemas se convierten en canales privilegiados de datos sensibles. La promesa de comodidad se paga con una concentración inédita de información íntima en infraestructuras opacas, a menudo sin garantías claras sobre qué se recoge, qué se infiere y con quién se comparte. La autonomía técnica amplifica el riesgo: cuanto más puede hacer un sistema por sí solo, mayor es el impacto de un diseño defectuoso o de incentivos mal alineados.

El resultado es una tecnología que avanza en una dirección muy concreta. Por un lado, sigue sin comprender el mundo que describe, porque confunde lenguaje con conocimiento. Por otro, se vuelve cada vez más eficaz modelando a las personas, porque ahí sí existe un incentivo económico claro. La IA no está ‘aprendiendo’ en un sentido humano o cognitivo; está afinando mecanismos de extracción, predicción y adaptación conductual.

Nada de esto es inevitable. No es una ley de la naturaleza, sino la consecuencia de decisiones técnicas y de negocio. Existen alternativas: arquitecturas que integran modelos del mundo reales, sistemas diseñados para minimizar la captura de datos, marcos regulatorios que prioricen derechos fundamentales. Pero avanzar en esa dirección exige abandonar dos confusiones persistentes: que hablar bien equivale a entender y que ofrecer asistencia justifica cualquier nivel de vigilancia.

Si no se corrigen estos supuestos, el futuro próximo será coherente con el presente: herramientas que prometen ayuda mientras consolidan dependencia, que exhiben inteligencia superficial mientras profundizan en el conocimiento íntimo de las personas, y que normalizan la cesión de privacidad a cambio de beneficios marginales. El verdadero desafío no es hacer que la IA suene más convincente, sino decidir si queremos que entienda el mundo o que nos modele a nosotros. ▀

El verdadero desafío no es hacer que la IA suene más convincente, sino decidir si queremos que entienda el mundo o que nos modele a nosotros