

Evaluación en tiempos de la IA

Jorge Gallego,

Banco Interamericano de Desarrollo

Junio 6, 2025

Introducción

- **La era de la inteligencia artificial (IA)**

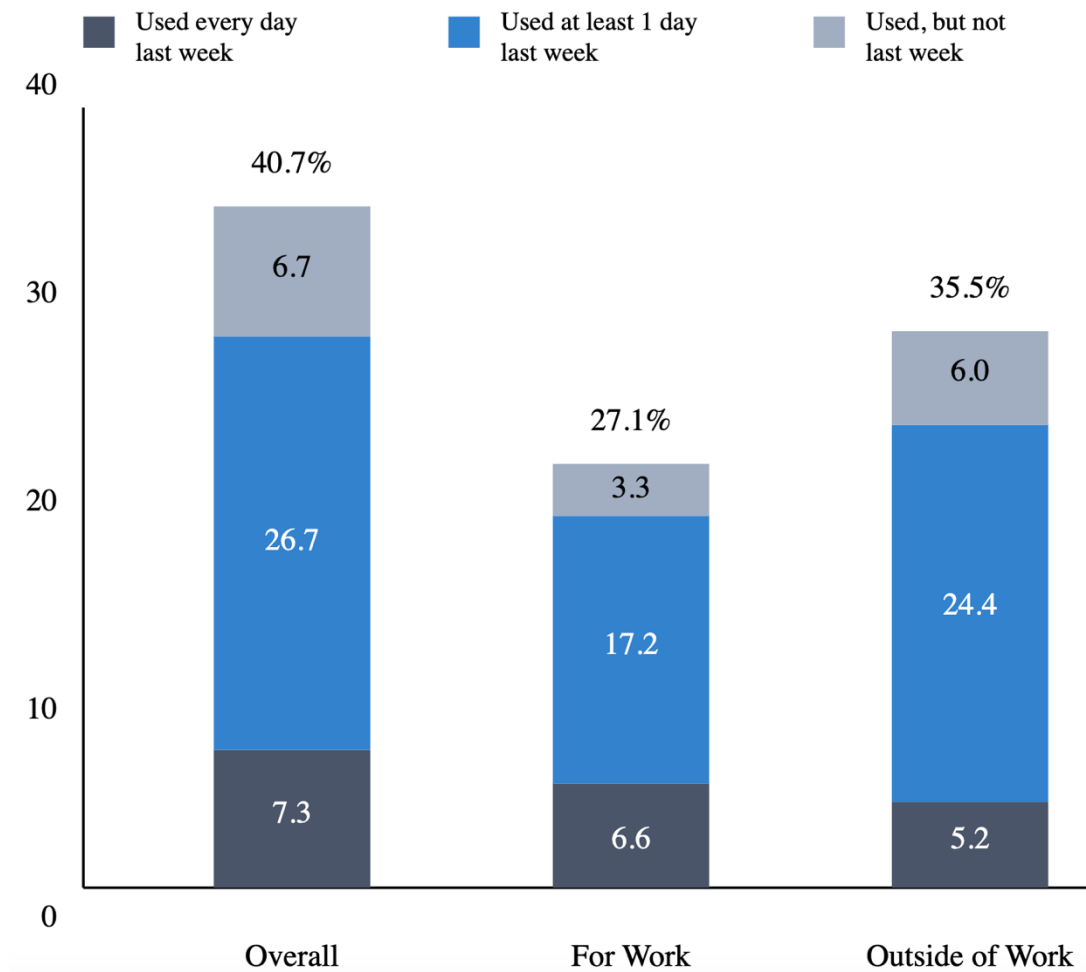
- Dic 2024 ChatGPT reportó 300m de usuarios semanales globales
- Adopción más rápida que para PCs e internet
- Altos impactos sobre productividad (Dell'Acqua et al. 2023)

- **IA y Evaluación**

- IA debe complementar a evaluadores, no reemplazarlos
- IA para procesar muchos datos y automatizar labores repetitivas
- Y dar mejores respuestas a preguntas viejas y nuevas
- Pero hay muchos riesgos y desafíos en el camino

El desafío es ser más eficientes, rigurosos, y tener más alcance a lo largo del ciclo de evaluación

Uso de Gen AI en el trabajo y casa en US
(Diciembre 2024)



Fuente: Bick et al. (2025)



IA Tradicional Y Evaluación

1. Problemas de predicción de política

- Modelos predictivos ayudan a hacedores de política a entender ex ante impactos potenciales de las políticas
- Aplicaciones: salud, ambiente, educación, crimen, etc.

