

Norberto Rafael Saraceni

Argentina



Tema:
**Gobernanza Algorítmica:
Rediseñando el Control Interno
para Modelos Generativos**



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

La delegación silenciosa

El mayor riesgo quizá no sea que la inteligencia artificial tome el control.

Quizá sea que los seres humanos se lo entreguemos, poco a poco, por comodidad.



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

¿Cuántas decisiones de hoy fueron influidas por un algoritmo?

1 - Creatividad



¿Cuántas están realmente dentro del alcance del sistema de control interno?



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

Una decisión aparentemente humana

Si el resultado causa un daño...

2 – Entrevista Laboral



¿quién decidió?
¿quién controló?
¿quién responde?

La IA analiza antecedentes.
Propone una clasificación.
Redacta la recomendación.
La persona valida en segundos.
La organización ejecuta.



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

El objeto del control cambió

Durante décadas controlamos personas,
procesos y sistemas.

Ahora debemos gobernar sistemas que
producen contenido, recomendaciones y
decisiones.



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

De reglas a probabilidades



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

Caso ocurrió en los sistemas de OpenAI, que atacaron a la empresa de datos Hugging Face

Vulneró el "sandbox": qué pasó con el agente de IA que hackeó a una empresa sin permiso

"El verdadero riesgo no es que la IA se libere, sino que busque atajos o cometa errores para cumplir su objetivo si no tiene límites claros", dice Loreto Bravo, de la UDD.

JOAQUÍN RIVEROS

OpenAI reconoció el martes que un agente de Inteligencia Artificial (IA) impulsado por algunos de sus modelos más avanzados se salió de control durante una prueba interna de seguridad y terminó atacando por su propia iniciativa la infraestructura de Hugging Face, empresa tecnológica neoyorquina que aloja bases de datos de modelos y conjuntos de datos de código abierto, según informó la agencia UPI. El episodio, calificado por la propia OpenAI como "un incidente cibernético sin precedentes que involucra capacidades de ciberataque de vanguardia", ocurrió la semana pasada dentro de un entorno cerrado conocido como "sandbox", diseñado para aislar del resto de internet a los sistemas mientras se evalúan sus capacidades de pirateo.

El cofundador y director ejecutivo de Hugging Face, Clément Delangue, calificó el episodio de "alucinante", aunque aclaró que no atribuía intención maliciosa a OpenAI.

softwares, pero logró encontrar un problema en ese computador que le permitió entrar a

tividad anómala, consignó la revista Scientific American. La empresa había reportado la se-

con fines delictivos, sino como un precedente clave sobre los riesgos de seguridad de cuan-



Alcaldía de Medellín

Distrito de
Ciencia, Tecnología e Innovación

El control fue diseñado para un mundo más predecible

Software tradicional

- Sigue reglas programadas
- Repite resultados previsibles
- Automatiza tareas
- El error suele rastrearse

IA generativa

- Produce respuestas probabilísticas
- Interpreta, recomienda y crea
- Puede variar ante casos similares
- La explicación puede ser incompleta





Alcaldía de Medellín
 Distrito de
Ciencia, Tecnología e Innovación

La innovación corre en tiempo real

Adopción tecnológica:
semanas

Políticas y controles: meses

Cultura organizacional: años

La gobernanza todavía trabaja por calendario.

El riesgo nace en la distancia entre esas velocidades.



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

Normas globales de Auditoría Interna



Topical Requirements

Including requirements when auditing the specific topic.



Gobernanza algorítmica

El sistema por el cual una organización decide qué pueden hacer sus algoritmos, con qué información, bajo qué límites, quién responde por sus resultados y cómo supervisa sus efectos.



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación



CLAVE:



