

Instituto Tecnológico de Costa Rica

Escuela de Ingeniería en Computación



IC6831 Aseguramiento De La Calidad Del Software

Grupo 40

Investigación - Hacer testing en apps de IA

Profesora:

Ericka Solano Fernández

Equipo 1:

Valery Mishel Carvajal Oreamuno – 2022314299

Luis Felipe Calderon Perez – 2021048663

Sergio Chavarría Avilés – 2021077879

Kevin Yadir Calvo Rodriguez – 2023367224

Andrew Denilson Lopez Herrera – 2021062132

Enero, 2025

Índice

Índice.....	1
Introducción.....	2
Estado del arte.....	3
Antecedentes.....	3
Desarrollo del tema.....	5
¿Cómo se garantiza la calidad de la IA?.....	5
La importancia del AC en la IA.....	6
Desafíos para realizar prueba la IA.....	6
Herramientas.....	7
1. Frameworks de Pruebas para IA.....	7
2. Herramientas de Testing de Datos.....	7
3. Evaluación de Modelos.....	7
4. Simuladores y Entornos Virtuales.....	8
5. Pruebas en Producción.....	8
6. Herramientas Éticas y de Gobernanza.....	8
7. Pruebas de Estrés y Robustez.....	8
Casos de éxito/fracaso.....	9
Caso de éxito: Simulaciones extensas en Waymo.....	9
Caso de fracaso: Uso de IA en la toma de órdenes en el drive-through de McDonald's.....	9
Caso de fracaso: El bot de Twitter creado por Microsoft, Tay.....	9
Conclusiones.....	11
Bibliografía.....	12
Anexos.....	15

Introducción

En los últimos años, el ritmo acelerado de la vida moderna ha llevado a las personas a buscar soluciones más rápidas y eficientes para realizar sus tareas diarias. La inteligencia artificial (IA) ha surgido como una herramienta clave en este contexto, permitiendo reducir el tiempo necesario para completar actividades complejas como la edición de videos, la generación de imágenes o texto, y la resolución de problemas. Estas aplicaciones han transformado sectores enteros, facilitando una mayor productividad y creatividad.

Sin embargo, el uso masivo de la IA también plantea desafíos significativos. Según Romero (2024), OpenAI reportó que su IA generativa, ChatGPT, alcanzó 200 millones de usuarios activos en solo una semana, reflejando la creciente preferencia por las consultas a sistemas de IA en lugar de motores de búsqueda tradicionales como Google. Esta popularidad subraya la necesidad crítica de garantizar la precisión y la calidad de las respuestas generadas. Sin procesos adecuados de aseguramiento de calidad, las IA podrían propagar información imprecisa o falsa, generando una gran ola de desinformación con consecuencias potencialmente graves.

Por ello, el presente documento abordará la importancia del testing en aplicaciones de IA, examinando los antecedentes, los desafíos que enfrentan los ingenieros de calidad, las herramientas más utilizadas y casos de éxito y fracaso en el desarrollo de IA. El objetivo es destacar cómo un enfoque robusto de aseguramiento de calidad puede mitigar riesgos y potenciar el impacto positivo de estas tecnologías.

Estado del arte

Primeramente se introducirá el concepto de testing y su relevancia en la ingeniería de software, destacando su papel en la identificación de defectos y la estabilización del software (Hoyos, s.f.). Asimismo, se describen las aplicaciones de la inteligencia artificial (IA), que emplean tecnologías como el aprendizaje automático (ML) y el procesamiento del lenguaje natural (NLP) para realizar tareas que típicamente requieren intervención humana (Sevilla et al., 2024).

Antecedentes

La metodología adoptada se basó en una revisión de literatura científica en línea, enfocándose en publicaciones de los últimos cinco años, seleccionadas por su calidad y relevancia.

El avance acelerado de la IA ha incrementado la importancia de los procesos de prueba y control de calidad en sus diversas aplicaciones. La ética en la IA es un área crítica, donde las pruebas deben asegurar el cumplimiento de derechos humanos, privacidad de datos y transparencia (Ayling & Chapman, 2022).

Se destaca la implementación de frameworks respaldados por regulaciones como el GDPR, involucrando a stakeholders que aportan perspectivas esenciales para identificar y abordar sesgos (Ayling & Chapman, 2022). La importancia de una aproximación multidisciplinaria es subrayada para desarrollar una IA ética.