2. Acceso a nuevos y mejores datos

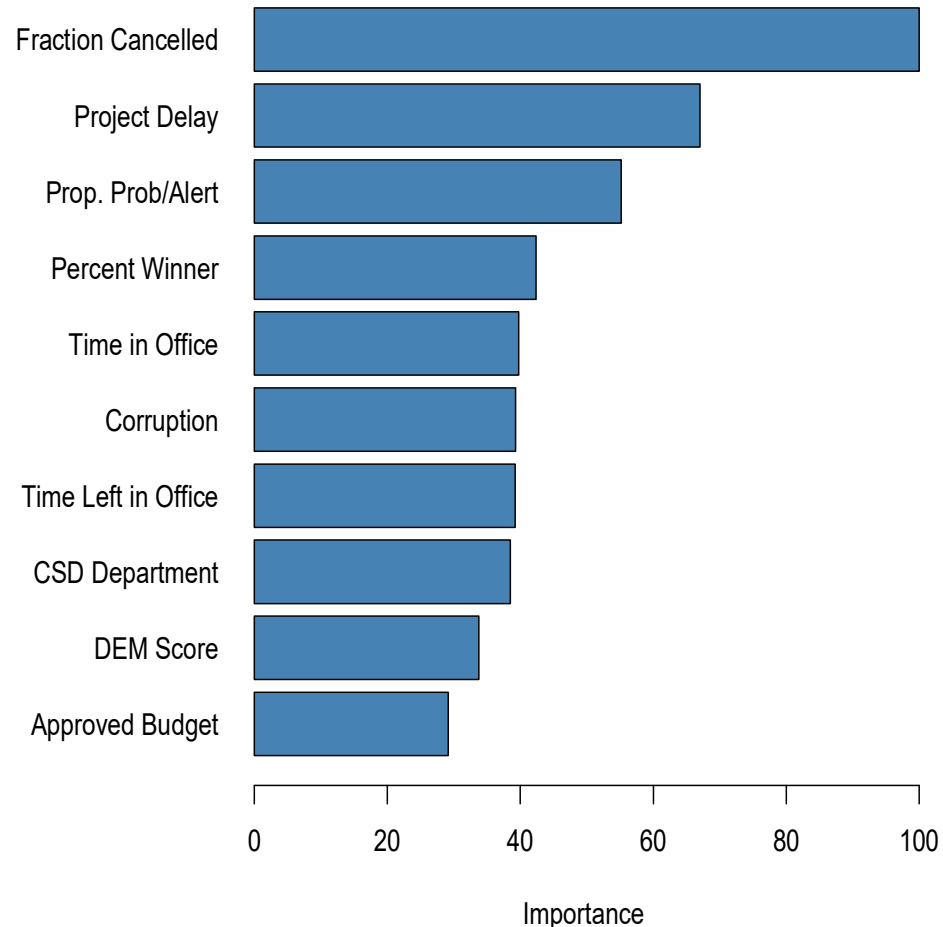
- Imágenes satelitales, datos GPS, posts de redes sociales, teléfonos móviles, tarjetas de crédito, e-gobierno, encuestas automáticas...

3. Inferencia causal fortalecida

- Nuevos y poderosos experimentos (en línea)
- Métodos cuasi-experimentales fortalecidos
- Efectos de tratamiento heterogéneos

¿Qué predice la efectividad en el desarrollo?