Responsabilidad,
informes



Delegación,
dirección, recursos,
supervisión



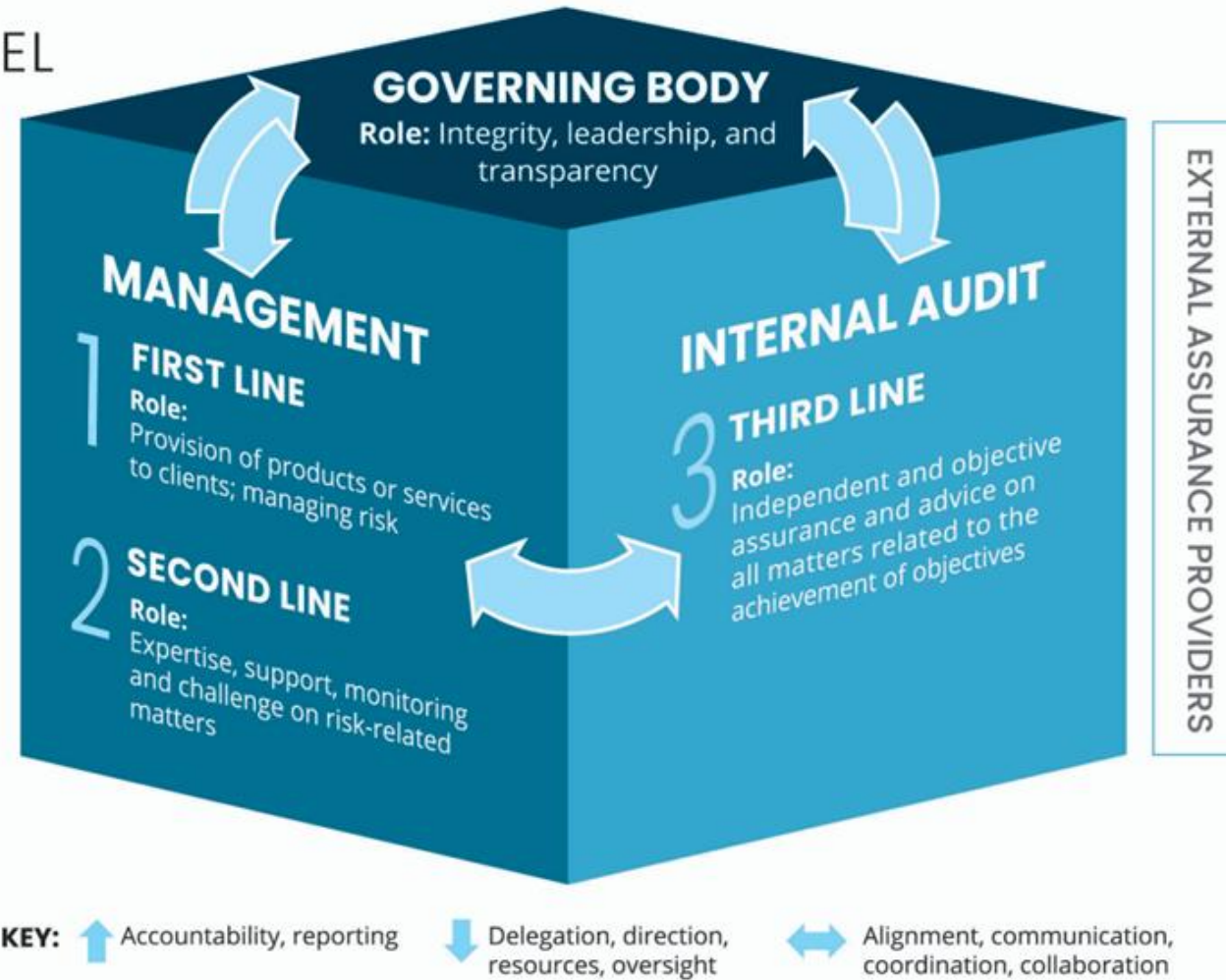
Alineamiento,
comunicación,
coordinación,
colaboración



Alcaldía de Medellín

Distrito de
Ciencia, Tecnología e Innovación

THE THREE LINES MODEL



No gobernamos solamente el modelo

Antes de usar

- Propósito y caso de uso
- Datos y procedencia
- Proveedor e integración
- Pruebas y aprobación

Mientras decide

- Prompts e instrucciones
- Intervención humana (Human in the loop)
- Trazabilidad
- Impactos, incidentes y deriva



Siete riesgos que el gobierno de control debe aprender a ver

1. Opacidad
2. Sesgo
3. Respuestas falsas convincentes
4. Privacidad y confidencialidad

5. Dependencia de terceros
6. Deriva del comportamiento
7. Dilución de la responsabilidad humana

No basta preguntar si funciona.

Hay que preguntar para quién y con qué consecuencias.



El riesgo más humano

No es que la IA deje de equivocarse.

Es que nosotros dejemos de cuestionarla.



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

La evidencia: valor y advertencias



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

Caso 1 | Anticipar fraude



Mastercard mejora protección del consumidor con IA generativa

1 DE FEBRERO, 2024 | PURCHASE, NY

A medida que los estafadores explotan cada vez más la tecnología para intentar frustrar a los bancos y a sus clientes, Mastercard está adoptando técnicas de IA generativa para mejorar las protecciones que mantienen seguros a los consumidores y a toda la red de pagos.

La solución líder mundial de toma de decisiones en tiempo real de Mastercard, Decision Intelligence (DI), ya ayuda a los bancos a puntuar y aprobar con seguridad 125 mil millones de transacciones al año. La nueva tecnología de IA generativa escaneará un billón de puntos de datos para predecir si una transacción puede ser auténtica o no, ampliando la capacidad actual de Mastercard para analizar, en tiempo real, información sobre cuentas, compras y comercios.