En las aplicaciones de salud la detección de enfermedades infecciosas, la IA ha optimizado significativamente el análisis de datos complejos. Se describen las etapas esenciales para el desarrollo de estas aplicaciones, que incluyen la evaluación de datos, validación de métodos, generalización y pruebas en condiciones reales (Ayling & Chapman, 2022).

Se han identificado retos en las pruebas de IA del desarrollo de pruebas para IA, incluyendo la identificación de datos de prueba adecuados, incertidumbre algorítmica, falta de medidas efectivas de efectividad, y dificultades en la especificación comprobable (Sugali

et al., 2021). También se aborda la incomprensibilidad inherente a ciertos modelos de IA, complicando la evaluación de la cobertura del código (Sugali et al., 2021).

En las innovaciones de aseguramiento de calidad se han presentado nuevos modelos de aseguramiento de calidad, como un modelo de tres etapas para el aprendizaje automático, que enfatiza la planificación y monitoreo de resultados (Fujii et al., 2020). Se destacan metodologías específicas para sistemas como interfaces de usuario por voz y conducción autónoma, enfocándose en la robustez y retroalimentación continua para mejorar el desempeño (Fujii et al., 2020).

Desarrollo del tema

Así como cualquier aplicación de software, en la inteligencia artificial también se necesita asegurar la calidad de este y mostrar que funciona de manera correcta. Los ingenieros están acostumbrados a realizar pruebas a software que funciona de manera determinista, pero cuando se creó la IA generativa, se tuvo que buscar otras maneras de realizar pruebas.

¿Cómo se garantiza la calidad de la IA?

En primer lugar, se debe decir que llegar al 100% de precisión en las salidas o respuestas de la inteligencia artificial es imposible, por lo que antes de iniciar las pruebas, se debe decidir el porcentaje de precisión realista al que se desea llegar. En este porcentaje se debe contemplar las expectativas y limitaciones del modelo de IA.

Ahora, uno de los problemas que se presentan al realizar pruebas de calidad a la IA es que la inteligencia está constantemente aprendiendo, por lo que, los ingenieros de AC (Aseguramiento de la Calidad) deben tener una gestión continua de las pruebas que realizan en los modelos para asegurarse que la IA haya aprendido de manera correcta.

La manera en la que los ingenieros prueban la IA es la siguiente: se deben de diseñar pruebas, se debe seleccionar una parte de los datos de entrenamiento, se deben realizar esas pruebas y se mide el desempeño del modelo, revisar si las tareas se completaron de manera correcta según los estándares indicados y se observa el comportamiento de los datos y si el resultado es el esperado. Si se encuentran errores importantes durante el proceso de validación, se devuelve a la etapa de desarrollo y se vuelve a iniciar otra prueba de calidad, pero con otro conjunto de datos. Además, los ingenieros deben asegurarse que los datos proporcionados por la IA (en el caso de la IA generativa) sean diversos, precisos y representen los escenarios reales.

Por la propia naturaleza de la inteligencia artificial, esta debe ser probada en producción, es decir, el entorno final donde la aplicación está plenamente operativo y accesible para los usuarios finales. Esto se debe al crecimiento en los datos que utiliza la IA para aprender de manera constante. Por esa razón, los ingenieros deben estar constantemente realizando pruebas incluso en el periodo de producción.

Actualmente, el proceso de AC para la IA tiene cinco enfoques: integridad de datos, robustez del modelo, calidad del sistema, agilidad del proceso y expectativas del cliente.

La importancia del AC en la IA

La IA es ampliamente utilizada actualmente, de hecho se ha convertido el principal apoyo de muchas personas, pero ¿Cómo se hace para que la información dada sea realmente útil? En esta ocasión no se hablará sobre los procesos de AC para que la información generada por la IA sea útil, sino, se hablará de la importancia que tiene el AC para la IA.

La principal razón por la que una IA debe llevar un proceso de AC es por la información generada, como se dijo anteriormente, la información debe ser precisa, pero también debe ser útil para la persona que la consulta. De nada sirve información precisa si el modelo de IA no es capaz de proporcionar lo que se le solicita.

Lo más probable es que el equipo de AC se encuentre con muchas situaciones en el que la IA no responda adecuadamente a las consultas que se le realizan. Para esas situaciones el equipo de AC debe intervenir para modificar lo que sea necesario para que la IA pueda contestar correctamente. Este proceso puede tener una duración no estimable para el equipo. Todo dependerá de la rigurosidad de las pruebas que hagan los ingenieros.

Desafíos para realizar pruebas en la IA

Como en toda actividad profesional, siempre existen los desafíos y el medir la calidad de una IA no es la excepción. Los desafíos más comunes son las siguientes:

- **Conjunto de Datos de Entrenamiento:** Los modelos de IA requieren muchos datos para ser entrenados, estos datos pueden ser difíciles de conseguir o pueden conseguirse, pero pueden ser datos con baja calidad, que podrían provocar que la IA dé respuestas erróneas.
- **Complejidad Algorítmica:** Lo más difícil de la IA es comprender cómo funciona, si el algoritmo que utiliza la IA es bastante compleja, puede provocar errores humanos por no comprender de manera correcta la IA.
- **Sistemas Supervisados y No Supervisados:** Los sistemas supervisados se realiza el entrenamiento del modelo con datos etiquetados, es decir, la IA sabe qué datos es, porque ya hay una etiqueta que lo indique. El sistema no supervisado utiliza datos no etiquetados para su entrenamiento. Depende de cada equipo de AC el sistema a utilizar para que el modelo se entrene de manera efectiva.
- **Integración de Componentes de Terceros:** Muchos sistemas de IA utilizan componentes de terceros, por ejemplo modelos pre-entrenados, el equipo de AC debe asegurar la compatibilidad entre los componentes de terceros y el modelo que están desarrollando.

- **Toma de Decisiones Transparencia:** Los equipos de AC deben asegurarse que las decisiones de la IA, así como las respuesta de esta misma, son comprendidas de manera clara por los usuarios finales. Además, esto se debe mostrar a través de un diseño de UI/UX reflexivo, que permita apreciar de manera clara la toma de decisiones de la IA.
- **Velocidad de Aprendizaje Cambiable:** La velocidad con la que aprende una IA es impredecible, esta depende de muchos factores como la cantidad de datos, la calidad de los datos o la complejidad de los datos. La velocidad con la que aprende el modelo puede afectar directamente el rendimiento de este mismo.
- **Modelos subajuste y sobreajuste:** La subadaptación pasa cuando un modelo es demasiado simple y no puede entender los patrones de los datos. Esto hace que no funcione bien y que no pueda manejar situaciones más complejas del mundo real. Por otro lado, el sobreajuste ocurre cuando un modelo funciona muy bien con los datos de entrenamiento, pero le va mal con datos nuevos o de validación. Aunque parece preciso en el entrenamiento, no logra adaptarse a otros casos o situaciones.
- **Riesgos de usar Modelos Pre-Entrenados:** Introducir modelos pre-entrenados al modelo que se está desarrollando puede provocar problemas heredados. Además, el equipo debe asegurar la compatibilidad entre ambos modelos para evitar futuros problemas en las respuestas de la IA.
- **Concepto Drift:** El drift conceptual sucede cuando las características de los datos o de la variable que se quiere predecir cambian con el tiempo. Estos cambios pueden ser lentos o repentinos, y afectan cómo funcionan los modelos de IA que fueron entrenados con un conjunto de datos diferente. Por eso, es importante monitorear constantemente los datos y ajustar los modelos para que sigan siendo efectivos.

Herramientas

La implementación y validación de modelos de inteligencia artificial requieren herramientas especializadas que permitan abordar los desafíos únicos de esta tecnología, como la evaluación de datos, el rendimiento del modelo y la ética. A continuación, se presentan algunas de las herramientas más utilizadas en este ámbito:

1. Frameworks de Pruebas para IA

- a. TensorFlow Extended (TFX): Toolkit para el desarrollo de pipelines de aprendizaje automático. Incluye componentes como TFMA que permite analizar el rendimiento del modelo en diferentes subconjuntos de datos, asegurando que funcione bien para todos los usuarios (Google, 2023).
- b. MLflow: Rastrea y organiza los modelos de aprendizaje automático, permite comparar el desempeño y garantiza la reproducibilidad en diferentes entornos (Zaharia et al., 2018).

2. Herramientas de Testing de Datos

- a. Great Expectations: Framework para validación y documentación de datos, asegura que los conjuntos de datos cumplan con estándares de calidad (Great Expectations, 2023).
- b. Pandera: Biblioteca de Python para validar la estructura de los datos y detectar anomalías antes de usarlos en el entrenamiento del modelo (Nielsen, 2022).