Importancia de variables para predecir desempeño de proyectos

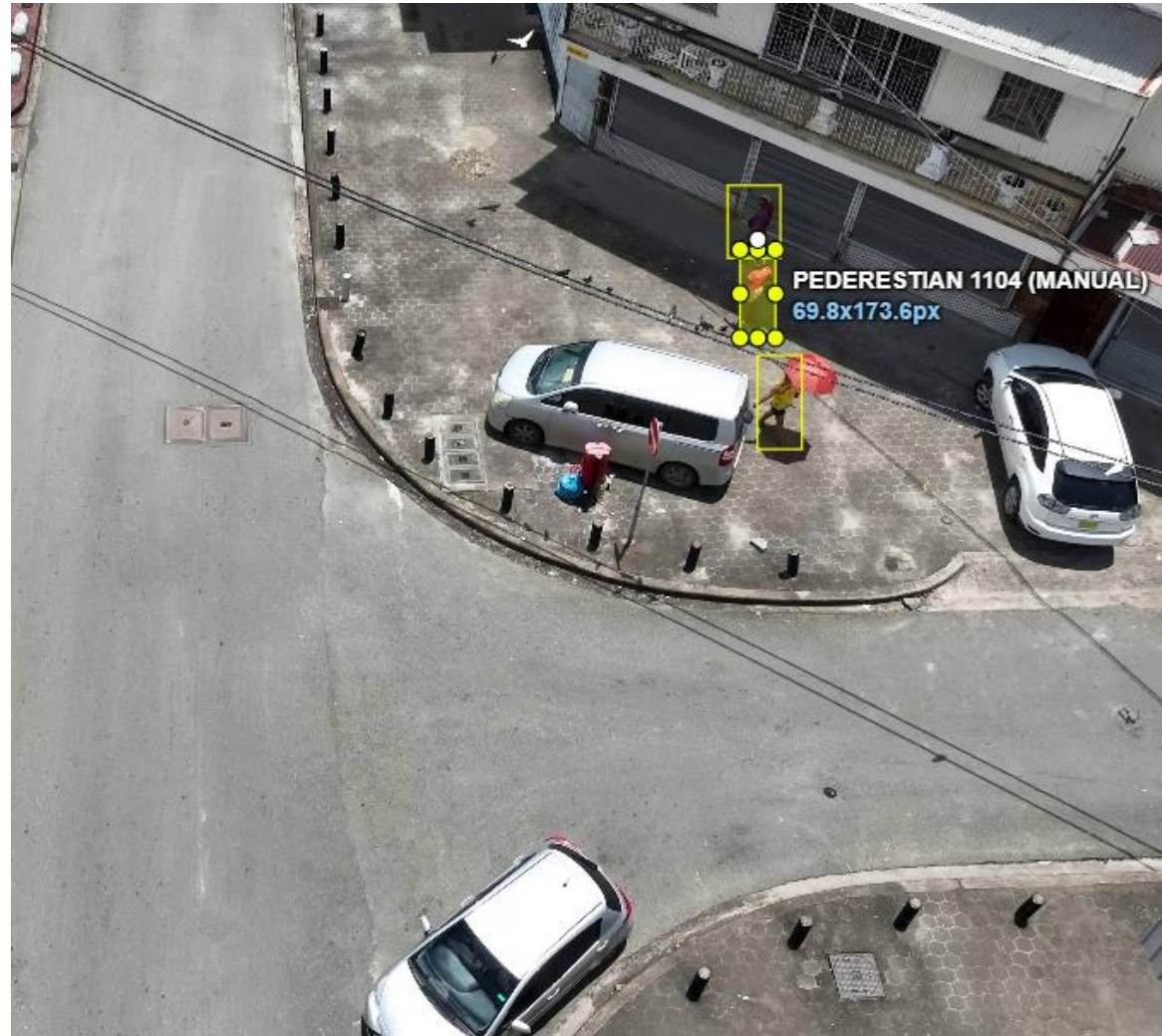


Fuente: OVE (2024)

- Predecir proyectos exitosos/no exitosos al cierre
- Entrenamos modelos de ML (random forest, GBM, redes neuronales, lasso) usando rica base
- 423 proyectos cerrados y 354 características
 - Características de proyectos
 - Condiciones económicas
 - Condiciones políticas
 - Características agencia implementadora
 - Características equipo del Banco
- Niveles aceptables de poder predictivo (7 de 10 proyectos correctamente clasificados)
- Principales predictores de proyectos no exitosos
- Siguiendo paso: Sistema de alertas tempranas