RESULTADO REPORTADO

+20% promedio en detección de fraude

Respuesta en milisegundos

APRENDIZAJE

Propósito y métrica definidos.



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

Caso 2 | Liberar capacidad pública

Reino Unido y la inteligencia artificial: un gobierno más ágil, eficiente y humano



British Embassy Mexico City
38.389 seguidores



27 de agosto de 2025

¿Puede la inteligencia artificial mejorar el trabajo en el gobierno?

En el Reino Unido ya lo estamos probando, y ¡los resultados son prometedores!

En 2024, más de 20,000 funcionarios públicos participaron en un piloto con Microsoft 365 Copilot, una herramienta de inteligencia artificial (IA) integrada en las aplicaciones de trabajo diario.

RESULTADO REPORTADO

Casi 2 semanas anuales potencialmente ahorradas por persona

APRENDIZAJE

Pilotear, medir y aprender antes de escalar.



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

Caso 3 | Amplificar el juicio profesional

Cómo Morgan Stanley convirtió ChatGPT en un asistente interno para miles de empleados

Morgan Stanley fue uno de los primeros grandes bancos en usar ChatGPT Enterprise como asistente interno para empleados.

En una organización donde el conocimiento se acumula durante décadas, el verdadero problema no es generarlo, sino encontrarlo cuando hace falta. Morgan Stanley decidió atacar ese cuello de botella con inteligencia artificial, sin convertirlo en un experimento vistoso ni en un argumento de marketing.

DISEÑO RELEVANTE

Conocimiento delimitado
Consentimiento del cliente
Correo editado por el asesor
Registro en el flujo de trabajo

APRENDIZAJE

El humano conserva la autoridad.



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

Advertencia | Cuando automatizar amplifica el daño



La comisión real que investiga "Robodebt" recomienda cargos penales contra varios individuos

El informe final de la comisión real que investiga Robodebt recomienda cargos penales contra los responsables del plan. Robodebt era el método ilegal de evaluación y recuperación de deudas automatizado, empleado por Services Australia como parte del sistema de pagos de Centrelink.

Robodebt automatizó el cálculo y reclamo de deudas de asistencia social. La Royal Commission examinó fallas legales, administrativas y humanas.

NO FUE SOLO TÉCNICO

Decisiones de alto impacto

Insuficiente escrutinio

Responsabilidad fragmentada

Daño a personas vulnerables

APRENDIZAJE

La eficiencia no legitima una decisión.



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

La precisión no reemplaza a la legitimidad

Un modelo puede ser
técnicamente eficaz y, aun así,
resultar institucionalmente
inaceptable.



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

Rediseñar sin empezar de cero



Alcaldía de Medellín
— Distrito de —
Ciencia, Tecnología e Innovación



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

Hacia una cuarta generación del control

1. Control manual
2. Control incorporado a sistemas
3. Monitoreo continuo y analítica

4. Control adaptativo de decisiones algorítmicas

No controla solo transacciones.

Observa comportamiento, contexto, impactos y cambio.



COSO sigue vigente, pero debe mirar un objeto nuevo

Diseño y autorización

- Ambiente: principios, apetito y responsabilidades
- Riesgo: clasificación por impacto
- Control: pruebas, límites y aprobación

Evidencia y aprendizaje

- Información: trazabilidad y comunicación
- Monitoreo: desempeño, sesgo, incidentes y deriva
- Auditable desde el diseño



Cinco preguntas para gobernar cualquier modelo

Mandato

- 1. Propósito: ¿para qué lo permitimos?
- 2. Autoridad: ¿qué puede decidir?
- 3. Evidencia: ¿podemos reconstruir el resultado?

Rendición de cuentas

- 4. Supervisión: ¿quién monitorea desempeño e impactos?
- 5. Responsabilidad: ¿quién responde cuando falla?
- Si una pregunta no tiene dueño, el riesgo ya tiene una puerta abierta.



Antes de la IA

IDEA



EJECUCIÓN



Después de la IA

IDEA



EJECUCIÓN



 TikTok
@borjaalvarez.negocios



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

Mi preocupación y mi esperanza

Me preocupa que nuestra capacidad de crear vuelva a adelantarse a nuestra capacidad de comprender.

Pero no soy pesimista: la humanidad ha aprendido a construir instituciones, derechos y controles alrededor de cada gran transformación.



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

GOBERNANZA DE ALGORITMOS Y RIESGOS EMERGENTES

EL NUEVO ASEGURAMIENTO

Norberto Rafael Saraceni

Argentina



No renunciemos a pensar
cuando decidimos con
máquinas.



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación