



► El proyecto fue desarrollado bajo una colaboración de más de 100 profesionales y 60 instituciones de 15 países.

# LatamGPT: presentan modelo de inteligencia artificial entrenada con datos latinoamericanos y hecha en Chile

Francisco Corvalán

Este martes se realizó el lanzamiento oficial de LatamGPT, el primer modelo de lenguaje abierto desde y para América Latina y el Caribe, con datos, idiomas y contextos propios de la región. El proyecto, coordinado por el Centro Nacional de Inteligencia Artificial (CENIA) de Chile, marca un hito y posiciona a esta parte del continente no solo como usuaria, sino como creadora de inteligencia artificial (IA).

LatamGPT surgió, según expresaron sus responsables, como respuesta a una brecha estructural en el desarrollo global de modelos de lenguaje y la subrepresentación histórica del español y el portugués en estos motores de IA, que concentran cerca del 4% y el 2% de los datos utilizados para entrenar estos sistemas. Junto a eso está la ausencia de contextos culturales, sociales e históricos latinoamericanos en la arquitectura de los modelos dominantes. Frente a ese escenario, el nuevo modelo fue entrenado desde su origen con información latinoamericana, permitiéndole comprender mejor sus matices lingüísticos, memorias colectivas e identidades.

El proyecto fue desarrollado bajo una colaboración de más de 100 profesionales y 60 instituciones de 15 países. En su construcción participaron el Ministerio de Ciencia de Chile, el Banco de Desarrollo de América Latina y el Caribe, Amazon Web Services y el centro tecnológico Data Observatory, junto a universidades, organismos públicos, centros de investigación y aliados internacionales. El resultado es un modelo concebido como bien público, con código abierto, orientado a democracia

**Se trata del primer modelo** de lenguaje abierto entrenado con datos latinoamericanos, el cual busca reducir brechas históricas en IA y fortalecer la soberanía tecnológica regional, con foco en cultura, idiomas y decisiones públicas.

tizar el acceso a la IA y habilitar su uso en educación, gestión pública, investigación e innovación productiva.

Para el director del CENIA, Álvaro Soto, la motivación de fondo del proyecto es estratégica. "La IA es una tecnología disruptiva, nos va a cambiar la vida, y dejar que esa tecnología se haga en otra parte del mundo no es una opción", afirmó. A su juicio, depender completamente de desarrollos externos en un contexto geopolítico incierto es "un poco ingenuo", por lo que resulta clave que la región "levante las capacidades para estar en posición de opinar y decidir sobre esta tecnología".

Soto subrayó que la noción de soberanía tecnológica no implica aislamiento, sino capacidad efectiva de participación. "Significa levantar capacidades locales para sentarse a la mesa, tener propiedad y poder opinar sobre la inteligencia artificial". En esa línea, advirtió que los modelos de punta actuales saben mucho más de otras partes del mundo que de nuestra región.

Desde el punto de vista técnico, LatamGPT fue construido sobre la arquitectura base Llama 3.1, con 70 mil millones de parámetros, y entrenado mediante un proceso de pre y post entrenamiento.

Rodrigo Durán, gerente del CENIA, enfatizó que LatamGPT no busca competir, sino complementar a otros modelos exis-

tentes. "Sabemos que nuestro modelo está por debajo del rendimiento en tareas de razonamiento complejo, pero la virtud que tiene es que actúa mejor en un criterio sobre América Latina".

En ese sentido, sostuvo que el modelo "responde con mayor exactitud y alucina menos" cuando se le consulta por cultura, políticas públicas o literatura latinoamericana, lo que abre oportunidades concretas en educación, investigación académica y toma de decisiones públicas. Para los gobiernos, añadió, el carácter abierto del modelo es clave. "Permite mayor trazabilidad, lo que genera legitimidad y confianza en la ciudadanía", aclaró.

El director ejecutivo del Data Observatory, Rodrigo Roa, destacó que uno de los principales desafíos fue levantar la infraestructura necesaria para entrenar y ajustar el modelo. "Estamos creando un repositorio nacional de datos para que este modelo siga entrenándose y siga su camino hacia el futuro", señaló, y agregó que el valor estratégico no está solo en el modelo, sino en su base: los datos. "Cómo el dato se transforma en información y cómo, con esa información, podemos tomar decisiones a nivel político y estratégico".

Roa también relevó la eficiencia del proyecto, ya que LatamGPT se desarrolló con un presupuesto aproximado de 2,5 millo-

nes de dólares, muy por debajo de lo que destinan las grandes compañías tecnológicas, que puede ser entre 10 y 20 veces más. Aunque advirtió que para escalar el proyecto será necesario seguir combinando inversión pública y privada.

El ministro de Ciencia, Aldo Valle, definió a esta iniciativa como una "de carácter político y diplomático", orientada a fortalecer la cooperación regional. "Si no cooperamos, tenemos muy pocas posibilidades de alcanzar los niveles de desarrollo que exige la inteligencia artificial". También espera que la próxima ministra, Ximena Lincolao, siga impulsando iniciativas como estas.

Durante el lanzamiento de esta tecnología, el Presidente Gabriel Boric enmarcó el hito en un contexto geopolítico más amplio. "Si no estamos en la mesa, estamos en el menú", dijo, citando al primer ministro de Canadá, Mark Carney, advirtiendo que en IA "el punto de partida importa mucho", porque el crecimiento es exponencial. A su juicio, LatamGPT responde a una preocupación de fondo: cómo la reproducción de información moldea la cultura y puede reforzar hegemonías históricas.

Boric ejemplificó la asimetría de información entre Europa y América Latina en los modelos actuales, y sostuvo que LatamGPT es clave para revertirla. Por ello, espera que pronto pueda existir como aplicación para consultar interrogantes referidas a la región. El programa aún está en marcha blanca y los creadores dicen que todavía falta para que pueda descargarse en Android o iOS, puesto que primero es necesario contar con la infraestructura necesaria para entregar esta gran cantidad de información a todos los usuarios. ●