Videos de Drones para Conteo de Peatones

Imagen de dron en corredor vial de Paramaribo





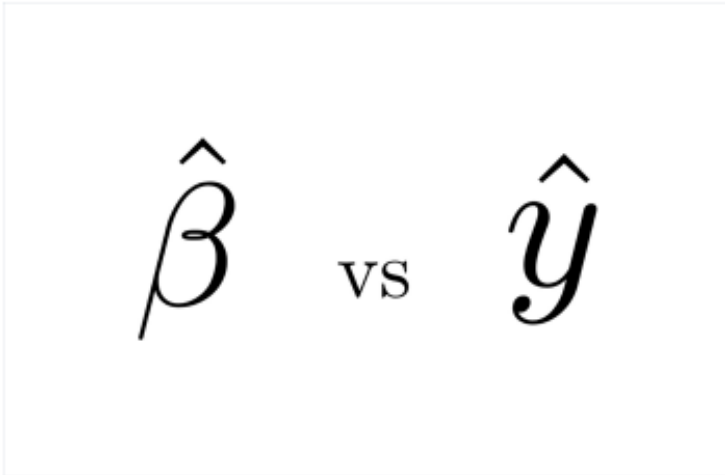
Tweet



Sendhil Mullainathan 
@m_sendhil



Our paper on how machine learning fits in econometric toolbox w/[@jannspiess](#)
J.Econ.Perspectives bit.ly/2pF5OkN
1pic summary:



 AEA Journals

5/3/17, 9:20 AM

Tweet your reply



Home



Explore



Notifications



Messages



Me

Inferencia causal fortalecida

1. Nuevos experimentos

- RCTs en línea (A/B testing, experimentos de encuesta)
- Experimentos con bandidos multi-armados

2. Diseños cuasi-experimentales fortalecidos

- Estimadores ML doblemente robustos
- Variables instrumentales
- Emparejamiento

3. Efectos de tratamiento heterogéneos

- Bosques causales
- Aprendizaje de política

Impactos del DEF en Desempeño de Proyectos



CONFIDENTIAL
INTERNAL USE
PUBLIC UPON CONSIDERATION

Evaluation of the Development Effectiveness Framework

Annex IV Analyzing the Impact of the Development Effectiveness Framework on Project Performance: Cancellations and Project Delays

Office of Evaluation and Oversight (OVE)
Inter-American Development Bank
Washington, D.C.
April 2024

- Evaluamos el impacto del Sistema de M&E del Banco sobre el desempeño de los proyectos
- Usamos modelos de machine learning causal para ello
- Resultados: Cancelaciones financieras y retrasos de implementación
- Modelos lasso sobre 1428 proyectos cerrados
- El DEF reduce las cancelaciones financieras pero aumenta los retrasos



Gen AI: Funciones Útiles

1. Resumen

- Condensación de grandes volúmenes de texto
- Estadísticas descriptivas de datos cuantitativos

2. Transformación

- Traducción, revisión gramatical, corrección de estilo
- Pre-procesamiento de datos, extracción de características

3. Expansión

- Generación de texto
- Aumento de datos

4. Inferencia y análisis

- Clasificación, topic modeling, análisis de sentimiento
- Code interpreter: descriptivo, correlacional, causal, predictivo

Diseño

PNAS

BRIEF REPORT

POLITICAL SCIENCES

OPEN ACCESS

ChatGPT outperforms crowd workers for text-annotation tasks

Fabrizio Gilardi ¹, Meysam Alizadeh, and Maël Kubli

Edited by Mary Waters, Harvard University, Cambridge, MA; received March 27, 2023; accepted June 2, 2023

July 18, 2023 | 120 (30) e2305016120 | <https://doi.org/10.1073/pnas.2305016120>

Identificación de Cartera



Evidencia Evaluativa

SCIENCE ADVANCES | RESEARCH ARTICLE

COMPUTER SCIENCE

Generative AI enhances individual creativity but reduces the collective diversity of novel content

Anil R. Doshi^{1*} and Oliver P. Hauser^{2,3*}

Generación de Hipótesis

Recolección de datos

unconv/gpt4v-browsing

Web Scraping with GPT-4 Vision API and Puppeteer

1 Contributor 0 Issues 522 Stars 244 Forks



Extracción datos no-estructurados

American Journal of Qualitative Research
2024, Vol. 8 No. 2, pp. 153-160
<https://doi.org/10.29333/ajqr/14487>
© 2024 AJQR. <http://www.ajqr.org>

AMERICAN
JOURNAL OF QUALITATIVE
RESEARCH
ISSN: 2576-2141

Using Chat GPT to Clean Qualitative Interview Transcriptions: A Usability and Feasibility Analysis

Zachary Taylor¹
The University of Southern Mississippi, USA

Limpieza de transcripciones

Content

Entrevista Team Leaders - ICPR Panama.docx

Notes AI notes Mentions Transcript

Download

Gallego Duran, Jorge Andres 2:06 PM
¿Listo entonces, Eh?

2:06 PM
Gallego Duran, Jorge Andres started transcription

Gallego Duran, Jorge Andres 2:06 PM
¿Sin más, sin más preámbulo, pues vamos a empezar, vamos a empezar la entrevista, Eh?

La idea de esta entrevista es hablar de de este, de este, de este proyecto.

El 1442101, el del Grupo Ortiz Waterland, Sanitat.

¿Guion hace City?

¿Pues que es 1 de los de los de los proyectos más interesantes que tenemos en la en la cartera de de NG antes, pues de de de de empezar con preguntas específicas, les quería darle Gunnar? Sí.

Generación de Transcripciones

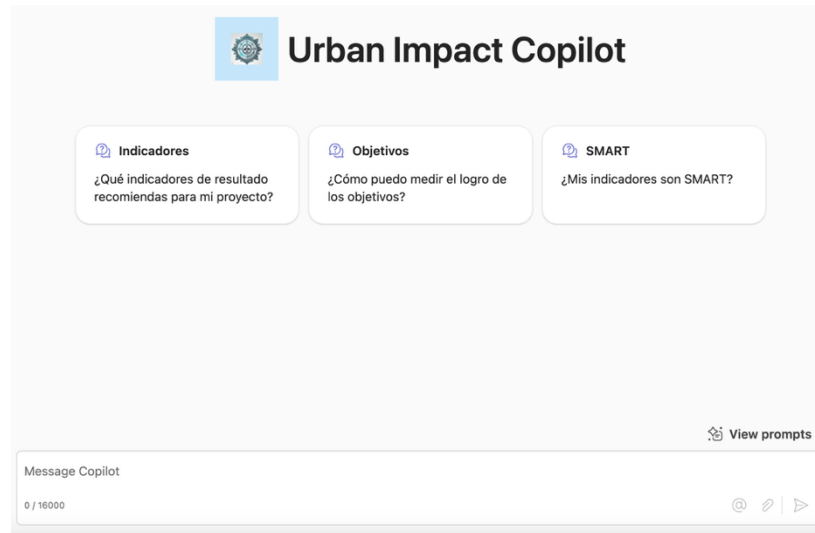
CESifo
WORKING
PAPERS

10666
2023
September 2023

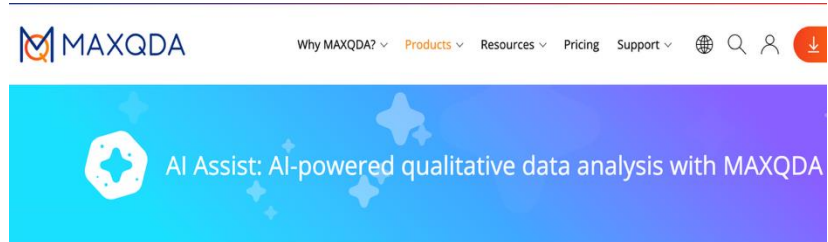
Conducting Qualitative
Interviews with AI
Felix Chopra, Ingar Haaland

Conducción de entrevistas

Inferencia y análisis



Asistentes IA



Meet AI Assist, MAXQDA's virtual research assistant—an add-on to support your analytical efforts. Enhance your research while maintaining the [highest standards of data protection](#) and control over your analytical work.

Discover what is possible with AI Assist

Análisis datos cualitativos



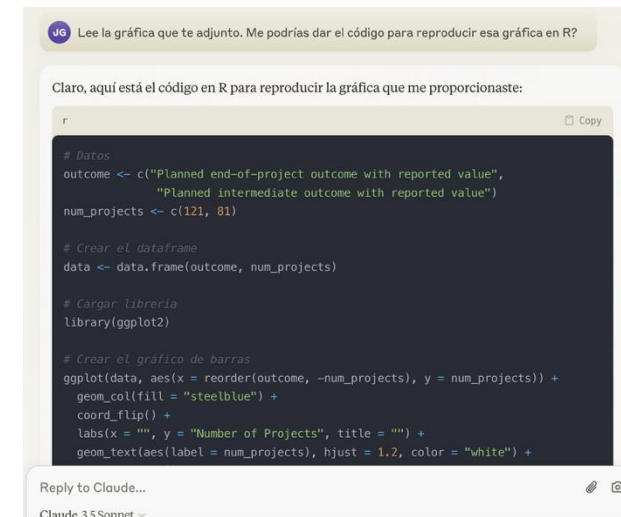
CONFIDENTIAL
INTERNAL USE
PUBLIC UPON CONSIDERATION

Policy-Based Lending at the IDB

Annex III Depth of Policy Actions: A Machine Learning Approach

Office of Evaluation and Oversight (OVE)
Inter-American Development Bank
Washington, D.C.
March 2024

Clasificación de texto



Código

Agente IA: Urban Impact

Copilot Studio My agents > Urban Impact Copilot Update ... X

Describe **Configure** Try it

Template: None

Details

Name *
Urban Impact Copilot
20/30

Description *
AI de Indicadores de Resultado e Impacto para Vivienda y Desarrollo Urbano
74/1000

Detalles del agente

Describe **Configure** Try it

Instructions *

Este agente recomienda al usuario indicadores de resultado para sus proyectos de vivienda y desarrollo urbano. Con base en la descripción de los objetivos de desarrollo del proyecto, sus componentes, intervenciones y cualquier otra información relevante, el GPT propone indicadores SMART (Specific, Measurable, Achievable, Relevant, Time-bound) que sirvan para medir el logro del objetivo. Debe especificar la unidad de medida del indicador, la fuente y la periodicidad del mismo. Los indicadores pueden provenir de fuentes primarias (como encuestas) o secundarias (como registros administrativos gubernamentales). Ideal si se proponen indicadores basados en fuentes alternativas de información y big data, como imágenes satelitales, datos de movilidad GPS, redes sociales, portales web, etc. El documento "Matrices de resultado HUD"

2121/8000

Instrucciones

Urban Impact Copilot

Indicadores ¿Qué indicadores de resultado recomiendas para mi proyecto?

Objetivos ¿Cómo puedo medir el logro de los objetivos?

SMART ¿Mis indicadores son SMART?

View prompts

Message Copilot
0 / 16000

Interface de Chat

Knowledge

SharePoint
Select folders and file types .docx, .doc, .txt, .pptx, .ppt, and .html. or add a SharePoint site. You can add up to 20 knowledge sources. [Learn more](#)

Browse or Enter URL of a SharePoint site

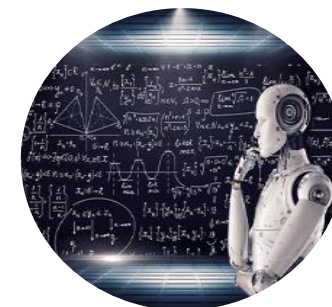
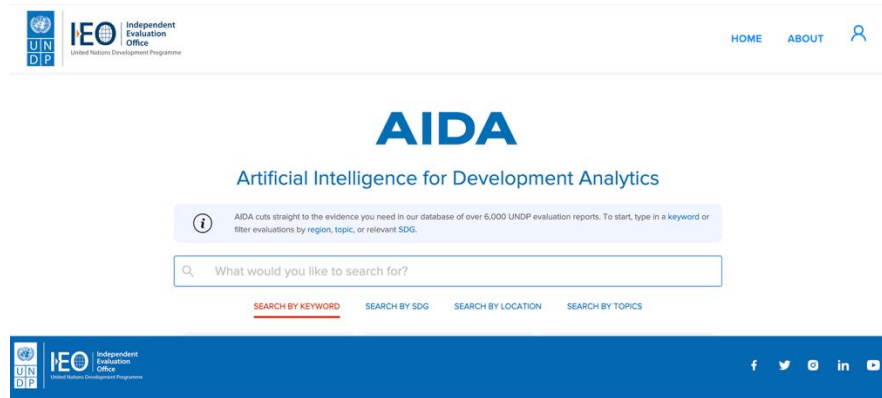
Web content

Referenced from SharePoint

- Matrices de resultado HUD.pdf Preparing ...
- Big-Data-para-el-desarrollo-urbano-sostenible.pdf Preparing ...

Conocimiento

Diseminación



**LESSONS LEARNED
GPT**

Technology **GENERATIVE AI**



Technology

Legal Ethics

Legal Industry

Litigation

Data Privacy



A lawyer used ChatGPT to cite bogus cases. What are the ethics?



By Karen Sloan ▾

May 30, 2023 5:15 PM EDT · Updated a day ago



Riesgos y desafíos

1. Ética

- Sesgos potenciales de los modelos afectando la evaluación
- Asegurar el consentimiento informado en el uso de la IA

2. Seguridad

- Riesgo de fuga de información sensible
- Necesidad de encriptado avanzada y algoritmos seguros

3. Desplazamiento de empleos

- Automatización puede reducir demanda de ciertos empleos
- Requisito de nueva habilidades para evaluadores

4. Interpretabilidad y precisión

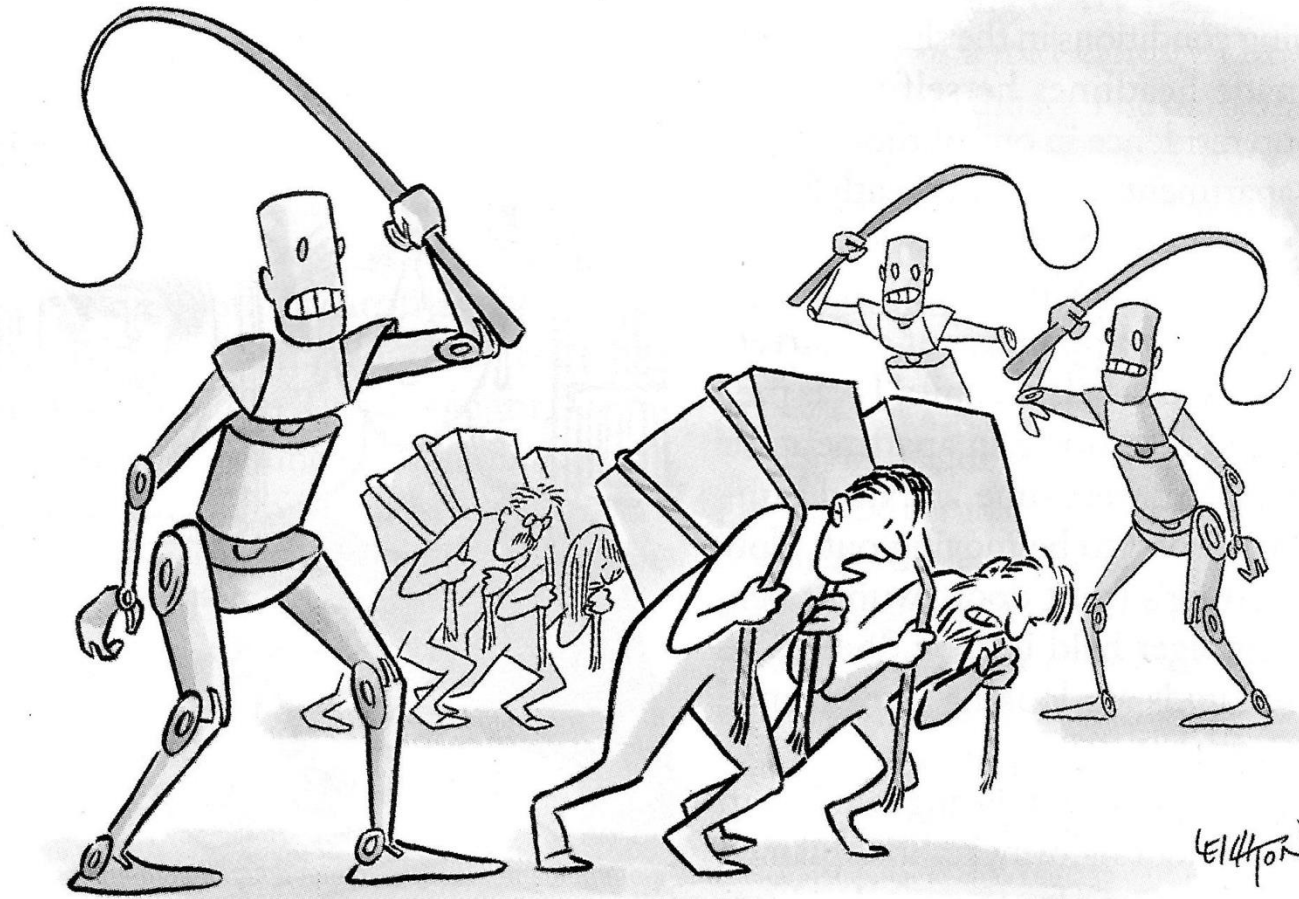
- Modelos IA pueden ser cajas negras, complicando interpretación
- Alucinación, desinformación y manipulación

5. Escalabilidad

- Manejar datos masivos requiere capacidades computacionales

6. Cumplimiento legal y regulatorio

- Cumplimiento con la ley es un gran desafío



“To think this all began with letting autocomplete finish our sentences.”

Gracias!