3. Evaluación de Modelos

- a. Apache Benchmarks for Machine Learning (MLBench): Permite realizar pruebas de benchmarking (proceso para comparar el rendimiento de sistemas o modelos frente a estándares o competidores para mejorar su eficiencia) en modelos de IA bajo diferentes datasets (Apache Software Foundation, 2022).
- b. Explainability Tools (SHAP y LIME): Ayudan a entender cómo el modelo toma decisiones y detectan posibles comportamientos inesperados. Estas herramientas son esenciales para garantizar la transparencia de los modelos (Lundberg & Lee, 2017; Ribeiro et al., 2016).

4. Simuladores y Entornos Virtuales

- a. CARLA Simulator: Entrenar y prueba modelos de conducción autónoma en entornos controlados (Dosovitskiy et al., 2017).
- b. Unity ML-Agents: Permite crear entornos simulados personalizados para entrenar modelos de IA en aplicaciones como videojuegos o robots (Unity Technologies, 2023).

5. Pruebas en Producción

- a. Seldon Core: Implementa, monitorea y realiza pruebas en modelos de IA en entornos de producción, capacidades de detección de deriva de datos y monitorización de métricas (Seldon, 2023).
- b. Prometheus y Grafana: Permiten rastrear el comportamiento del modelo en producción y detectar anomalías en tiempo real (Prometheus Authors, 2023; Grafana Labs, 2023).

6. Herramientas Éticas y de Gobernanza

- a. IBM AI Fairness 360 (AIF360): Toolkit que evalúa y mitiga sesgos en los modelos de IA, tiene herramientas para analizar la equidad y generar reportes detallados (Bellamy et al., 2018).
- b. Microsoft Responsible AI Toolbox: Toolkit que ayuda a desarrollar modelos de IA responsables, centrándose en la transparencia, interpretabilidad y ética (Microsoft, 2023).

7. Pruebas de Estrés y Robustez

- a. Adversarial Robustness Toolbox (ART): Permite probar la robustez de los modelos de IA frente a ataques adversariales, asegurando su desempeño en condiciones adversas (Nicolae et al., 2018).
- b. DeepCT (Deep Check Testing): Evalúa la capacidad de un modelo para manejar casos extremos y datos anómalos (DeepCT, 2023).

Casos de éxito/fracaso

Caso de éxito: Simulaciones extensas en Waymo

Waymo es una empresa que desarrolla vehículos autónomos. Para asegurar la mejor experiencia, Waymo realiza pruebas en miles de millones de entornos simulados, reuniendo cantidades de datos (Waymo, s.f.). Estos casos simulados permiten exponer sus vehículos a casos extremos poco frecuentes, además de probar y entrenar su IA en escenarios que serían demasiado riesgosos para recrearlos en carreteras reales (Berkovich, 2024). Estas simulaciones han hecho que Waymo sea uno de los líderes en el ámbito de la conducción autónoma, con muy pocas quejas entre sus usuarios.

Caso de fracaso: Uso de IA en la toma de órdenes en el drive-through de McDonald's

En el 2021, McDonald's colaboró con IBM para implementar inteligencia artificial en la toma de órdenes del drive-through de unos 100 establecimientos en los Estados Unidos (Sky News, 2024). Este sistema utilizaba reconocimiento de voz para procesar las órdenes de los clientes. Sin embargo, la confiabilidad de la tecnología fue puesta en duda cuando el público compartió videos de errores en los pedidos en las redes sociales. Se determinó que no se hicieron suficientes pruebas en entornos reales, por lo que la inteligencia artificial no reconocía correctamente las palabras de las personas cuando tenían distintos acentos, o cuando había algo de ruido de fondo (Williams, 2024). En junio del 2024, McDonald's anunció que dejará de operar estos servicios en todos sus restaurantes (Williams, 2024). Este caso resalta la necesidad de no apresurarse a desplegar este tipo de tecnologías, ya que podría afectar negativamente la percepción del público. Es más importante sacar algo bien hecho que sacar algo rápido.

Caso de fracaso: El bot de Twitter creado por Microsoft, Tay

En el 2016, Microsoft anunció a Tay, un bot de Twitter que podía publicar sus “pensamientos”. Tay fue diseñada para aprender más sobre el lenguaje a lo largo del tiempo a partir de su entorno, lo que le permite tener conversaciones sobre cualquier tema (Ganesan, s.f.). Sin embargo, este aspecto no fue probado lo suficiente. Algunos usuarios de Twitter comenzaron a tuitearle frases políticamente incorrectas, enseñándole mensajes cosas malas al bot. Este bot terminó haciendo varios comentarios racistas y sexistas, por lo que terminó siendo apagada tan solo 16 horas después de haber sido desplegada (Ganesan, s.f.). Esto resalta la falta de pruebas al bot. Si Tay hubiera estado sujeta a diferentes tipos de

interacciones en sus tweets durante la etapa de pruebas, los peligros de desplegar a Tay se habrían vuelto obvios (Ganesan, s.f.). Este caso resalta la necesidad de probar con todos los datos que sean posibles, con tal de encontrar algún caso que previamente no se haya considerado.

Conclusiones

Las pruebas en inteligencia artificial son esenciales para garantizar la funcionalidad, ética y adaptabilidad de los sistemas. Dado que los modelos de IA son dinámicos y aprenden continuamente, las pruebas deben realizarse tanto en etapas iniciales como durante la producción. Las herramientas como TensorFlow Extended y MLflow facilitan la identificación de sesgos y errores, mientras que los casos estudiados destacan la importancia de una validación rigurosa antes de desplegar sistemas en el mundo real.

Es fundamental que las pruebas no solo evalúen la precisión del modelo, sino también la robustez frente a escenarios adversos y la transparencia en la toma de decisiones. Las herramientas de interpretabilidad, como SHAP y LIME, juegan un rol crucial para garantizar que los modelos sean comprensibles tanto para ingenieros como para usuarios finales. Además, los casos de éxito y fracaso demuestran que el aseguramiento de calidad efectivo puede prevenir errores costosos y mejorar la confianza del usuario. Una supervisión continua en entornos de producción es vital, especialmente para aplicaciones de IA que aprenden constantemente de nuevos datos.

En el futuro, se requiere avanzar hacia metodologías que no sean únicamente más efectivas y adaptativas, sino también que promuevan una IA transparente, segura y éticamente responsable. Esto implica fomentar la colaboración interdisciplinaria, integrar principios de ética desde el diseño y establecer marcos regulatorios globales que aseguren el cumplimiento de buenas prácticas en el testing de IA.

Bibliografía

- Apache Software Foundation. (2022). Apache Benchmarks for Machine Learning (MLBench). Recuperado de <https://mlbench.apache.org/>
- Ayling, J., & Chapman, A. (2022). Putting AI ethics to work: Are the tools fit for purpose? AI and Ethics, 2(3), 405–429. <https://doi.org/10.1007/s43681-021-00084-x>
- Bellamy, R. K. E., Dey, K., & Hind, M. (2018). AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted bias in machine learning models. Recuperado de <https://aif360.mybluemix.net/>
- Berkovich, A. (2024). How Edge Case Detection Contributes to Safer AI in Autonomous Vehicles. Akridata. Recuperado de: <https://akridata.ai/blog/edge-case-detection-safer-ai-autonomous-vehicles/>
- Besedina, E., & Project Manager. (s/f). The role of QA & testing in AI projects. Integrio Systems. <https://integrio.net/blog/the-role-of-qa-and-testing-in-ai-projects>
- Chamorro, P. (2021, enero 6). How quality assurance (QA service) works with AI. BairesDev. <https://www.bairesdev.com/blog/how-quality-assurance-works-with-ai/>
- Claessens, M., Oria, C. S., Brouwer, C. L., Ziemer, B. P., Scholey, J. E., Lin, H., Witztum, A., Morin, O., Naqa, I. E., Van Elmpt, W., & Verellen, D. (2022). Quality assurance for AI-based applications in radiation therapy. Seminars in Radiation Oncology, 32(4), 421–431. <https://doi.org/10.1016/j.semradonc.2022.06.011>
- Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017). CARLA: An open urban driving simulator. Proceedings of the 1st Annual Conference on Robot Learning. Recuperado de <https://carla.org/>
- El Español. (2024, 30 de agosto). *El éxito brutal de ChatGPT: alcanza 180 millones de usuarios semanales, el doble que hace un año*. Omicrono. Recuperado de https://www.elespanol.com/omicrono/software/20240830/exito-brutal-chatgpt-alcanza-millones-usuarios-semanales-doble-hace-ano/882161787_0.html
- Fujii, G., Hamada, K., Ishikawa, F., Masuda, S., Matsuya, M., Myojin, T., Nishi, Y., Ogawa, H., Toku, T., Tokumoto, S., Tsuchiya, K., & Ujita, Y. (2020). Guidelines for quality assurance of machine learning-based artificial intelligence. International Journal of Software Engineering and Knowledge Engineering, 30(11-12), 1589–1606. <https://doi.org/10.1142/S0218194020400227>
- Ganesan, K. (s.f.). What went wrong with Tay, the Twitter bot that turned racist? Opinosis Analytics. Recuperado de: <https://www.opinosis-analytics.com/blog/tay-twitter-bot/>

Google. (2023). TensorFlow Extended (TFX). Recuperado de <https://www.tensorflow.org/tfx>

Grafana Labs. (2023). Grafana. Recuperado de <https://grafana.com/>

Great Expectations. (2023). Data quality made easier. Recuperado de <https://greatexpectations.io/>

Hoyos, P. (s.f.). Qué es el testing de software y por qué es tan importante en el desarrollo de software. PacifiTIC. <https://pacifitic.org/que-es-el-testing-de-software-y-por-que-es-tan-importante-en-el-desarrollo-de-software/>

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.

Microsoft. (2023). Responsible AI Toolbox. Recuperado de <https://www.microsoft.com/en-us/ai/responsible-ai>

Nicolae, M.-I., Sinn, M., Tran, M. N., Buesser, B., & Rawat, A. (2018). Adversarial Robustness Toolbox v1.0. arXiv preprint arXiv:1807.01069.

Nielsen, J. (2022). Pandera: Statistical data validation for humans. Recuperado de <https://pandera.readthedocs.io/>

Prometheus Authors. (2023). Prometheus monitoring. Recuperado de <https://prometheus.io/>

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

Seldon. (2023). Seldon Core: Simplifying Machine Learning Deployment. Recuperado de <https://www.seldon.io/>

Sevilla, A. K. R. N. (2024, diciembre 6). ¿Cuáles son las mejores aplicaciones de IA en 2024? Guru. <https://www.getguru.com/es/reference/best-ai-apps#:~:text=%C2%BFQu%C3%A9%20es%20una%20aplicaci%C3%B3n%20de,IA%20en%20la%20siguiente%20secci%C3%B3n>

Sky News (2024). McDonald's ends AI drive-thru trial after order mishaps. Recuperado de: <https://news.sky.com/story/mcdonalds-ends-ai-drive-thru-trial-after-order-mishaps-13155091>

Sugali, K., Sprunger, C., & Inukollu, V. N. (2021). Software testing: Issues and challenges of Artificial Intelligence & Machine Learning. *International Journal of Artificial Intelligence & Applications*. <https://doi.org/10.5121/IJAIA.2021.12107>

Unity Technologies. (2023). Unity ML-Agents. Recuperado de <https://unity.com/ml-agents>

Waymo (s.f.). Waymo Driver. Recuperado de: <https://waymo.com/intl/es/waymo-driver/>

Williams, L. (2024). Explainer: Is the McDonald's AI rollout failure a lesson to companies jumping on the AI bandwagon? Verdict. Recuperado de:

<https://www.verdict.co.uk/explainer-is-macdonalds-ai-roll-out-failure-a-lesson-to-companies-jumping-on-the-ai-bandwagon/?cf-view&cf-closed>

Zaharia, M., Chen, A., Davidson, A., Ghodsi, A., Hong, M., Konwinski, A., ... & Xin, R. (2018). Accelerating the machine learning lifecycle with MLflow. Proceedings of the 4th ACM SIGMOD Workshop on Data Management for End-to-End Machine Learning.

Anexos

Anexo 1: Link a la presentación

https://www.canva.com/design/DAGcJgaBNmo/nyF7TvExMo_xr1R6l2Si_A/edit?utm_content=DAGcJgaBNmo&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

Anexo 2: Preguntas para el foro

1. ¿Cuáles podrían ser las consecuencias para los usuarios finales si no se realiza un testing adecuado en aplicaciones de IA?
2. ¿Qué tipos de pruebas son más efectivas para garantizar una evaluación rigurosa de las aplicaciones de IA?
3. ¿Es ético lanzar una aplicación de IA sin un testing exhaustivo que garantice su precisión y veracidad? ¿Por qué?