

**Hola, Neo.** Reserva hoy tus unidades

MacBook Neo

LenovoTU NEGOCIO TIENE GRANDES IDEAS.
IMPÚLSALAS CON LA TECNOLOGÍA LENOVO.Encuentra en Intcomex
las soluciones Lenovo que se adaptan
a cada desafío de tu negocio.

INTCOMEX

22/09/2026

Auditar antes de acelerar: El control que necesitan los asesores, algoritmos y agentes de IA

Recomendar

Twitter



José Lagos Docente UEjecutivos Facultad de Economía y Negocios Universidad de Chile.

Por José Lagos Docente UEjecutivos Facultad de Economía y Negocios Universidad de Chile.

El desarrollo de inteligencia artificial ha entrado en una etapa en la que avanzar más rápido ya no significa necesariamente avanzar mejor. En su artículo *We Must Pace the Frontier*, Dario Amodei, director ejecutivo de Anthropic, sostiene que las capacidades de los modelos están progresando a una velocidad que amenaza con superar nuestra capacidad para comprenderlos, alinearlos y controlarlos. Su propuesta no consiste en detener la innovación, sino en acompañar el aumento de capacidades con mecanismos verificables de seguridad, evaluación independiente y supervisión.

Este planteamiento también resulta aplicable a cualquier organización que diseñe asesores de IA generativa, Gems, Skills, algoritmos de machine learning o agentes capaces de ejecutar acciones. En este contexto, la organización debe demostrar que funciona dentro de límites aceptables y la función de auditoría se transforma en una condición para avanzar.

Las soluciones de IA no presentan la misma exposición. Un asistente que redacta borradores tiene un impacto diferente al de un modelo que rechaza créditos o un agente que modifica registros. Sin embargo, todos requieren evidencia sobre su propósito, datos, instrucciones, comportamiento, controles y responsables.

En este caso, la auditoría debe comenzar antes de producción. Esto implica definir el caso de uso, los usuarios autorizados, las decisiones permitidas, los datos procesados y las acciones prohibidas. También exige métricas, pruebas de aceptación, trazabilidad, monitoreo y criterios de suspensión. Acelerar sin estas definiciones genera deuda técnica, ética, regulatoria y operacional.

Los asesores de IA generativa suelen percibirse como aplicaciones simples porque operan mediante instrucciones y lenguaje natural. No obstante, pueden revelar información personal o confidencial, entregar respuestas incorrectas, reproducir sesgos, citar fuentes inexistentes o inducir decisiones erróneas. Su auditoría debe revisar el propósito y alcance; la calidad y seguridad de las instrucciones; las fuentes; los permisos; la retención de conversaciones; la resistencia frente a prompt injection; y las salvaguardas para contenidos sensibles. El usuario debe saber que interactúa con una IA, conocer sus limitaciones y poder escalar el caso a una persona.

En este aspecto, no basta con revisar el prompt principal. Es necesario probar el comportamiento mediante escenarios normales, extremos y adversarios. Esto incluye las solicitudes ambiguas, intentos de extraer instrucciones internas, carga de datos personales, peticiones contrarias a la ética y preguntas fuera del dominio autorizado. El resultado debería documentarse en una ficha que identifique riesgos, controles, pruebas, limitaciones, propietario y versión.

Por el contrario, la auditoría de algoritmos de machine learning debe cubrir obtención de datos, preparación, entrenamiento, validación, aprobación, implementación y monitoreo. Un modelo puede mostrar buena exactitud general y, simultáneamente, perjudicar a determinados grupos o fallar en los casos de mayor impacto. Por eso se debe evaluar la calidad y representatividad de los datos, la legitimidad de las variables, la fuga de información y el uso de proxies de características sensibles. Las métricas deben responder al objetivo y costo del error: precisión, recall, F1, AUC, falsos positivos y negativos, además de indicadores de equidad. La auditoría también debe verificar reproducibilidad, versionamiento, explicabilidad, aprobación independiente y monitoreo de drift. Cuando cambian el negocio o la población, un modelo válido puede degradarse silenciosamente. Deben existir umbrales de alerta, recalibración y retiro, junto con revisión humana para decisiones de alto impacto.

Finalmente, el desarrollo e implementación de agentes de IA introducen un riesgo adicional: planifican, utilizan herramientas, consultan sistemas y ejecutan acciones. Un error puede propagarse rápidamente entre correos, bases de

**Edición
Agosto 2026**
Revisela aquí

Noticias Relacionadas

**56% de los chilenos está seguro de que detectaría un contenido falso generado por IA** (22/09/2026)**Ingram Micro celebró las tradiciones chilenas junto a sus asociados en Providencia y Lo Boza** (21/09/2026)**ASRock presenta AI Center para impulsar la implementación de IA en las empresas** (21/09/2026)**Encuentro aborda los desafíos del phishing y la IA para las industrias** (21/09/2026)

datos, plataformas financieras o infraestructuras críticas.

La auditoría debe reconstruir la cadena completa: objetivo, razonamiento operativo, herramientas, credenciales, acciones y resultados. Debe aplicarse mínimo privilegio, separar ambientes, limitar operaciones, exigir aprobación humana en acciones críticas e incorporar detención inmediata. Los registros deben explicar quién autorizó al agente, qué información utilizó y por qué actuó. Las pruebas adversariales deben considerar desviación del objetivo, instrucciones maliciosas externas, uso indebido de herramientas, coordinación inesperada y persistencia no autorizada. También se requieren límites de tiempo, costo, acciones y sistemas accesibles.

La tesis de Amodei conduce a una conclusión empresarial clara: la velocidad de implementación debe estar condicionada por la capacidad de control. A mayor autonomía, impacto y acceso a información o sistemas, mayor debe ser la profundidad de evaluación independiente.

Una política de IA es insuficiente si no se traduce en inventarios, responsables, evaluaciones de impacto, pruebas, monitoreo y evidencia. Auditar no significa frenar la innovación, sino impedir que avance a ciegas. La organización madura no es la que implementa más IA, sino aquella que puede demostrar qué utiliza, para qué, bajo qué límites y quién responde cuando algo falla. En la nueva frontera de la inteligencia artificial, gobernar el ritmo será tan importante como dominar la tecnología. La auditoría constituye el mecanismo que permite convertir prudencia en evidencia y confianza en una propiedad verificable del sistema.

Indice de Noticias



Contáctenos

📍 **Dirección:** José Manuel Infante
919, Of. 203,
Providencia, Chile

☎ **Teléfono:** (562) 2433 5500

✉ **Email:** info@emb.cl

Visite también:

EMB CONSTRUCCIÓN

GERENCIA

HSEC

INDUSTRIA

NG LOGISTICA