Curso de fundamentos algorítmicos y ética en IA: de la teoría a la práctica

Manual

Número de Proyecto :2022-1-ES01-KA220-HED-000085257

Índice de contenidos

CU1 | Ética de la IA: un enfoque práctico 2

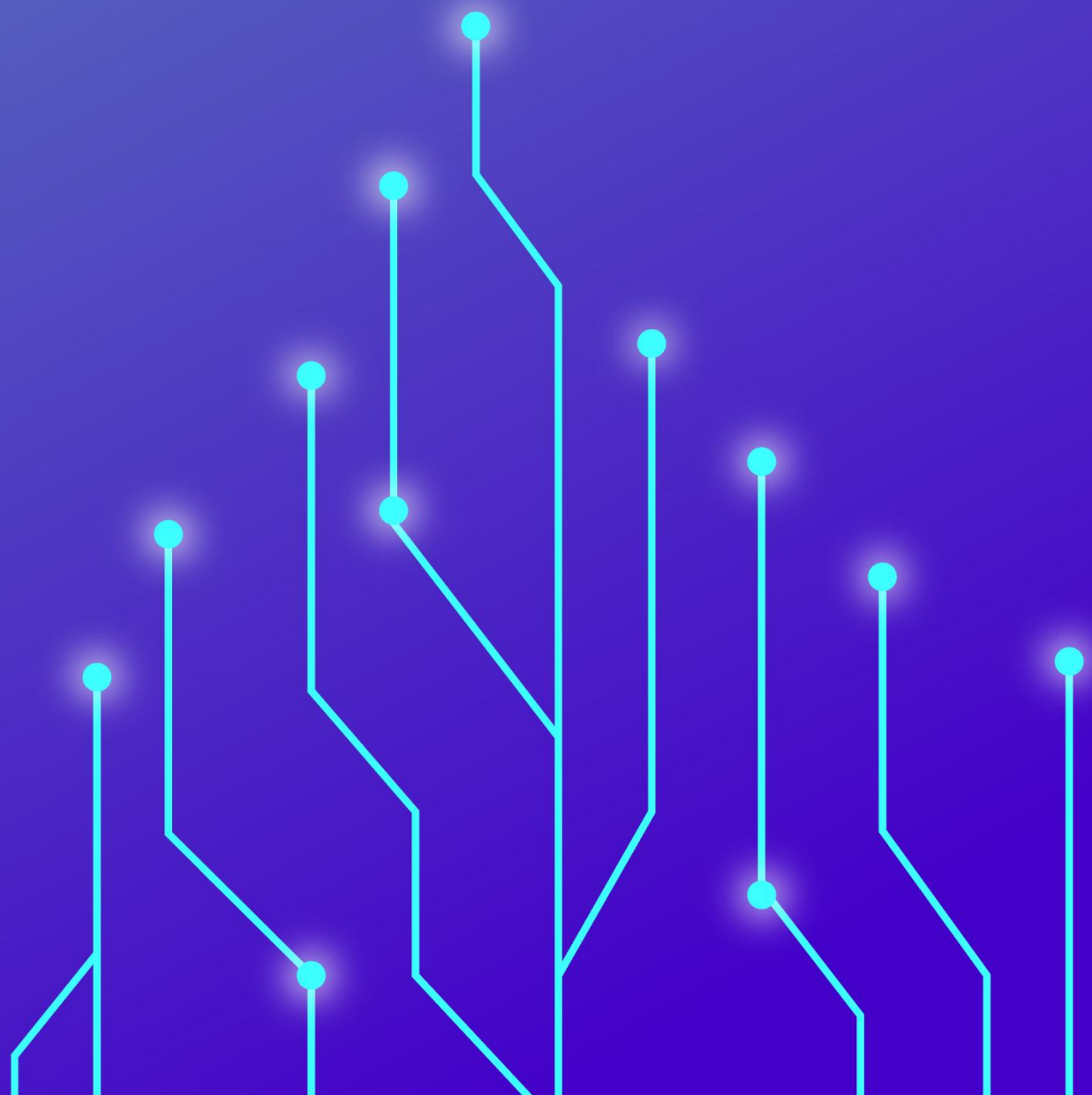
CU2 | Privacidad y conveniencia de la IA..... 39

CU3 | Los algoritmos y sus limitaciones 67

CU4 | Imparcialidad y sesgo de los datos..... 113

CU5 | Casos de estudio y proyectos 132

CU1 | Ética de la IA: un enfoque práctico



Índice de contenidos

1. Introducción 4

2. Definición de principios, marcos y directrices éticas 5

3. Esfuerzo común para dar forma al panorama ético 6

4. Principios éticos 7

5. Marcos éticos, directrices y conjuntos de herramientas 12

6. De los principios a la práctica; Marco de IA confiable de HLEG 14

7. De los principios al derecho: LEY DE IA DE LA UE 16

8. Ciclo de vida del desarrollo de la IA 18

9. Participación de las partes interesadas en el ciclo de vida del desarrollo de la IA 19

10. Ciclo de vida del desarrollo de la IA: definición del problema y comprensión del negocio 23

11. Ciclo de vida del desarrollo de la IA: etapa de diseño 25

12. Ciclo de vida del desarrollo de la IA: recopilación, comprensión y preparación de datos 27

13. Ciclo de vida del desarrollo de la IA; Desarrollo y formación de modelos 28

14. Ciclo de vida del desarrollo de la IA: evaluación y pruebas del modelo 30

15. Ciclo de vida del desarrollo de la IA: implementación 32

16. Ciclo de vida del desarrollo de la IA: funcionamiento y supervisión 34

17. Conclusión 35

18. Referencias 35

1. Introducción

CONSEJOS Y RECOMENDACIONES PARA LOS

Este manual tiene como objetivo guiarlo a través del material de la unidad del curso. Los PPT de soporte contienen información similar a la que cubrió el E-learning, pero en forma más detallada. Los PPT son extensos y no están destinados a ser cubiertos en su totalidad durante las sesiones interactivas. En su lugar, puede preguntar al comienzo de las sesiones interactivas qué temas fueron más difíciles de comprender para los estudiantes y usar solo esas diapositivas de apoyo durante la sesión. Puedes preguntar esto usando herramientas como *Mentimeter o Kahoot* como ejemplo, agregando los temas del curso del contenido de la unidad de competencia descritos en viñetas al final de la sección de introducción. Además, se pretende que el estudio de caso basado en CU también se maneje durante la sesión interactiva. Cualquier sugerencia adicional en el manual es solo una guía; Siéntase libre de usarlo como mejor le parezca.

¿Qué pasaría si la inteligencia artificial (IA) tuviera el poder de decidir si obtienes un préstamo? ¿O qué pasaría si la IA juzgara quién será condenado a prisión? ¿Cuáles son las consecuencias si un coche autónomo tuviera que decidir si salvar a un peatón o salvar la vida de los que están dentro del coche? ¿Qué pasaría si un sistema de contratación basado en IA decidiera si contratar a una candidata femenina o masculina como capitán de vuelo?

La creciente integración de la IA en diversos aspectos de la sociedad plantea importantes preocupaciones éticas. Desde las decisiones sobre préstamos, sentencias penales y vehículos autónomos, hasta las prácticas de contratación, la dependencia de los algoritmos de IA plantea preguntas sobre principios éticos como la equidad, la transparencia y la responsabilidad. Las tecnologías de IA tienen el potencial de perpetuar sesgos, discriminación y desigualdades si no se desarrollan y regulan cuidadosamente. Garantizar que los sistemas de IA prioricen las consideraciones éticas es esencial para promover la justicia, la igualdad y la confianza en sus aplicaciones en diferentes ámbitos. (UNESCO, 2023c)

Esta unidad profundizará en los principios, marcos y directrices éticas que afectan al desarrollo de las soluciones de IA. El ciclo de vida del desarrollo de la IA se examina desde el planteamiento del problema hasta el funcionamiento y la supervisión de la solución de IA. En cada paso del proceso se tienen en cuenta los principios relacionados y la participación de las partes interesadas. En el nivel alto, esta unidad del curso maneja los siguientes temas.

- La ética de la IA y su importancia en el desarrollo de soluciones de IA

- Principios éticos de la IA
- Marcos éticos, directrices y conjuntos de herramientas
- Grupo de expertos de alto nivel; IA confiable – Marco
- Ley de IA de la UE
- Ciclo de vida del desarrollo de la IA
- Participación de las partes interesadas en el ciclo de vida del desarrollo de la IA
- Ciclo de vida del desarrollo de IA con principios éticos relacionados y partes interesadas

2. Definición de principios, marcos y directrices éticas

La inteligencia artificial (IA) está impactando en todas las facetas de la vida, afectando al trabajo, el ocio y ofreciendo soluciones a problemas globales como el cambio climático y el acceso a la atención médica. Ante la integración generalizada de la IA en las economías y las sociedades, se plantea la cuestión de los marcos políticos e institucionales adecuados necesarios para dirigir el desarrollo y la aplicación de la IA, garantizando el beneficio social.

Las amplias consideraciones éticas en el campo de la IA han impulsado la creación y publicación de numerosos enfoques para garantizar la aplicación ética de la inteligencia artificial. (Prem, 2023) Los principios, marcos y directrices éticas son de uso común y ampliamente reconocidos como la estructura central en la toma de decisiones éticas en muchas disciplinas. Proporcionan un enfoque integral para navegar por las cuestiones éticas, desde el establecimiento de valores fundamentales hasta la orientación de acciones específicas. Aunque a menudo están entrelazados, pueden describirse de la siguiente manera:

Los principios éticos proporcionan un marco moral para evaluar y resolver dilemas éticos. Los principios éticos forman los valores y creencias fundamentales que sirven de brújula para la toma de decisiones éticas y proporcionan una base para determinar lo que está bien y lo que está mal en diferentes contextos. (Prem, 2023; Zhou y Chen, 2022)

Los marcos éticos son enfoques o modelos sistemáticos que ayudan a las personas y organizaciones a navegar por los problemas éticos. Los marcos éticos son modelos integrales o métodos estructurados que facilitan el examen y la resolución de cuestiones éticas. Proporcionan un marco para evaluar las consecuencias éticas de una acción y guían el proceso de toma de decisiones, asegurando que las consideraciones éticas se tengan en cuenta sistemáticamente. (Prem, 2023; Zhou y Chen, 2022)

Las directrices éticas son reglas, estándares y mejores prácticas específicas que proporcionan orientación práctica sobre el comportamiento ético en diversos

contextos. Las directrices éticas son reglas detalladas y puntos de referencia que proporcionan consejos prácticos sobre cómo aplicar los principios éticos en diferentes escenarios. Estas directrices ayudan tanto a las personas como a las organizaciones a hacer frente eficazmente a los problemas éticos y a mantener un comportamiento ético en sus operaciones y en su toma de decisiones. (Prem, 2023; Zhou y Chen, 2022)

3. Esfuerzo común para dar forma al panorama ético

Los principios, marcos y directrices éticas provienen de muchos campos y organizaciones diferentes, incluidas empresas privadas, instituciones de investigación y organizaciones del sector público, lo que representa un esfuerzo común para dar forma al panorama ético de la inteligencia artificial. (Jobin et al., 2019; Morley et al., 2020)

Los documentos de orientación y los informes sobre la IA ética son ejemplos de herramientas políticas no legislativas o de «Derecho indicativo». A diferencia del "hard law", que incluye regulaciones legales establecidas por los cuerpos legislativos para imponer o prohibir ciertos comportamientos, las directrices éticas carecen de fuerza legal, pero tienen poder persuasivo. Estos documentos están concebidos para apoyar los procesos de adopción de decisiones en esferas específicas y se ha observado que ejercen una influencia considerable en las prácticas. (Jobin et al., 2019)

Ejemplos de documentos que definen los principios éticos de la inteligencia artificial son:

Los Principios de IA de la OCDE promueven un uso innovador y fiable de la IA que respete los derechos humanos y los valores democráticos. Adoptados en mayo de 2019, establecen estándares para la inteligencia artificial que pretenden ser prácticos y flexibles. (OCDE, 2019)

La Academia de Inteligencia Artificial de Pekín (BAAI) publicó los "Principios de IA de Pekín", un esquema para guiar la investigación y el desarrollo, la implementación y la gobernanza de la IA. ("Academia de Inteligencia Artificial de Beijing - Principios de IA de Beijing", 2019; *Principios de Inteligencia Artificial de Beijing*, 2022)

Los principales actores de la industria, como Google, IBM, Microsoft e Intel, han desarrollado sus propias pautas éticas que reflejan sus perspectivas y áreas de operación únicas. (Jobin et al., 2019; Morley et al., 2020)

Los organismos gubernamentales no se han quedado atrás, con documentos clave como la Declaración de Montreal y contribuciones de organismos como el Comité Selecto de la Cámara de los Lores sobre Inteligencia Artificial y el Grupo de Expertos de la Comisión Europea, cada uno involucrado en la regulación y la política ética de la IA. (Morley et al., 2020)

Instituciones académicas y grupos de expertos, como el Future of Life Institute, IEEE y AI4People, han profundizado el debate con investigación científica y modelos teóricos. (Floridi et al., 2018; Jobin et al., 2019; Morley et al., 2020)

La combinación de estos esfuerzos refleja un movimiento global hacia una IA responsable, y cada contribución forma una pieza del rompecabezas más amplio de la construcción de sistemas de IA éticos que estén en línea con los valores y normas de la sociedad.

CONSEJOS Y RECOMENDACIONES PARA LOS

Puede pedir a los alumnos que consulten las fuentes mencionadas anteriormente o algunas de ellas, y que analicen las diferencias en los principios que presentan.

4. Principios éticos

El panorama de los principios éticos en IA es rico y diverso, con numerosos principios propuestos por una variedad de actores, incluidas instituciones académicas, líderes de la industria, organismos gubernamentales y organizaciones internacionales. La proliferación de estos principios refleja un creciente consenso sobre la necesidad de una orientación ética en el desarrollo y la aplicación de tecnologías de inteligencia artificial. (Jobin et al., 2019; Morley et al., 2020; Prem, 2023)

Jobin et al (2019) llevaron a cabo una revisión exhaustiva de varios principios éticos de la IA y los destilaron para identificar temas y elementos comunes que se repiten en diferentes fuentes. Entre las directrices que revisaron, encontraron un consenso emergente sobre los principios éticos clave para la inteligencia artificial. Aunque no cuentan con un apoyo universal, los principios de convergencia incluyen la apertura, la equidad y la justicia, la responsabilidad, la reparación, la privacidad, la benevolencia, la libertad y la autonomía, la confianza, la sostenibilidad, la dignidad humana y la solidaridad.

Los elementos comunes identificados se refieren a un conjunto de preocupaciones y aspiraciones comunes en la comunidad mundial. Apuntan hacia un consenso emergente sobre los valores que muchos creen que deberían sustentar el desarrollo y el despliegue de sistemas de inteligencia artificial. Este acuerdo preliminar constituye un terreno común fundamental que puede servir como punto de partida para establecer expectativas y evaluar los resultados. (Morley et al., 2020)

Morley et al (2020) han reconocido este convenio colectivo como un punto de partida fundamental. Argumentan que estos principios acordados caracterizan a una IA éticamente adaptada de la siguiente manera:

- Beneficioso y respetuoso con las personas y el medio ambiente (beneficencia).
- Robusto y seguro (no maleficencia).
- Respetuoso de los valores humanos (autonomía).
- Justo (justicia).
- Explicable, responsable y comprensible (explicabilidad).

Beneficencia

El principio de beneficencia es una de las consideraciones éticas clave en el desarrollo de la inteligencia artificial, y enfatiza la importancia de los sistemas de inteligencia artificial que no solo benefician a las personas y a la sociedad, sino que también contribuyen al medio ambiente. Según Floridi et al. (2018), el objetivo es garantizar que la inteligencia artificial funcione de una manera que promueva el bienestar general y la salud ambiental.

Jobin et al. (2019) enumeran varias acciones clave para promover eficazmente la inteligencia artificial:

- Alinear la IA con los valores humanos y garantizar que los sistemas de IA apoyen y refuercen estos valores en lugar de socavarlos.
- Involucrar a las partes interesadas de la IA en su proceso de desarrollo y tener en cuenta sus necesidades y comentarios para mejorar el impacto de los sistemas de IA en sus vidas.
- Trabajar activamente para minimizar y resolver posibles conflictos de intereses, utilizando los comentarios de los usuarios para guiar el desarrollo y la aplicación de la IA.
- Crear nuevas formas de medir el bienestar humano.

CONSEJOS Y RECOMENDACIONES PARA LOS

Discuta con los estudiantes los impactos ambientales y sociales. Por ejemplo, los algoritmos de TikTok pueden proporcionar contenido dañino si el usuario muestra interés en él, lo que provoca efectos negativos para la salud mental. Además, la cantidad significativa de electricidad utilizada para crear modelos de lenguaje de gran tamaño puede dejar una huella de dióxido de carbono considerable.

No maleficencia

El principio ético de la no maleficencia está dedicado a la prevención del daño. Según Floridi et al. (2018), la no maleficencia abarca medidas proactivas contra el abuso intencional por parte de los humanos y las acciones negativas no intencionales de los sistemas de IA que pueden afectar el comportamiento humano. El objetivo principal es garantizar que la IA no provoque efectos nocivos, independientemente de su origen.

Para gestionar los riesgos y protegerse contra los daños en los sistemas de IA, Jobin et al. (2019) han identificado los siguientes enfoques:

- Utilizar la gestión estratégica de riesgos para anticipar y reducir los riesgos potenciales relacionados con los sistemas de IA.
- Implementación de una combinación de medidas técnicas y políticas de gobernanza que abarquen todo el ciclo de vida de la IA, con el objetivo de prevenir daños antes de que ocurran.
- Ciertos daños pueden ser inevitables a pesar de todos los esfuerzos, lo que requiere una evaluación de riesgos exhaustiva. Esto incluye medidas para reducir y mitigar los riesgos y asignar claramente la responsabilidad cuando se produce un daño.

CONSEJOS Y RECOMENDACIONES PARA LOS

Discuta con los estudiantes los posibles daños causados por la IA. Eg. La herramienta de reclutamiento de IA de Amazon mostró un sesgo contra las candidatas. Verifique para obtener más detalles, por ejemplo. de aquí: Insight: Amazon desecha la herramienta secreta de reclutamiento de IA que mostró prejuicios contra las mujeres | Reuters

Autonomía

La autonomía en el contexto de la IA se centra en mantener un equilibrio entre la toma de decisiones humana y la influencia de la IA. Los seres humanos deben mantener el control de la toma de decisiones y elegir cuándo tomar sus propias decisiones y cuándo dejar que la IA decida. Sin embargo, los humanos siempre deben tener la capacidad de recuperar el control de la IA si es necesario, lo que significa que tienen la última palabra. (Floridi et al., 2018)

Jobin et al (2019) identifican una colección de medidas para promover la libertad y la autonomía en los sistemas de IA:

- Mejorar la transparencia y la previsibilidad de los sistemas de IA para que los usuarios puedan comprender y anticipar el comportamiento de la IA.

- Mejorar la comprensión de las tecnologías de IA por parte del público en general.
- Implementación de protocolos sólidos de notificación y consentimiento para garantizar la autonomía individual al interactuar con los sistemas de IA.
- Evitar la recogida y distribución de datos personales sin el consentimiento expreso e informado de las personas interesadas, respetando su privacidad e independencia.

CONSEJOS Y RECOMENDACIONES PARA LOS

Discuta con los estudiantes los posibles problemas de autonomía con la IA. Eg. Facebook proporciona herramientas para que los usuarios ejerzan control sobre su contenido, pero ¿puede la complejidad y la opacidad del algoritmo socavar a veces la verdadera autonomía del usuario?

Justicia

La justicia es un principio polifacético que pretende utilizar la IA como herramienta para corregir las injusticias del pasado, garantizar que los beneficios de la IA se distribuyan de manera justa y proteger contra la aparición de nuevos daños o perturbaciones sociales. (Floridi et al., 2018)

La realización de la justicia en la IA, tal y como analizan Jobin et al. (2019), incluye varias medidas proactivas:

- Establecer estándares y normas técnicas.
- Aumentar la transparencia de los derechos y las regulaciones.
- Implementación de pruebas, seguimiento y auditoría, especialmente en unidades de protección de datos.
- Fortalecimiento del estado de derecho, incluidas las apelaciones y los recursos legales.
- Se implementan cambios sistémicos, como la supervisión gubernamental, equipos versátiles y una sociedad civil inclusiva, con un enfoque en la distribución justa de los beneficios.

CONSEJOS Y RECOMENDACIONES PARA LOS

Compruebe, por ejemplo, el caso del chatbot Tay de Microsoft: en 2016, el chatbot racista de Microsoft reveló los peligros de la conversación en línea - IEEE Spectrum. Refiriéndose a la justicia; El algoritmo priorizó injustamente el

contenido sensacionalista sobre el discurso significativo y constructivo, lo que llevó a una representación desequilibrada de puntos de vista y voces en la plataforma, lo que creó interrupciones sociales.

Explicabilidad

La explicabilidad es crucial para crear y mantener la confianza de los usuarios en los sistemas de IA. Esto significa que los procesos deben ser transparentes, las capacidades y el propósito de los sistemas de inteligencia artificial deben comunicarse abiertamente y las decisiones deben explicarse a los afectados directa e indirectamente. (Grupo de Expertos de Alto Nivel sobre IA (AI HLEG), 2019) Floridi et al (2018) también consideran que la explicabilidad es fundamental en la aplicación de otros principios éticos.

Para aumentar la transparencia en los sistemas de IA, Jobin et al. (2019) enumeran varios enfoques:

- Se anima a los desarrolladores y usuarios de IA a compartir más información sobre sus sistemas, desmitificando los procesos que hay detrás de la toma de decisiones de IA.
- Explicaciones de las operaciones y decisiones de la IA de forma que sean fácilmente comprensibles para los no expertos, lo que garantiza que las operaciones de la IA puedan ser auditadas por humanos que no sean necesariamente expertos en la materia.
- El uso de auditorías como una forma de garantizar la transparencia en la IA, lo que puede ayudar a identificar y abordar posibles sesgos o problemas éticos.
- Establecimiento de mecanismos que supervisen el rendimiento de la IA, compromiso con las partes interesadas y el público para recopilar diversas perspectivas, y mecanismos para apoyar la denuncia de irregularidades.

CONSEJOS Y RECOMENDACIONES PARA LOS

Pregunte a los estudiantes si se sienten lo suficientemente informados sobre las decisiones tomadas por las soluciones de IA que están utilizando. Por ejemplo, la herramienta de reclutamiento de IA de Amazon carecía de transparencia en su proceso de toma de decisiones. Los candidatos y los gerentes de contratación no recibieron explicaciones claras de por qué se rechazaba a ciertos candidatos, lo que dificultaba desafiar o abordar los sesgos subyacentes en el sistema.

¡NOTA! Al final de la sección de principios, utilice el estudio de caso proporcionado como trabajo en grupo.

5. Marcos éticos, directrices y conjuntos de herramientas

Una vez que se han identificado los principios éticos, estos deben transformarse en marcos y estrategias viables para guiar las innovaciones basadas en la IA y la implementación práctica de la ética de la IA.

Según Ayling y Chapman (2022), los marcos éticos traducen los principios éticos en medidas prácticas. Este proceso implica la creación de directrices y procedimientos que los desarrolladores y adoptantes de tecnologías de IA puedan adoptar, incorporando así consideraciones éticas en cada paso, desde el diseño hasta las aplicaciones del mundo real.

Prem (2023) amplía esto al afirmar que muchos marcos éticos de IA buscan activamente anticipar y abordar posibles problemas y riesgos éticos. Al identificar de forma proactiva estos desafíos, los marcos guían a los desarrolladores en la creación de soluciones de IA. Este enfoque proactivo es esencial en un campo que está evolucionando tan rápidamente como la inteligencia artificial, donde pueden surgir nuevas consideraciones éticas a medida que la tecnología se desarrolla y se integra más en diferentes aspectos de la vida cotidiana.

Un marco ético proporciona un plan integral para establecer normas regulatorias y operativas que armonicen con las normas éticas. Como señalan Floridi y Cowls (2019), dicho marco es fundamental para la elaboración de legislación, la formulación de normas, el establecimiento de normas técnicas y la identificación de las mejores prácticas que son aplicables en una amplia gama de industrias y regiones geográficas.

Prem (2023, p. 701) define los componentes esenciales de los marcos éticos de la siguiente manera:

- Conceptos básicos relacionados con la discusión de los aspectos éticos
- Principios éticos (por ejemplo, valores)
- **Preocupación por** cómo el uso y desarrollo de los sistemas de inteligencia artificial se adhieren a los principios éticos.
- **Soluciones** para abordar las preocupaciones (por ejemplo, estrategias, reglas y directrices)

Son muy necesarias las herramientas y directrices para aplicar los principios éticos en el diseño, la implementación y el despliegue. (Zhou et al., 2020) Los conjuntos de herramientas y directrices éticas son herramientas importantes para dirigir la innovación de la IA en una dirección que esté en consonancia con las normas y los valores éticos. Se considera que los conjuntos de herramientas y directrices éticas son importantes para cerrar la brecha entre los principios éticos abstractos y su integración concreta en las prácticas de IA.

Si bien la importancia de los conjuntos de herramientas y las directrices es clara, traducir los principios en herramientas concretas es un desafío y muchos marcos existentes carecen de contexto de aplicación. (Prem, 2023, p. 702) La razón de la

gran brecha entre los principios y la práctica se debe a la falta de estandarización, la variabilidad, la complejidad y la comprensión subjetiva de los principios más bien abstractos.

El rápido ritmo reciente del desarrollo de la IA ha llevado a un crecimiento sustancial tanto en el número como en la diversidad de las herramientas disponibles, con la constante aparición de nuevos desarrollos. Morley et. al (2019) y Prem (2023) han revisado cientos de herramientas diferentes para abordar los desafíos éticos en las soluciones de IA. Esas herramientas se dividen por las etapas de diseño de soluciones de IA (definidas posteriormente a partir del capítulo H) y por diferentes principios éticos. Sin embargo, muchas herramientas se perciben como difíciles de usar o se considera que requieren una cantidad sustancial de esfuerzo por parte de sus usuarios.

Las listas compiladas por Morley et. al (2019) y Prem (2023) muestran claramente que la disponibilidad de las herramientas y métodos se divide de manera desigual entre los diferentes principios y etapas de diseño. Por ejemplo, las herramientas para medir la beneficencia de la solución son numerosas en las primeras etapas del desarrollo, mientras que las herramientas de explicabilidad se posicionan principalmente en la etapa posterior del proceso de desarrollo. Consulta las herramientas a través de este enlace: Tipología ética aplicada a la IA - Google Docs.

Los métodos utilizados también varían según la etapa de desarrollo. Los marcos, las bibliotecas y los algoritmos suelen desempeñar un papel más importante en la fase ex-ante, concretamente en la fase de diseño de la solución. Sin embargo, algunos marcos alcanzan tanto la fase ex ante como la ex post (UNESCO, 2023b, p. 5) En la fase de prueba, de nuevo, se presentan predominantemente auditorías y métricas, mientras que en la cara post ante se utilizan principalmente declaraciones, etiquetas y licencias como enfoques. (Prem, 2023, p. 709)

El estudio también reveló que solo unas pocas herramientas abordaban el impacto de la solución en un individuo o en una sociedad en su conjunto. Esto podría deberse a la razón por la que es difícil cambiar el complejo comportamiento humano en herramientas de diseño simples y generales. Sin embargo, sería esencial abordar el impacto humano y social para lograr la aceptación y la preferencia social de las soluciones de IA. (Morley et al., 2020, p. 2156)

6. De los principios a la práctica; Marco de IA confiable de HLEG

Aunque existe consenso en que la IA debe ser ética, continúan los debates sobre la definición exacta de "IA ética" y las obligaciones éticas y los criterios técnicos necesarios para implementarla. (Leijnen et al., 2020)

La Comisión Europea ha puesto en marcha una estrategia integral para el avance de la inteligencia artificial (IA) en Europa, haciendo especial hincapié en garantizar que la IA desarrollada y utilizada no solo sea avanzada desde el punto de vista tecnológico, sino que también se adhiera a las normas éticas y sea segura.

Para aplicar eficazmente esta estrategia y sus principios subyacentes, la Comisión Europea creó el Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial (GEAN IA). A este grupo se le encomendó la tarea de desarrollar directrices éticas y recomendaciones de políticas e inversiones en materia de IA, destinadas a guiar el desarrollo, el despliegue y el uso éticos de la IA en toda Europa, fomentando la innovación y salvaguardando los valores y derechos fundamentales. [Grupo de Expertos de Alto Nivel sobre IA (AI HLEG), 2019]

El marco de IA fiable está diseñado para garantizar que las tecnologías de IA se desarrollen y desplieguen de manera éticamente sólida, respetando los derechos fundamentales y garantizando la solidez y la fiabilidad. El marco de referencia para una IA fiable se compone de tres componentes, cada uno diseñado para guiar el desarrollo y la implementación de la IA de manera ética y fiable. [Grupo de Expertos de Alto Nivel sobre IA (AI HLEG), 2019]

Fundamentos de una IA fiable

Este segmento establece los principios básicos en los que se basa una IA fiable, adoptando un enfoque basado en los derechos fundamentales. Describe los principios éticos esenciales para garantizar que los sistemas de IA se desarrollen y utilicen de manera ética y sólida.

Los principios éticos forman la base sobre la que se debe construir una IA fiable, garantizando el cumplimiento de las normas éticas. Estos principios éticos son: (i) Respeto a la autonomía humana, (ii) Prevención de daños, (iii) Equidad y (iv) Explicabilidad.

Realización de una IA fiable

Sobre la base de los principios éticos esbozados, la segunda parte del marco traduce estos principios en siete requisitos clave que se espera que los sistemas de IA incorporen y mantengan a lo largo de su ciclo de vida. Esta transformación de los principios éticos en requisitos prácticos garantiza que la ética fundamental

no sean solo ideales teóricos, sino que se integren activamente en todas las etapas del desarrollo y el funcionamiento de un sistema de IA.

La lista de requisitos incluye: Agencia y supervisión humanas: derechos fundamentales, agencia humana y supervisión humana; Robustez técnica y seguridad: resistencia a los ataques y seguridad, plan de respaldo y seguridad general, precisión, fiabilidad y reproducibilidad; Privacidad y gobernanza de datos: respeto de la privacidad, la calidad y la integridad de los datos, y el acceso a los datos; Transparencia: trazabilidad, explicabilidad y comunicación; Diversidad, no discriminación y equidad: evitar sesgos injustos, accesibilidad y diseño universal, y participación de las partes interesadas; Bienestar social y medioambiental: sostenibilidad y respeto al medio ambiente, impacto social, sociedad y democracia; Rendición de cuentas: auditabilidad, minimización y presentación de informes sobre el impacto negativo, compensaciones y compensación.

Evaluación de la IA fiable

La tercera parte del marco proporciona una lista concreta y completa para la evaluación de la IA confiable, diseñada para hacer operativos los requisitos antes mencionados.

Conocida como la "Lista de Evaluación para una IA Confiable (ALTAI)", esta herramienta proporciona una lista de verificación completa y dinámica que sirve como hoja de ruta para que los desarrolladores e implementadores integren estos principios en sus aplicaciones de IA. ALTAI facilita este proceso al ofrecer un conjunto de pasos concretos para la autoevaluación, asegurando así que las tecnologías de IA se desarrollen de una manera que maximice los beneficios para el usuario y minimice la exposición a riesgos innecesarios. (Grupo de Expertos de Alto Nivel sobre IA (AI HLEG), 2020)

Esta lista de evaluación no exhaustiva sirve como guía práctica para los profesionales de la IA, ya que ofrece orientación específica sobre cómo implementar estos requisitos de manera efectiva. Es importante destacar que esta evaluación debe ser adaptable, lo que permite la personalización de acuerdo con la aplicación específica de cada sistema de IA, lo que garantiza la pertinencia y la eficacia en diversos contextos.

7. De los principios al derecho: LEY DE IA DE LA UE

La Ley de IA regulará la aplicación de la inteligencia artificial en la UE, lo que supondrá la primera legislación amplia sobre IA a nivel mundial. (Parlamento Europeo, 2023) El objetivo de AI ACT es mejorar el funcionamiento del mercado interior y promover la adopción de una IA fiable y centrada en el ser humano, garantizando al mismo tiempo un alto nivel de protección de la salud, la seguridad y los derechos fundamentales, incluida la democracia, el Estado de Derecho y la protección del medio ambiente frente a los efectos nocivos de los sistemas de inteligencia artificial en la Unión, y apoyando la innovación. (Instituto del Futuro de la Vida, 2024) Se prevé que la Ley entrará en vigor en 2026. (Hillard y Gulley, 2023)

La Comisión Europea ha organizado las regulaciones de IA dentro de un marco basado en el riesgo, creando un sistema de categorías de riesgo. Las aplicaciones de IA se regularán según el nivel de riesgo que presenten. (Comisión Europea, 2023). El enfoque basado en el riesgo descrito en la Ley de IA de la UE clasifica los sistemas de IA en cuatro niveles en función de su impacto potencial: riesgo inaceptable, riesgo alto, riesgo limitado y riesgo bajo o mínimo. (Comisión Europea, 2023; Hillard y Gulley, 2023)

Riesgo mínimo

La mayoría de los sistemas de IA entran en esta categoría, ya que presentan un riesgo mínimo o nulo para los derechos o la seguridad de los ciudadanos. Algunos ejemplos son los sistemas de recomendación habilitados para IA y los filtros de spam. Las empresas que operan estos sistemas no están obligadas a cumplir con requisitos estrictos, pero pueden comprometerse voluntariamente con códigos de conducta adicionales para mejorar la transparencia y la rendición de cuentas.

Riesgo limitado

Esta categoría se refiere a los sistemas de IA que representan un riesgo limitado para los derechos y la seguridad de los usuarios. Algunos ejemplos son los chatbots y los deep fakes. Si bien no están sujetos a requisitos regulatorios estrictos, los proveedores deben asegurarse de que los usuarios sepan cuándo interactúan con los sistemas de IA y etiquetar el contenido generado por IA en consecuencia. Esto garantiza la transparencia y la concienciación sobre el uso de las tecnologías de IA.

Alto riesgo

Los sistemas de IA identificados como de alto riesgo deben cumplir requisitos estrictos para mitigar los riesgos potenciales. Estos requisitos incluyen sistemas sólidos de mitigación de riesgos, conjuntos de datos de alta calidad, registro de actividades, documentación detallada, información clara para el usuario, supervisión humana y sólidas medidas de ciberseguridad. Los sandboxes

regulatorios se establecen para fomentar la innovación responsable y facilitar el desarrollo de sistemas de IA que cumplan con las normas. Algunos ejemplos de sistemas de IA de alto riesgo son las infraestructuras críticas, los dispositivos médicos, los sistemas de acceso a las instituciones educativas y determinadas aplicaciones en la aplicación de la ley y el control de fronteras.

Riesgo inaceptable

Se prohíben los sistemas de IA que supongan una amenaza clara para los derechos fundamentales. Esto incluye sistemas que manipulan el comportamiento humano para socavar el libre albedrío, como los juguetes que fomentan el comportamiento peligroso en los menores, o aquellos que permiten la "puntuación social" del gobierno o de las empresas. También se prohíben las aplicaciones policiales predictivas y ciertos usos de los sistemas biométricos, como el reconocimiento de emociones en los lugares de trabajo.

CONSEJOS Y RECOMENDACIONES PARA LOS

1) Discuta con los estudiantes si creen que la ley de la UE será más estricta que otras leyes y qué causará. Alimento para las reflexiones, por ejemplo, de este artículo: https://news.bloomberglaw.com/artificial-intelligence/eu-embraces-new-ai-rules-despite-doubts-it-got-the-right-balance
2) Si pensamos en la reciente discusión sobre TikTok y lo dañino que es, ¿en qué categoría de riesgo entraría?

8. Ciclo de vida del desarrollo de la IA

Como se explora en el Capítulo E, los principios éticos específicos se vuelven más relevantes en ciertas etapas del desarrollo de sistemas de IA, y se utilizan varias herramientas en consecuencia. Es crucial comprender el proceso de desarrollo desde una perspectiva amplia. (Prem, 2023, p. 703) En este curso, adoptamos el término "ciclo de vida de desarrollo de la IA" para encapsular esta visión integral de las etapas de desarrollo.

La vertiginosa industria de las soluciones de IA a veces es testigo de fracasos de proyectos debido a procesos de gestión de proyectos inadecuados. Un ciclo de vida de IA bien definido, que tenga en cuenta todos los pasos de desarrollo, puede mejorar la tasa de éxito de las soluciones de IA. Si bien existen diferentes versiones de las descripciones del ciclo de vida de la IA, nuestro enfoque sintetiza los modelos propuestos por Morley et al. (2020), Prem (2023) y el modelo CRISP de Data Science Process Alliance (Hotz, 2018) presentado por Rochel y Evequoz (2021).

El ciclo de vida del desarrollo de IA describe el recorrido que realiza un sistema de IA desde su concepción inicial hasta su implementación y mantenimiento finales. Cada fase es vital para garantizar un resultado exitoso. El ciclo de vida del diseño es iterativo, lo que significa que puede ser necesario volver a visitar las etapas anteriores si ciertas variables no funcionan como se esperaba. Aunque las consideraciones éticas pueden variar en las diferentes etapas, es esencial integrarlas en cada fase del proceso (Saltz, 2023).

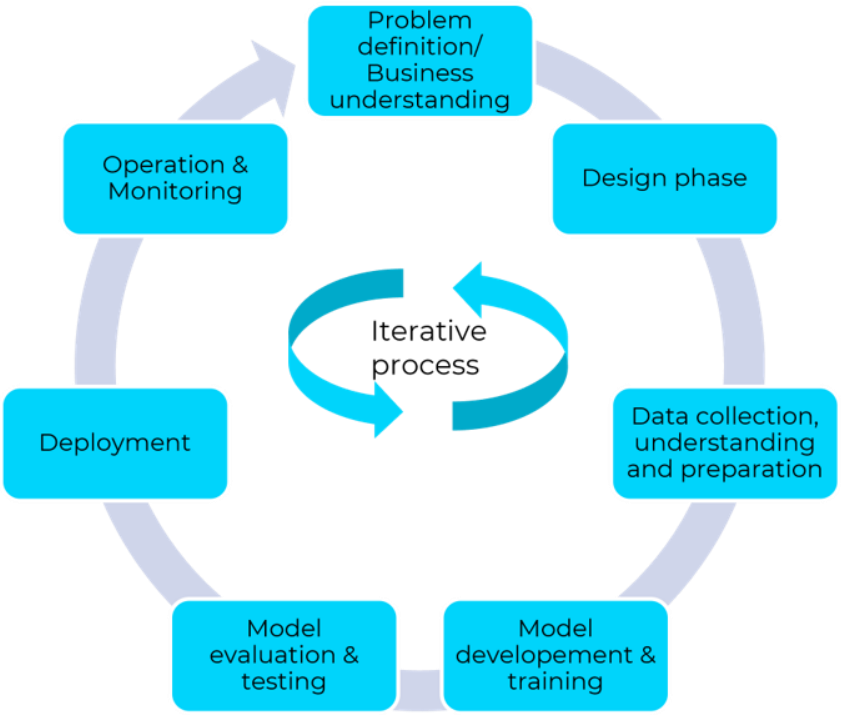


Foto: Autores del documento. Modificado como combinación de los diferentes modelos descritos por la Data Science Process Alliance (2023), Morley et.al' (2019), Prehm's (2022) y Rochel et. Todos los artículos (2020).

9. Participación de las partes interesadas en el ciclo de vida del desarrollo de la IA

Involucrar a una amplia gama de partes interesadas en el proceso de desarrollo de IA es crucial para evaluar a fondo los impactos éticos de un proyecto de IA desde múltiples perspectivas. Consultar a las partes interesadas, incluidas las directamente afectadas por la solución de IA, es esencial para construir sistemas fiables (Morley et al., 2020) y para realizar una evaluación exhaustiva de los impactos del proyecto (UNESCO, 2023a, p. 15). En este capítulo se ofrece una visión general de las diferentes partes interesadas que participan en los proyectos de desarrollo de IA y se describen sus funciones en las distintas etapas del ciclo de vida del desarrollo de la IA.

Las partes interesadas se dividen en partes interesadas en el desarrollo, partes interesadas en el uso y partes interesadas externas. Las partes interesadas en el desarrollo y el uso son partes interesadas activamente involucradas, es decir, los responsables de la toma de decisiones que tienen el poder de decidir y los diseñadores con la experiencia. Las partes interesadas externas no participan en el proyecto de IA, pero pueden afectar o verse afectadas por él o ser consultadas (Miller, 2022).

A continuación, se examinan las funciones clave de las distintas partes interesadas en el desarrollo de una inteligencia artificial y se clasifican en tres categorías: desarrollo, uso y partes interesadas externas. Además, se exploran las responsabilidades y contribuciones específicas de cada grupo en las diferentes etapas del ciclo de vida del desarrollo de la IA. Comprender estos roles es esencial para navegar por las complejidades del desarrollo de IA y garantizar el éxito del proyecto.

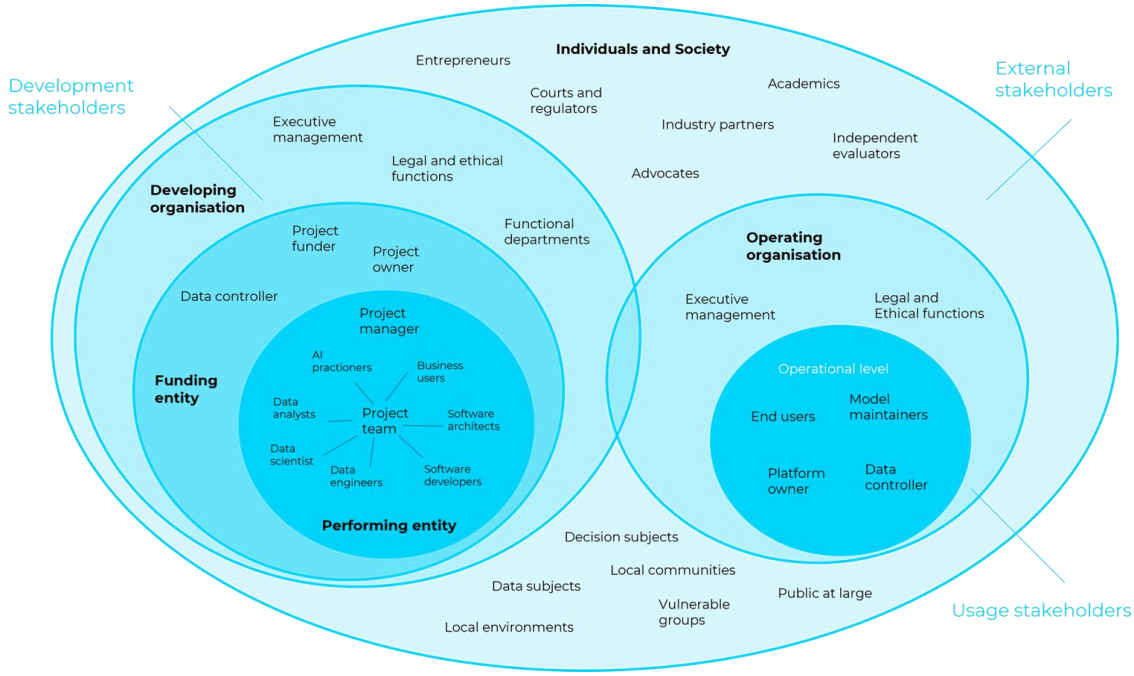


Imagen: G.J. Miller (2022)

Partes interesadas en el desarrollo

En el ámbito del desarrollo de proyectos de IA, Miller (2022) divide a las partes interesadas en tres categorías principales: la organización, la entidad financiadora y la entidad ejecutora, cada una de las cuales desempeña un papel distinto en el ciclo de vida del proyecto. La organización se beneficia de los resultados del proyecto de IA, aprovechando los avances e innovaciones para su crecimiento y éxito. La entidad financiadora, crucial para el apoyo financiero, no solo financia, sino que también gobierna el proyecto, asegurando que las inversiones se alineen con los objetivos del proyecto y los estándares de gobernanza. La entidad ejecutora es fundamental, ya que se encarga de generar los resultados del proyecto y de elaborar el sistema de IA, que encarna la ejecución técnica del proyecto. (Miller, 2022)

Dentro de la organización, las funciones abarcan desde la organización en desarrollo, que patrocina y define el alcance del proyecto, hasta la gestión ejecutiva, que supervisa la trayectoria del proyecto. Los departamentos funcionales proporcionan un apoyo esencial, mientras que las funciones legales y éticas garantizan que el proyecto se alinee con los estándares regulatorios y las normas éticas. El grupo de gestión de proyectos desempeña un papel de apoyo,

ayudando a los gerentes de proyectos con tareas administrativas para agilizar la ejecución del proyecto (Zwikael y Meredith, 2018).

La entidad financiadora abarca funciones como el financiador del proyecto, que suministra los recursos financieros, y el propietario del proyecto, que imparte dirección estratégica y potencialmente preside el comité directivo, asegurando la alineación del proyecto con los objetivos estratégicos. En escenarios en los que el financiador no está directamente involucrado, se delega la responsabilidad para garantizar una supervisión continua. El controlador de datos, un rol a menudo asociado con la entidad financiadora, administra el uso de datos, abordando el cumplimiento de marcos regulatorios y legales estrictos para evitar posibles sanciones (Miller, 2022).

En el corazón de la entidad ejecutora se encuentran los gerentes de proyecto y el equipo del proyecto, responsables de entregar los productos del proyecto según el plan aprobado. Esta colaboración garantiza que el proyecto avance desde su concepción hasta su realización, y que cada miembro del equipo contribuya al desarrollo y la implementación exitosa del sistema de IA (Zwikael y Meredith, 2018).

Esta organización estructurada de las partes interesadas y sus funciones subraya el enfoque colaborativo y multifacético necesario para el desarrollo y la implementación exitosos de proyectos de IA, destacando la importancia de roles y responsabilidades claros y el cumplimiento de las normas legales y éticas.

Partes interesadas en el uso

En el contexto de los proyectos de IA, las partes interesadas en el uso desempeñan un papel crucial, especialmente a nivel organizativo, donde la organización operativa funciona como una parte interesada del cliente que adquiere la solución de IA. Al igual que la organización de desarrollo, la organización operativa abarca roles como la gestión ejecutiva y las funciones legales y éticas, destacando las estructuras paralelas tanto en la fase de desarrollo como en la operativa (Miller, 2022).

A nivel operativo, la etapa de uso involucra a una variedad de actores clave, incluidos los custodios de datos, los usuarios finales, los propietarios de plataformas y los mantenedores de modelos. Los usuarios finales, ya sean individuos o grupos, interactúan con el sistema de IA ya sea a través de contratos de trabajo o servicios con la organización operativa o como consumidores, interactuando directamente con las funcionalidades del sistema (Miller, 2022).

Los mantenedores de modelos asumen la tarea crítica de gestionar, actualizar y mejorar continuamente los modelos de aprendizaje automático que impulsan el sistema de IA, garantizando su eficiencia y eficacia. Los controladores de datos se encargan de definir los objetivos y métodos para el procesamiento de datos personales dentro del sistema, asegurando el cumplimiento de las leyes de

protección de datos. Por último, los propietarios de plataformas son responsables de la infraestructura general del sistema de IA y de su despliegue, manteniendo los aspectos fundamentales sobre los que opera la IA (Miller, 2022).

Grupos de interés externos

Miller (2022) clasifica a las partes interesadas externas en los proyectos de IA en tres subgrupos principales: individuos, sociedad y representantes, cada uno con distintas funciones y contribuciones al proyecto.

El subgrupo de individuos abarca a los sujetos de datos, los sujetos de decisiones y los trabajadores que interactúan con la organización en desarrollo u operativa a través de acuerdos como contratos de servicio o de trabajo. Estas partes interesadas incluyen a las personas directamente involucradas o afectadas por el proyecto. Estas partes interesadas desempeñan un papel crucial como sujetos de datos, sujetos de decisiones y trabajadores dentro del proyecto (Miller, 2022).

El subgrupo de la sociedad capta el impacto más amplio en la población en general, incluidas las comunidades locales, el medio ambiente y los grupos vulnerables, como las personas con discapacidades, los grupos minoritarios, los menores y el público en general. Este segmento está directamente influenciado por el desarrollo y la utilización del sistema de IA, lo que enfatiza el alcance social de los proyectos de IA (Miller, 2022).

Los representantes, que forman el tercer subgrupo, se distinguen de los representantes formales. Si bien los representantes incluyen grupos y organizaciones como empresarios y evaluadores independientes (reporteros, auditores o revisores) que pueden no verse directamente afectados por el sistema de IA pero desempeñan un papel en su desarrollo o evaluación, los representantes formales, como los tribunales, las leyes, los reguladores y los sindicatos, representan la supervisión legal y regulatoria y la gobernanza dentro del proyecto (Miller, 2022; Rodrigues, 2020).

Participación de las partes interesadas en el desarrollo de soluciones de IA

La co-creación se refiere al compromiso colaborativo con las partes interesadas, incluidos los usuarios finales o los clientes, a lo largo del proceso de diseño. Este enfoque facilita el intercambio de ideas entre participantes con diversas funciones y experiencias. (Fundación de Diseño de Interacción - IxDF, 2021; Sanín, 2020). La práctica de la co-creación se puede llevar a cabo a través de talleres facilitados que abarcan una variedad de actividades, como rompehielos, juegos de rol, mapeo de viajes, lluvia de ideas y creación de prototipos. Al adoptar la co-creación, los diseñadores pueden comprender profundamente las necesidades y deseos de los clientes o usuarios finales, creando así soluciones que tienen en cuenta las diversas perspectivas de los usuarios.

10. Ciclo de vida del desarrollo de la IA: definición del problema y comprensión del negocio

CONSEJOS Y RECOMENDACIONES PARA LOS

En esta sección se explicará en detalle el ciclo de vida de desarrollo de una solución de IA, sus implicaciones éticas y las partes interesadas relacionadas. Para fomentar una discusión vívida, utilice un caso de ejemplo para recorrer todo el ciclo de vida. A modo de ejemplo, puede utilizar el sistema de recomendación basado en IA sobre qué comprar a través de este enlace: https://www.datascience-pm.com/ai-lifecycle/#An_Example_Ai_Project_Life_Cycle

Como cualquier proyecto, el ciclo de vida de la IA también debe comenzar con la definición del problema. El planteamiento del problema bien definido proporciona una comprensión profunda de las necesidades del cliente y, por lo tanto, una dirección para el proyecto. En este punto, se debe realizar un estudio de viabilidad para comprender si la IA es la solución más adecuada para abordar el problema. Además, la investigación y la participación de las partes interesadas son necesarias para comprender el problema a fondo. (De Silva y Alahakoon, 2022, pp. 4-5)

Principios éticos en la etapa de definición de problemas y comprensión del negocio

En la fase inicial de un proyecto, es crucial tener en cuenta los principales aspectos éticos, ya que sientan las bases para el camino del proyecto y el posterior impacto social. Particularmente en esta etapa, los principios de beneficencia y no maleficencia se consideran primordiales. (Prem, 2023, p. 703)

En cuanto a la beneficencia, la atención se centra en evaluar si el sistema de IA fomenta activamente el bienestar individual y el beneficio social. Los objetivos del sistema deben ser transparentes, buscando beneficios tangibles, respetando al mismo tiempo los derechos fundamentales y teniendo en cuenta la sostenibilidad ambiental. Es esencial involucrar a una variedad de partes interesadas, asegurando que las perspectivas de los afectados por la solución de IA sean escuchadas e integradas. (De Silva y Alahakoon, 2022, p. 5; Morley et al., 2020, p. 2151)

Por el contrario, la no maleficencia implica garantizar que el sistema de IA no inflija daño a las personas o a las comunidades. Esto requiere un enfoque proactivo para identificar y mitigar los posibles efectos adversos, salvaguardando atentamente el bienestar físico y mental de las personas afectadas por el sistema. (De Silva y Alahakoon, 2022, p. 5; Morley et al., 2020, p. 2151) Las preocupaciones generales de seguridad deben abordarse de manera proactiva e incorporarse en el diseño del sistema (UNESCO, 2023b, p. 18).

Desde el punto de vista de la justicia, es imperativo examinar las implicaciones más amplias del sistema de IA en las instituciones, la democracia y las estructuras sociales. Es vital considerar si el proyecto podría crear inadvertidamente sesgos, favoreciendo a ciertos grupos sobre otros, o invadir la autonomía individual. Cada uno de estos aspectos exige una consideración cuidadosa para garantizar que el sistema de IA contribuya a una sociedad justa y equitativa. (Morley et al., 2020, p. 2151)

Stakeholders en la etapa de definición de problemas y comprensión del negocio

En esta etapa, es esencial involucrar al menos a los usuarios finales, que son los principales usuarios de la solución de IA. Además, los desarrolladores de IA deben incluirse desde el principio para obtener una comprensión profunda del problema que tendrán la tarea de resolver. Los especialistas en ética o los científicos sociales también deben participar para evaluar los posibles impactos humanos y sociales de las soluciones. También es crucial incluir a los expertos en la materia para aumentar el conocimiento sobre la industria en la que se aplica la IA. (De Silva y Alahakoon, 2022)

11. Ciclo de vida del desarrollo de la IA: etapa de diseño

Durante la etapa de diseño, el trabajo fundamental realizado en la etapa de definición del problema se convierte en un plan integral para la implementación del proyecto del sistema de IA. Esta fase incluye la definición de los objetivos del proyecto y los resultados deseados, el desarrollo de un plan de proyecto con plazos y consideraciones presupuestarias, y la comprensión de los requisitos de las partes interesadas. (De Silva y Alahakoon, 2022, p. 4)

Las actividades clave incluyen la elaboración de una especificación de requisitos que formule las funciones del sistema de IA y las métricas para el éxito del proyecto. Además, se refinan las estrategias para la minería de datos, incluidas, por ejemplo, las decisiones sobre el uso de modelos previamente entrenados. A lo largo de esta etapa, las consideraciones éticas son integrales y deben tenerse en cuenta en todos los aspectos del diseño del sistema. (De Silva y Alahakoon, 2022, p. 3)

Principios éticos en la etapa de diseño

Durante la fase de diseño de la solución de IA, es crucial emplear estrategias proactivas de "qué pasaría si" para prever y mitigar los posibles desafíos a lo largo del ciclo de vida de la solución de IA. Al plantear preguntas críticas como "¿Qué pasa si los datos utilizados están sesgados?" y explorar los mecanismos necesarios para detectar y abordar el comportamiento injusto, los desarrolladores pueden abordar estos problemas de manera preventiva. El establecimiento de medidas concretas es imperativo para dotar a los desarrolladores de los conocimientos necesarios para evitar sesgos relacionados, por ejemplo, con la raza, el color, la ascendencia, el género, la edad, el idioma, la religión, la opinión política, el origen nacional o étnico, el origen social, la situación económica, las condiciones de nacimiento o la discapacidad. Además, es esencial establecer canales de comunicación sólidos, protocolos de denuncia y medidas que sean rápidas para reaccionar a los sesgos y los ajustes que puedan requerir. (Morley et al., 2020, págs. 2151, 2155; UNESCO, 2023b, pp. 13, 17)

Además de estas precauciones, los requisitos de diseño deben articular claramente los métodos para aumentar la transparencia del proceso, mejorando así la explicabilidad. Este enfoque subraya la importancia de hacer que las operaciones del sistema de IA sean comprensibles y responsables (Morley et al., 2020, pp. 2151, 2155; UNESCO, 2023b, pp. 13 y 17).

Además, Prem (2023, p. 703) enfatiza el valor de involucrar a las partes interesadas durante esta etapa y el papel fundamental de los mecanismos de supervisión humana. Esta inclusión garantiza que una amplia gama de perspectivas contribuya al desarrollo de la IA, promoviendo un proceso de diseño más equitativo e informado.

Partes interesadas en la etapa de diseño

En esta fase es importante involucrar a los arquitectos de software, ya que desempeñan un papel crucial en el establecimiento de las bases técnicas de la solución de IA y tienen en cuenta tanto los requisitos actuales como la escalabilidad futura (Peter, 2024). Los diseñadores de UX son esenciales para crear interfaces que faciliten la adopción y la satisfacción de los usuarios y garantizar que los usuarios puedan interactuar sin problemas con el sistema de IA (White, 2023). Los profesionales de la IA deben colaborar estrechamente con los diseñadores de UX para integrar los componentes de IA sin problemas en la interfaz de usuario. Los expertos en ética abordan las preocupaciones éticas en el desarrollo de la IA. Se necesitan funciones legales y éticas para reducir los riesgos legales asociados con los proyectos de IA.

Según De Silva y Alahakoon (2022), se recomienda utilizar un enfoque de diseño participativo que involucre a las personas y comunidades afectadas por el modelo de IA y sus decisiones, fomentando la transparencia, la inclusión y la alineación con los valores sociales a lo largo del ciclo de vida del proyecto.

12. Ciclo de vida del desarrollo de la IA: recopilación, comprensión y preparación de datos

Después de establecer el objetivo y diseñar una estrategia, la siguiente fase consiste en adquirir y preparar datos relevantes. Esto incluye:

- Recopilación de datos
- Definiéndolo y categorizándolo
- Realización de análisis exploratorios
- Validando su pertinencia
- Selección de información esencial
- Limpiarlo, reconstruirlo, integrarlo y formatearlo para que se ajuste a los objetivos del proyecto

Abordar y compensar los valores faltantes es crucial. Es esencial comprender los supuestos en los que se basan los datos existentes y garantizar la claridad de cómo se utilizan los conjuntos de datos en los procesos de toma de decisiones. Mantener la calidad de los datos es primordial, especialmente cuando los datos están incompletos o son ambiguos. (Rochel y Evéquoz, 2021, pp. 614-617)

Si bien esta fase requiere mucho tiempo, también es fundamental. La fiabilidad y el rendimiento de la solución de IA resultante están directamente relacionados con la calidad y la precisión de los datos subyacentes. (Saltz, 2023)

Principios éticos en la etapa de recopilación, comprensión y preparación de datos

La fase de recopilación de datos conlleva riesgos inherentes, en particular en lo que respecta al principio de no maleficencia. Garantizar la privacidad y confidencialidad de los datos es fundamental durante esta fase. Los sistemas de IA deben mantener la privacidad de los datos de forma coherente, desde el inicio hasta todo el ciclo de vida de la solución. Las regulaciones legales, como la Ley Global de Protección de Datos, suelen regular los asuntos de privacidad (de Silva y Alahakoon, 2022, p. 5).

Los sesgos dentro de los datos pueden conducir a la sobrerrepresentación de ciertos grupos demográficos, por lo que es esencial examinar y garantizar la calidad e integridad de los datos. Además, los conjuntos de datos están expuestos a amenazas cibernéticas, lo que requiere medidas de seguridad sólidas para protegerlos de posibles ataques. (UNESCO, 2023a, págs. 9 y 18).

Ser transparente sobre los procesos de recopilación de datos, incluidos los datos que se recopilan, con qué fines y cómo se utilizarán en el sistema de IA, también es válido en este punto.. (De Silva y Alahakoon, 2022, p. 6).

Partes interesadas en la etapa de recopilación, comprensión y preparación de datos

Involucrar a los científicos de datos durante esta fase es crucial, ya que desempeñan un papel fundamental en el discernimiento de patrones y tendencias dentro de los datos, transformándolos en conocimiento accionable (Miller, 2022). Del mismo modo, los ingenieros de IA son indispensables para evaluar y verificar la integridad de los datos (Rochel y Evéquoz, 2021), mientras que los analistas de datos mejoran el proyecto al comprender las características de los datos y asesorar sobre estrategias efectivas de preparación de datos. Además, la contribución de los ingenieros de datos es vital para la preparación y el mantenimiento de los datos, asegurando su calidad (Miller, 2022).

La inclusión de un delegado de protección de datos (DPD) es importante para garantizar el cumplimiento de las normas legales y éticas relativas a los datos personales (Supervisor Europeo de Protección de Datos, 2024).

Los especialistas en ética desempeñan un papel importante al integrar consideraciones éticas en todo el proceso de procesamiento y selección de datos. La participación de estas diversas partes interesadas equipa el proyecto de IA con datos de origen ético y meticulosamente preparados, estableciendo una base de confianza y fiabilidad para la solución de IA.

13. Ciclo de vida del desarrollo de la IA; Desarrollo y formación de modelos

Durante la fase de desarrollo y entrenamiento del modelo, la creación de un modelo de IA está específicamente dirigida a abordar el desafío preidentificado. Esta fase incluye la cuidadosa selección, configuración y uso previsto de herramientas algorítmicas, por lo que es esencial documentar y evaluar minuciosamente las implicaciones de estas elecciones, en particular identificando cualquier inconveniente potencial o compromiso necesario. La naturaleza iterativa de esta etapa es una característica definitoria, que implica múltiples rondas de refinamiento para mejorar la precisión y la eficacia del modelo.

El proceso de selección de herramientas algorítmicas a menudo implica navegar entre objetivos contradictorios. Un buen ejemplo es el desafío de filtrar los correos electrónicos no deseados, donde existe la necesidad de equilibrar el objetivo de minimizar el spam en las bandejas de entrada de los usuarios con el riesgo de categorizar incorrectamente los correos electrónicos legítimos como spam. Lograr el equilibrio óptimo es fundamental para mejorar la eficacia del modelo sin comprometer su fiabilidad o la confianza del usuario. (Rochel y Evéquoz, 2021, pp. 617-618; Saltz, 2023)

Principios éticos en la etapa de desarrollo y formación de modelos

Si bien los detalles técnicos a menudo tienen prioridad, esta etapa ofrece una buena oportunidad para garantizar la explicabilidad e interpretabilidad del modelo elegido, así como para evaluar si existen sesgos. (Prem, 2023, p. 703) El proceso de desarrollo debe ser claro y comprensible para todas las partes involucradas, lo que requiere una documentación exhaustiva de las fuentes de datos, las opciones de modelos y la lógica que informó estas decisiones. (Morley et al., 2020, p. 2151)

Cuando es necesario, surgen compensaciones, en particular porque la eficiencia a menudo implica compromisos, es esencial un enfoque sistemático para reconocer y evaluar abiertamente sus efectos. Los ingenieros de IA deben estar preparados para racionalizar sus elecciones, obligándolos a rendir cuentas y mitigar cualquier consecuencia adversa para los afectados. (Morley et al., 2020, p. 2151)

Actores en la etapa de desarrollo y formación del modelo

Los científicos de IA o aprendizaje automático desempeñan un papel fundamental en la transformación de las definiciones de problemas en prototipos iniciales de modelos de IA (De Silva y Alahakoon, 2022), lo que marca su participación crítica a medida que el proyecto avanza en la fase de desarrollo y entrenamiento del modelo. Durante esta fase, los ingenieros de IA se vuelven indispensables debido a su capacidad para identificar y emplear las herramientas algorítmicas más adecuadas que estén específicamente alineadas con los objetivos del proyecto (Rochel y Evéquoz, 2021). Además, los científicos de datos están en el centro de esta fase, dirigiendo el desarrollo y el refinamiento continuo de los modelos de aprendizaje automático hacia una mayor precisión y eficiencia.

La inclusión de expertos en ética es fundamental para garantizar la incorporación de consideraciones éticas en el proceso de desarrollo, con el objetivo de evitar cualquier posible impacto adverso. Además, los conocimientos proporcionados por los usuarios empresariales son inestimables, ya que garantizan que los modelos en evolución estén en concordancia con las estrategias empresariales y cumplan con los resultados esperados, cerrando así la brecha entre la innovación técnica y las aplicaciones empresariales prácticas.

14. Ciclo de vida del desarrollo de la IA: evaluación y pruebas del modelo

Una vez que se ha desarrollado y entrenado el modelo de IA, su eficacia debe probarse y evaluarse en función de objetivos y criterios de éxito predefinidos. Esta evaluación no solo mide el rendimiento del modelo, sino que también examina los supuestos subyacentes y los criterios de selección de los conjuntos de datos utilizados. En los casos en que el modelo no cumple con las expectativas, es posible que se requieran ajustes, ya sea en los parámetros del modelo, su arquitectura general o los conjuntos de datos en los que se basa. La comunicación transparente de cualquier deficiencia con el líder del proyecto y todas las partes interesadas relevantes es crucial, dado el impacto significativo potencial en las personas involucradas. En función de los resultados de la evaluación, el equipo debe decidir las acciones subsiguientes, que podrían implicar más iteraciones para el refinamiento o pasar a la fase de implementación, asegurándose de que cada paso esté claramente alineado y acordado. (Hotz, 2018; Rochel y Evéquoz, 2021, p. 619)

Principios éticos en la etapa de evaluación y prueba del modelo

En esta fase, es primordial probar el cumplimiento del modelo a las directrices éticas, asegurándose de que se alinea con los principios críticos establecidos al principio. El desarrollo y el perfeccionamiento de las auditorías y las métricas de auditabilidad son tareas esenciales. Es importante incorporar principios éticos clave como la justicia, la beneficencia y la no maleficencia. Evaluar la resiliencia del modelo a posibles amenazas de seguridad y establecer una estrategia para desafíos inesperados son pasos cruciales. Debe haber mecanismos claros para la adopción de medidas correctivas en caso de resultados adversos. Es necesaria una evaluación exhaustiva que abarque la privacidad, el impacto social, los posibles sesgos y más, con el compromiso de minimizar e informar de manera transparente cualquier efecto negativo. (De Silva y Alahakoon, 2022, p. 8; Morley et al., 2020, pág. 2151; Prem, 2023, p. 703)

La transparencia es esencial, no solo en lo que respecta al funcionamiento técnico del modelo, sino también en lo que respecta a las justificaciones de las decisiones humanas dentro del proyecto, lo que garantiza la explicabilidad (De Silva y Alahakoon, 2022, p. 8; Morley et al., 2020, p. 2151). Además, es crucial reconocer que los ingenieros podrían enfrentarse a la presión de acelerar el proceso de desarrollo para la implementación, lo que podría introducir más riesgos (Rochel y Evéquoz, 2021, p. 618).

Partes interesadas en la etapa de evaluación y prueba del modelo

En esta fase, los ingenieros de IA desempeñan un papel vital al interpretar los resultados de la modelización, ofreciendo información esencial sobre los resultados (Rochel y Evéquoz, 2021).

Los equipos de aseguramiento de la calidad (QA) son cruciales para confirmar que los modelos se adhieren a los estándares de calidad y están libres de errores o resultados no deseados. Involucrar a expertos en ética es clave para garantizar el despliegue ético de los modelos de IA, asegurando que se utilicen de manera responsable. Los defensores o representantes de los usuarios contribuyen significativamente al alinear los modelos de IA con las expectativas y requisitos de los usuarios, aumentando así la satisfacción y la aceptación de los usuarios.

Además, los equipos legales y de cumplimiento son indispensables para abordar los riesgos legales, asegurándose de que los modelos de IA cumplan con todas las leyes y regulaciones aplicables. (Moore, 2023)

15. Ciclo de vida del desarrollo de la IA: implementación

Después de su desarrollo exitoso, la solución de IA se despliega en un entorno de producción destinado a resolver problemas del mundo real. El alcance y el método de implementación dependen de los objetivos específicos de la solución y de los sistemas o aplicaciones para los que está diseñada. Por lo general, la implementación comienza con una fase piloto en la que participa un grupo selecto de expertos y usuarios pioneros que pueden proporcionar comentarios iniciales (De Silva y Alahakoon, 2022, p. 8).

La solución de IA podría integrarse con los sistemas existentes, servir de base para una nueva aplicación o servicio, o sus hallazgos podrían compartirse a través de métodos fuera de línea, como informes de gestión. Es crucial establecer un mecanismo de retroalimentación para monitorear su desempeño e impacto, facilitando mejoras continuas basadas en el uso en el mundo real (Saltz, 2023).

Para hacer frente a los riesgos potenciales, se implementan estrategias integrales de gestión de riesgos. Se implementan medidas operativas y de supervisión para garantizar que la solución de IA funcione de manera eficiente después de la implementación. Al final se lleva a cabo una revisión exhaustiva del proyecto, que da como resultado un informe detallado que resume los resultados del proyecto y las lecciones aprendidas (De Silva y Alahakoon, 2022, pp. 8-9).

Principios éticos en la etapa de implementación

Morley (2020) destaca que las consideraciones éticas a menudo no reciben suficiente atención durante la fase de implementación del proceso de diseño. Destilar comportamientos humanos complejos en herramientas que sean a la vez comprensibles y transparentes presenta desafíos significativos. Sin embargo, la creación de este tipo de herramientas es crucial para mejorar la confianza en los sistemas de IA. Por lo tanto, la explicabilidad emerge como un principio ético clave en esta etapa.

Además, garantizar la autonomía es esencial para informar a los usuarios finales cuando interactúan con un proceso impulsado por la IA. Esta transparencia es crucial para evitar una dependencia excesiva de los sistemas de IA. La UNESCO (2023b, p. 36) subraya la necesidad de contar con mecanismos eficaces en cuatro ámbitos clave: concienciación del sistema, procedimientos de auditoría sólidos, claridad en el razonamiento algorítmico y medidas para medir el impacto, como canales para apelaciones o quejas, que deben incluir protecciones para los denunciantes. Además, no se puede exagerar el papel fundamental de la supervisión humana, con la necesidad de disposiciones que permitan la intervención y supervisión humanas de los sistemas de IA para garantizar su despliegue ético (Morley et al., 2020, p. 2151).

Partes interesadas en la etapa de implementación

Los ingenieros de IA o aprendizaje automático son fundamentales en la transición del prototipo de modelo de IA a un servicio o solución completamente implementado, accesible para las partes interesadas y los usuarios finales (De Silva y Alahakoon, 2022).

Los expertos en ciberseguridad realizan una evaluación de riesgos para proteger el sistema de IA (Junklewitz et al., 2023, pp. 9-12). Los usuarios finales contribuyen significativamente al resaltar cualquier problema o deficiencia que podría no haber sido evidente durante las pruebas en entorno controlado. Además, los defensores o representantes de los usuarios son fundamentales para recopilar comentarios sobre la usabilidad y el rendimiento de la solución de IA implementada. Su función garantiza que la solución se alinee con las expectativas y los requisitos del usuario, facilitando las mejoras cuando sea necesario.

16. Ciclo de vida del desarrollo de la IA: funcionamiento y supervisión

El mantenimiento, las actualizaciones y las evaluaciones continuas son esenciales para la longevidad y la eficacia de un sistema de IA operativo. Se llevan a cabo revisiones constantes del rendimiento para garantizar que el sistema cumpla con los estándares esperados, mientras que el monitoreo continuo de las interacciones de los usuarios ofrece información valiosa sobre la aplicación en el mundo real y las áreas potenciales de mejora. Las actualizaciones periódicas con nuevos datos son fundamentales para mantener el modelo relevante y preciso. La retroalimentación de los usuarios es fundamental para el refinamiento continuo del modelo de IA, lo que destaca la importancia de las mejoras iterativas como parte del ciclo de mantenimiento, simbolizando el progreso en lugar del fracaso.

Además, medir el retorno de la inversión (ROI) del sistema y otras métricas críticas es vital para evaluar su éxito en el logro de objetivos predefinidos. (De Silva y Alahakoon, 2022, p. 9; Saltz, 2023)

Principios éticos en la etapa de operación y monitoreo

En esta fase, el monitoreo del desempeño se extiende más allá de las métricas técnicas para incluir el cumplimiento de los estándares éticos. La reevaluación periódica de todos los principios es fundamental, especialmente teniendo en cuenta los posibles cambios en los conjuntos de datos y la estructura del modelo. La participación de las partes interesadas es clave incluso después de la implementación, y sus aportes son esenciales para las mejoras continuas. La retroalimentación debe buscarse activamente de los usuarios, las organizaciones y la sociedad en general. Deben existir mecanismos estructurados para recopilar e incorporar sistemáticamente esta retroalimentación en los esfuerzos de mejora continua. (Grupo de Expertos de Alto Nivel sobre IA (AI HLEG), 2019, p. 19; UNESCO, 2023b, p. 15)

Actores en la etapa de operación y monitoreo

Durante la fase de operación y monitoreo, los equipos de DevOps y operaciones de TI son cruciales para monitorear el rendimiento del sistema y mantener la disponibilidad continua (Immersant Data Solutions, 2023). Los evaluadores independientes, como los certificadores de seguridad, pueden evaluar la seguridad de la solución de IA desarrollada (Miller, 2022). Además, los paneles de expertos, los comités directivos o los organismos reguladores realizan revisiones del proyecto, que abarcan consideraciones técnicas y éticas (De Silva y Alahakoon, 2022).

17. Conclusión

Esta unidad ha profundizado en los principios, marcos y directrices éticos que son fundamentales para dar forma al desarrollo de soluciones de IA. Exploró el ciclo de vida del desarrollo de la IA, desde el planteamiento inicial del problema hasta el funcionamiento y la supervisión de la solución de IA, centrándose en la integración de los principios relacionados y la participación de las partes interesadas en cada etapa.

El desarrollo ético de la IA pone de manifiesto la importancia de incorporar consideraciones éticas a lo largo de todo el ciclo de vida del desarrollo de la IA. Este enfoque garantiza que las tecnologías de IA no solo contribuyan positivamente a la sociedad, sino que también aborden eficazmente los riesgos potenciales, como los sesgos y los problemas de privacidad. Además, se hizo hincapié en la importancia de involucrar a un amplio espectro de partes interesadas, lo que ilustra que el desarrollo exitoso de la IA requiere aportaciones más allá de los conocimientos técnicos para incluir diversas perspectivas. Este enfoque inclusivo enriquece el proceso de desarrollo, asegurando que las soluciones de IA estén alineadas con los valores sociales y logren una mayor aceptación.

Además, se subrayó la necesidad de un seguimiento continuo y un perfeccionamiento iterativo de los sistemas de IA después de su implantación como esenciales para mantener su integridad ética y garantizar su éxito sostenido. Al adoptar una postura proactiva hacia la innovación, los desarrolladores pueden identificar y abordar los desafíos éticos a medida que surjan, navegando así por las complejidades del desarrollo de la IA con un compromiso con los estándares éticos y la innovación responsable.

18. Referencias

Ayling, J., & Chapman, A. (2022). Poner en práctica la ética de la IA: ¿Son las herramientas adecuadas para su propósito? IA y Ética, 2(3), 405–429. https://doi.org/10.1007/s43681-021-00084-x

Academia de Inteligencia Artificial de Pekín: Principios de IA de Pekín. (2019). Protección de datos y seguridad de datos, 10.

Principios de Inteligencia Artificial de Beijing. (2022, 10 de enero). Centro Internacional de Investigación para la Ética y la Gobernanza de la IA. https://ai-ethics-and-governance.institute/beijing-artificial-intelligence-principles/

De Silva, D., & Alahakoon, D. (2022). Un ciclo de vida de inteligencia artificial: desde la concepción hasta la producción. Patrones, 3(6), 100489. https://doi.org/10.1016/j.patter.2022.100489

Comisión Europea. (2023, 12 de septiembre). La Comisión acoge con satisfacción el acuerdo político sobre la Ley de IA [Texto]. Comisión Europea - Comisión

Europea. https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6473
Supervisor Europeo de Protección de Datos. (9 de abril de 2024). Delegado de Protección de Datos (DPO) | Supervisor Europeo de Protección de Datos. https://www.edps.europa.eu/data-protection/data-protection/reference-library/data-protection-officer-dpo_en
Parlamento Europeo. (2023, 8 de junio). Ley de IA de la UE: primer reglamento sobre inteligencia artificial. Temas | Parlamento Europeo. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence
Floridi, L., & Cowls, J. (2019). Un marco unificado de cinco principios para la IA en la sociedad. Revista de Ciencia de Datos de Harvard. https://doi.org/10.1162/99608f92.8cd550d1
Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People: un marco ético para una buena sociedad de IA: oportunidades, riesgos, principios y recomendaciones. Mentes y máquinas, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
Instituto del Futuro de la Vida. (2024). El explorador de la Ley de IA | Ley de Inteligencia Artificial de la UE. https://artificialintelligenceact.eu/ai-act-explorer/
Grupo de Expertos de Alto Nivel sobre IA (AI HLEG). (2019). Directrices éticas de la IA confiable. Grupo de Expertos de Alto Nivel sobre Artificial Inteligencia. Comisión Europea. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
Grupo de Expertos de Alto Nivel sobre IA (AI HLEG). (2020). Lista de evaluación de la inteligencia artificial confiable (ALTAI) para la autoevaluación. Comisión Europea.
Hillard, A., & Gulley, A. (2023, 9 de diciembre). ¿Qué es la Ley de Inteligencia Artificial de la UE? https://www.holisticai.com/blog/eu-ai-act
Hotz, N. (2018, 10 de septiembre). ¿Qué es CRISP DM? Alianza de Procesos de Ciencia de Datos. https://www.datascience-pm.com/crisp-dm-2/
Soluciones de datos inmersivas. (9 de junio de 2023). DevOps: Cerrando la brecha entre el desarrollo y las operaciones | LinkedIn. https://www.linkedin.com/pulse/devops-bridging-gap-between-development-operations/
Fundación de Diseño de Interacción - IxDF. (27 de octubre de 2021). ¿Qué es la Co-Creación? Fundación de Diseño de Interacción—IxDF. La Fundación del Diseño de Interacción. https://www.interaction-design.org/literature/topics/co-creation
Jobin, A., Ienca, M., & Vayena, E. (2019). El panorama global de las directrices éticas de la IA. Inteligencia Artificial de la Naturaleza, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2
Junklewitz, H., Hamon, R., André, A., Evas, T., Soler Garrido, J., & Sánchez Martín, J. I. (2023). Ciberseguridad de la inteligencia artificial en la Ley de IA: Principios rectores para abordar los requisitos de ciberseguridad para los sistemas de

IA de alto riesgo. Oficina de Publicaciones de la Unión Europea. https://data.europa.eu/doi/10.2760/271009
Leijnen, S., Aldewereld, H., van Belkom, R., Bijvank, R., & Ossewaarde, R. (2020). Un marco ágil para una IA fiable. 75–78.
Miller, G. J. (2022). Roles de los stakeholders en proyectos de inteligencia artificial. Liderazgo de proyectos y sociedad, 3, 100068. https://doi.org/10.1016/j.plas.2022.100068
Moore, S. (2023, 18 de diciembre). Qué es... un equipo de Legal y Cumplimiento Corporativo? | LinkedIn. https://www.linkedin.com/pulse/what-isa-corporate-legal-compliance-team-steven-moore-iczqf/
Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). Del qué al cómo: una revisión inicial de las herramientas, métodos e investigaciones éticas de la IA disponibles públicamente para traducir los principios en prácticas. Ética de la Ciencia y la Ingeniería, 26(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5
OCDE. (2019). Descripción general de los principios de IA. Resumen de los principios de IA de la OCDE. https://oecd.ai/en/principles
Pedro, A. (2024, 2 de octubre). Arquitectura de software de IA: Navegando por los aspectos esenciales del desarrollo de aplicaciones empresariales | LinkedIn. https://www.linkedin.com/pulse/ai-software-architecture-navigating-essentials-business-alexx-peter-uugnc/
Prem, E. (2023). De los marcos éticos de la IA a las herramientas: una revisión de los enfoques. IA y Ética, 3(3), 699–716. https://doi.org/10.1007/s43681-023-00258-9
Rochel, J., & Evéquoz, F. (2021). Entrar en la sala de máquinas: Un plan para investigar los pasos oscuros de la ética de la IA. IA Y SOCIEDAD, 36(2), 609–622. https://doi.org/10.1007/s00146-020-01069-w
Rodrigues, R. (2020). Cuestiones jurídicas y de derechos humanos de la IA: brechas, desafíos y vulnerabilidades. Revista de Tecnología Responsable, 4, 100005. https://doi.org/10.1016/j.jrt.2020.100005
Saltz, J. (1 de junio de 2023). ¿Qué es el ciclo de vida de la IA? Alianza de Procesos de Ciencia de Datos. https://www.datascience-pm.com/ai-lifecycle/
Sanín, A. (4 de septiembre de 2020). Tutoriales de co-creación. Diseño Globant. https://medium.com/design-globant/co-creation-how-tos-e25696f56d6f
UNESCO. (2023a). Evaluación de impacto ético. Una herramienta de la Recomendación sobre la Ética de la Inteligencia Artificial. UNESCO. https://doi.org/10.54678/YTSA7796
UNESCO. (2023b). Metodología de evaluación de la preparación. Una herramienta de la Recomendación sobre la Ética de la Inteligencia Artificial. UNESCO. https://doi.org/10.54678/YHAA4429
UNESCO. (2023c, 21 de abril). Inteligencia Artificial: Ejemplos de dilemas éticos | UNESCO. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics/cases
Blanco, C. (2023, 30 de noviembre). ¿Qué hace realmente un diseñador de UX? [Guía 2024]. https://careerfoundry.com/en/blog/ux-design/what-does-a-ux-designer-actually-do/

Zhou, J., & Chen, F. (2022). Ética de la IA: de los principios a la práctica. IA Y SOCIEDAD. https://doi.org/10.1007/s00146-022-01602-z

Zhou, J., Chen, F., Berry, A., Reed, M., Zhang, S., & Savage, S. (2020). Una encuesta sobre los principios éticos de la IA y sus implementaciones. Serie de Simposios IEEE 2020 sobre Inteligencia Computacional (SSCI), 3010–3017. https://doi.org/10.1109/SSCI47803.2020.9308437

Zwikael, O., & Meredith, J. R. (2018). ¿Quién es quién en el zoológico del proyecto? Los diez roles principales del proyecto. Revista Internacional de Operaciones y Gestión de la Producción, 38(2), 474–492. https://doi.org/10.1108/IJOPM-05-2017-0274

CU2 | Privacidad y conveniencia de la IA

Índice de contenidos

1. Introducción 41

2. El dilema de Sara: la disyuntiva entre privacidad y conveniencia 43

3. Comprender la privacidad en la IA 46

3.1. Privacidad de datos: Protección de datos personales contra accesos no autorizados 47

3.2. Seguridad de la información: medidas para proteger la integridad y confidencialidad de los datos 48

3.3. Autonomía personal: el derecho a controlar la información sobre uno mismo 48

4. Privacidad de datos 49

4.1. Importancia de la privacidad de los datos en la IA 49

4.2. Papel de las regulaciones como el GDPR 49

4.3. Aspectos clave de la privacidad de datos 50

5. Comprender la conveniencia en la IA 51

5.1. Ejemplos de conveniencia de la vida real 53

6. La interacción de la privacidad y la conveniencia 57

7. Importancia de la privacidad y la conveniencia 59

8. Navegar por los riesgos y beneficios 60

9. Estudio de caso: La IA en la vigilancia 62

10. Conclusión 65

1. Introducción

En la era digital, el rápido avance de la Inteligencia Artificial (IA) ha transformado innumerables aspectos de la vida cotidiana, remodelando la forma en que interactuamos con la tecnología y entre nosotros. En el corazón de esta transformación se encuentra la interacción dinámica entre dos factores críticos: la privacidad y la conveniencia. Esta UC profundiza en la compleja relación entre estos elementos, explorando los desafíos y oportunidades que presentan en la era de las tecnologías inteligentes.

A medida que integramos cada vez más la IA en nuestra vida personal y profesional, se hace imperativo examinar críticamente el equilibrio entre mejorar la conveniencia del usuario y salvaguardar la privacidad personal. Este CU está diseñado para proporcionar a los participantes una comprensión integral de cómo la privacidad y la conveniencia coexisten armoniosamente dentro de las aplicaciones de IA. Destaca tanto su potencial para mejorar la vida cotidiana como las consideraciones éticas que conllevan.

Los participantes obtendrán información sobre los principios fundamentales que rigen la privacidad de los datos, los dilemas éticos que plantea la IA y los marcos legales que dan forma al futuro de la implementación de la tecnología. Al explorar aplicaciones del mundo real y escenarios hipotéticos, este curso tiene como objetivo equipar a los alumnos con las herramientas para navegar por el panorama cambiante de la IA de manera responsable, asegurando que los avances tecnológicos mejoren las experiencias humanas sin comprometer los derechos individuales.

Este viaje a través de los reinos de la IA, la privacidad y la conveniencia desafiará a los participantes a pensar críticamente sobre el papel de la tecnología en la sociedad y su responsabilidad como futuros líderes en IA.

FUENTE DE LA IMAGEN | Generado por DALL-E

CONSEJOS Y RECOMENDACIONES PARA LOS

Una pregunta atractiva o un dato sorprendente para despertar el interés de los alumnos por explicar el panorama actual de la IA, haciendo hincapié en la importancia de la privacidad y la conveniencia.

Dato sorprendente

Un estudio reciente descubrió que los sistemas de IA pueden tomar decisiones sobre sus solicitudes de empleo, puntajes de crédito e incluso su atención médica, a menudo sin su participación o conocimiento directo. Esta integración de la IA plantea preguntas cruciales: ¿Cuánto control tenemos sobre nuestros datos y qué significa esto para nuestra privacidad? Profundicemos en el complejo mundo de la IA, donde la conveniencia puede venir a costa de nuestra privacidad.

Pregunta atractiva

¿Sabías que la persona promedio interactúa con la IA más de 20 veces al día sin darse cuenta? Desde recomendaciones de compra personalizadas hasta dispositivos domésticos inteligentes que gestionan nuestras rutinas diarias, la IA se integra a la perfección en nuestras vidas. Pero ¿con qué frecuencia considera qué información personal está intercambiando para esta conveniencia? Exploremos lo que realmente está en juego.

Además, el contenido se puede presentar como una encuesta.

Pregunta de la encuesta

"¿Qué servicios utilizas que conoces para recopilar tus datos con el fin de mejorar la funcionalidad y personalizar tu experiencia?"

Opciones

- Teléfonos inteligentes
- Plataformas de redes sociales
- Servicios de streaming
- Sitios web de comercio electrónico
- Dispositivos domésticos inteligentes
- Rastreadores de actividad física y aplicaciones de salud
- Asistentes de voz
- Aplicaciones de navegación y viajes
- Aplicaciones bancarias y financieras
- Servicios de correo electrónico

Pregunta de seguimiento

"¿Cómo de cómodo/a se siente con estos servicios que recopilan sus datos?"

Mucho || Algo || Neutro || Poco|| Muy poco

Este enfoque ayuda a identificar el conocimiento de la recopilación de datos en diferentes servicios y captura los sentimientos sobre la privacidad y el uso de datos personales, proporcionando una visión integral de las actitudes de los participantes hacia la privacidad de los datos.

2. El dilema de Sara: la disyuntiva entre privacidad y conveniencia

FUENTE DE LA IMAGEN | Generado por DALL-E

Imagínese a Sara, una estudiante universitaria que acaba de mudarse a su primer apartamento. Ansiosa por optimizar su ajetreada vida, Sara invierte en varios dispositivos inteligentes: un altavoz inteligente para administrar su horario y reproducir música, un termostato inteligente para controlar la temperatura de su apartamento y un reloj inteligente para monitorear sus rutinas de ejercicios.

CONSEJOS Y RECOMENDACIONES PARA PROFESORES

Consejo: haz preguntas en los puntos críticos para que te identifiques

FUENTE DE LAS IMÁGENES | Generado por DALL-E

Una mañana, el altavoz inteligente de Sara sugiere una nueva ruta a su campus universitario, lo que le permite ahorrar tiempo durante un período de mucho tráfico, un ejemplo perfecto de conveniencia. Más tarde, recibe un anuncio personalizado en su teléfono para una cafetería a lo largo de su nueva ruta, tentándola con su café con leche favorito. Inicialmente complacida, Sara se pregunta:

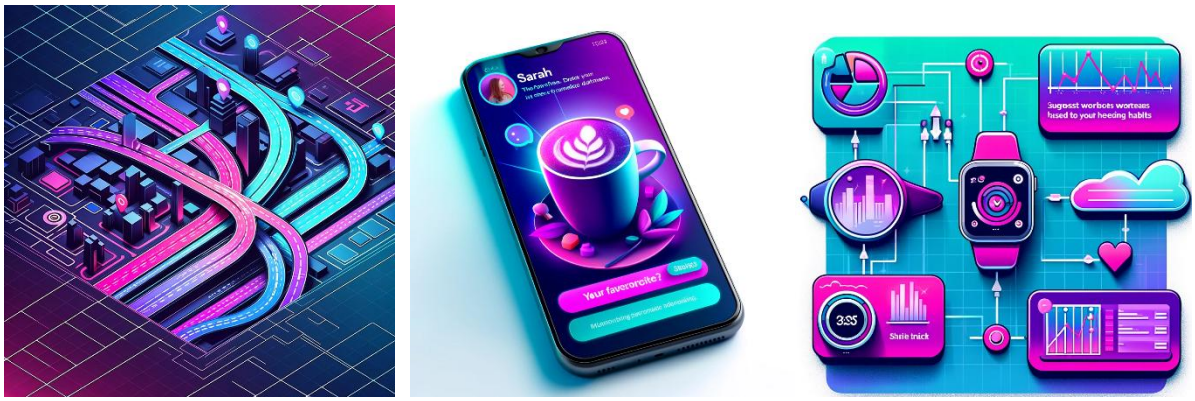
—¿Cómo lo supo?

Pregunta

"¿Crees que Sara consideró las implicaciones de privacidad antes de comprar estos dispositivos? ¿Por qué sí o por qué no?

Pregunta

- ¿Qué opinas de que los dispositivos tomen decisiones basadas en tus rutinas diarias? ¿Cuáles son las posibles compensaciones de privacidad?
- ¿Qué información se podría haber compartido para personalizar este anuncio? ¿Te sientes cómodo con este nivel de intercambio de datos para anuncios personalizados?



FUENTE DE LAS IMÁGENES | Generado por DALL-E

A medida que pasan los días, Sara observa más de estos sucesos. Su reloj inteligente sugiere planes de ejercicio basados en su nivel de condición física y compras recientes de comida rápida, registrados por su aplicación de pago. Su termostato se ajusta a su horario y parece anticipar cambios inusuales en su rutina, como quedarse en casa cuando está enferma.

Pregunta

"¿En qué momento la personalización se vuelve demasiado invasiva? ¿Dónde trazarías la línea para ti mismo?"

FUENTE DE LA IMAGEN | Generado por DALL-E

Si bien Sara aprecia la conveniencia que ofrecen estos dispositivos impulsados por IA, que hacen su vida más fluida y eficiente, comienza a sentirse incómoda con la cantidad de información personal que recopilan y cómo se usa. No accedió explícitamente a compartir su ubicación ni su historial de compras. El deleite de

Sara con las convenienciaes rápidamente se tiñe de preocupación por su privacidad.

Esta historia nos presenta el delicado equilibrio entre la privacidad y la conveniencia en la IA. A medida que las tecnologías de IA se integran más en nuestra vida cotidiana, ofrecen una conveniencia sin precedentes y plantean importantes problemas de privacidad. El reto consiste en encontrar un equilibrio en el que Sara y usuarios como ella puedan disfrutar de los beneficios de la IA sin sentir que su información personal se ve comprometida.

FUENTE DE LA IMAGEN | Generado por DALL-E

Pregunta:

- ¿Alguna vez la tecnología te ha hecho sentir similar a cómo se siente Sara? ¿Qué acciones tomó, si es que tomó alguna?
- ¿Qué le aconsejarías a Sara que hiciera en esta situación?
- ¿Qué medidas puede tomar para controlar mejor sus datos?

3. Comprender la privacidad en la IA

La privacidad puede entenderse en términos generales como el derecho de las personas a mantener su información personal fuera de la vista del público y a controlar sus datos e interacciones. Este derecho es reconocido como un derecho humano fundamental en muchos ordenamientos jurídicos y tratados internacionales, haciendo hincapié en la importancia de la autonomía y la dignidad personal.

En el ámbito de la IA, la privacidad adquiere una complejidad adicional. Los sistemas de IA a menudo se basan en grandes cantidades de datos para aprender y tomar decisiones. Estos datos pueden incluir información personal confidencial que, si se maneja incorrectamente, puede conducir a violaciones significativas de la privacidad. La integración de la IA en las tecnologías cotidianas, desde los asistentes personales y los dispositivos portátiles hasta el análisis predictivo en la atención sanitaria, significa que la protección de la privacidad requiere medidas tradicionales de protección de datos y nuevos enfoques para abordar los desafíos únicos que plantean las tecnologías de IA.

"La privacidad es un derecho humano fundamental reconocido en la Declaración Universal de Derechos Humanos de las Naciones Unidas, el Pacto Internacional de Derechos Civiles y Políticos y muchos otros tratados mundiales y regionales. En el contexto de la IA, la privacidad abarca varias dimensiones clave".

La importancia de la privacidad en el contexto de la IA es multifacética:

- **Confianza**: Las protecciones de privacidad efectivas son cruciales para mantener la confianza de los usuarios en las tecnologías de IA. Sin confianza, los usuarios pueden ser reacios a adoptar innovaciones potencialmente beneficiosas.
- **Obligaciones éticas**: Existe un imperativo ético de respetar y proteger la privacidad individual. Esto implica garantizar que los sistemas de IA no infrinjan los derechos personales y estén diseñados teniendo en cuenta la privacidad desde cero.
- **Cumplimiento**: Muchas jurisdicciones tienen regulaciones estrictas de protección de datos, como el GDPR en Europa, que impone obligaciones específicas a las organizaciones que procesan datos personales a través de sistemas de IA.

3.1. Privacidad de datos: Protección de datos personales contra accesos no autorizados

La privacidad de los datos en la IA se refiere a la protección de los datos personales contra el acceso, el uso o la divulgación no autorizados. Los sistemas de IA deben diseñarse para garantizar que los datos y la información susceptible se manejen de forma segura y solo las partes autorizadas accedan a ellos. Las medidas eficaces de privacidad de datos ayudan a prevenir el robo de identidad, el fraude financiero y los riesgos de explotación de datos personales. Entre ellos se encuentran:

- **Cifrado de datos**: Cifrado de datos tanto en reposo como en tránsito para evitar el acceso no autorizado.
- **Controles de acceso**: Limitar el acceso a los datos a las personas que los necesitan para realizar sus funciones laborales.
- **Anonimización de datos**: eliminar la información de identificación personal de los conjuntos de datos utilizados en la IA para reducir el riesgo de violaciones de la privacidad.

Por ejemplo, una empresa minorista en línea recopila datos de clientes para el procesamiento de pedidos y el marketing personalizado. La empresa utiliza el cifrado para proteger los datos de pago y la información personal de los clientes para garantizar la privacidad de los datos. Además, implementan estrictas políticas internas que limitan el acceso de los empleados a estos datos confidenciales solo a aquellos que los necesitan para procesar pedidos. También aseguran el consentimiento del cliente a través de procedimientos claros de inclusión voluntaria antes de enviar comunicaciones de marketing.

3.2. Seguridad de la información: medidas para proteger la integridad y confidencialidad de los datos

La seguridad de la información está estrechamente relacionada con la privacidad de los datos, pero se centra más en las medidas de protección que garantizan que los datos sigan siendo precisos, fiables y accesibles solo para los usuarios autorizados. En el contexto de la IA:

Integridad: Proteger los datos de cambios no autorizados que podrían comprometer su precisión.

Confidencialidad: Garantizar que la información personal se mantenga en secreto y solo sea accesible a ella las personas que estén autorizadas a acceder a ella.

Protocolos de seguridad: Implementación de protocolos robustos como capas de sockets seguros (SSL), firewalls y sistemas de detección de intrusos para protegerse contra ataques externos.

Por ejemplo, un proveedor de atención médica utiliza un sistema de registros médicos electrónicos (EHR) para almacenar datos del paciente. Para proteger estos datos, el proveedor emplea múltiples capas de medidas de seguridad. Esto incluye el uso de un almacenamiento de datos sólido y el cifrado de transferencias, la autenticación multifactor (MFA) para el acceso al sistema y auditorías de seguridad periódicas para identificar y mitigar las vulnerabilidades. Estas prácticas ayudan a garantizar la integridad y confidencialidad de los datos de los pacientes frente a las amenazas cibernéticas.

3.3. Autonomía personal: el derecho a controlar la información sobre uno mismo

La autonomía personal en la era digital, especialmente dentro de la IA, consiste en mantener el control sobre la información personal y las decisiones que se toman en función de esa información. Los sistemas de IA deben diseñarse para respetar y mejorar la autonomía personal mediante:

- **Consentimiento informado**: Garantizar que los usuarios estén completamente informados sobre cómo se utilizarán sus datos y obtener su consentimiento antes de recopilar datos.
- **Transparencia**: Proporcionar a los usuarios información clara sobre las prácticas de procesamiento de datos y la lógica detrás de las decisiones de IA.

- **Mecanismos de control**: Proporcionar a los usuarios mecanismos para controlar cómo se utiliza su información y para optar por no participar en el procesamiento de datos cuando lo deseen.

Por ejemplo, una aplicación de seguimiento de la actividad física del usuario recopila datos de actividad física para proporcionar consejos personalizados sobre la actividad física. La aplicación respeta la autonomía personal al informar claramente a los usuarios sobre los tipos de datos que recopila y cómo se utilizarán. También permite a los usuarios acceder a sus datos, corregir cualquier inexactitud y optar por no participar en la recopilación de datos para ciertas funciones. Este enfoque garantiza que los usuarios tengan control sobre su información personal y sus decisiones.

4. Privacidad de datos

La privacidad de los datos se refiere a la protección de los datos personales contra el acceso, el uso o la divulgación no autorizados. Abarca garantizar que las personas tengan control sobre su información y que se utilice de acuerdo con sus preferencias y regulaciones legales.

4.1. Importancia de la privacidad de los datos en la IA

La privacidad de los datos es primordial en la IA, donde se procesan grandes cantidades de datos personales. Sin la protección adecuada, la información confidencial de las personas puede ser vulnerable a violaciones y usos indebidos, lo que genera daños potenciales, como robo de identidad, pérdidas financieras o discriminación. Además, los sistemas de IA dependen en gran medida de los datos para tomar decisiones informadas, lo que significa que la calidad y la integridad de los datos tienen un impacto directo en el rendimiento y la fiabilidad de estos sistemas.

4.2. Papel de las regulaciones como el GDPR

Regulaciones como el Reglamento General de Protección de Datos (GDPR) son cruciales para hacer cumplir los principios de privacidad de datos en la IA. El GDPR establece requisitos estrictos para la recopilación, el procesamiento y el almacenamiento de datos personales, incluida la obtención del consentimiento explícito de las personas, la implementación de medidas de seguridad adecuadas y la provisión de mecanismos para que los interesados accedan y controlen sus datos. Al hacer cumplir estas regulaciones, el GDPR tiene como objetivo

salvaguardar los derechos de privacidad de las personas y responsabilizar a las organizaciones por el manejo de datos personales.

Por ejemplo, escenario de aplicación de salud: considere una aplicación de salud que recopile y almacene datos médicos de los usuarios, como su historial médico, diagnósticos y planes de tratamiento. Esta aplicación plantea riesgos significativos para la privacidad de los datos de los usuarios sin el consentimiento o las medidas de seguridad adecuados. Por ejemplo, si la aplicación carece de un cifrado sólido o controles de acceso, podría ser vulnerable a violaciones de datos, exponiendo información médica confidencial a partes no autorizadas. Además, si la aplicación comparte los datos del usuario con terceros sin el consentimiento explícito o el anonimato, podría violar las regulaciones de privacidad como el GDPR, lo que podría tener consecuencias legales. Este escenario ilustra la importancia de implementar medidas estrictas de privacidad de datos en las aplicaciones de IA para proteger la información confidencial de las personas y garantizar el cumplimiento de los requisitos normativos.

4.3. Aspectos clave de la privacidad de datos

La privacidad de los datos es fundamental para las prácticas digitales modernas, especialmente con la proliferación de tecnologías y servicios basados en datos. Estos son algunos elementos clave de la privacidad de los datos que ayudan a salvaguardar la información personal y garantizar el manejo ético de los datos:

- **Consentimiento:** El consentimiento implica obtener el permiso de las personas antes de recopilar, usar o compartir sus datos. Debe darse libremente, ser específica, informada y sin ambigüedades. Esto significa que las personas deben ser plenamente conscientes de lo que están consintiendo y comprender claramente el alcance y las implicaciones de las actividades de procesamiento de datos.
- **Limitación de la finalidad:** Este principio dicta que los datos deben recopilarse para fines específicos, explícitos y legítimos y no procesarse posteriormente de manera incompatible. La limitación de la finalidad ayuda a garantizar que los datos se utilicen solo de la manera que el interesado espera y consiente, evitando que los datos se utilicen de forma maliciosa o irresponsable.
- **Minimización de datos:** La minimización de datos significa que solo se recopilan y procesan los datos necesarios para el procesamiento. Este principio alienta a las organizaciones a limitar la recopilación de datos a lo que es directamente relevante y necesario para lograr un propósito específico. Esto reduce el riesgo de daños por violaciones de datos, ya que menos datos pueden verse comprometidos.
- **Rendición de cuentas:** La rendición de cuentas se refiere a la obligación de las organizaciones de asumir la responsabilidad de los datos que procesan y demostrar el cumplimiento de todos los principios de

protección de datos. Esto incluye mantener registros de las actividades de procesamiento de datos, implementar medidas de protección de datos y mostrar cómo se logra el cumplimiento. Este principio garantiza que las organizaciones sean transparentes sobre sus prácticas de tratamiento de datos y puedan ser consideradas responsables de sus actividades de tratamiento de datos.

- **Transparencia:** La transparencia exige que cualquier información y comunicación relacionada con el tratamiento de datos personales sea fácilmente accesible y comprensible para las personas. Las políticas de privacidad y los avisos de recopilación de datos deben utilizar un lenguaje claro y sencillo. La transparencia genera confianza, mejora la equidad y permite a los usuarios tomar decisiones informadas sobre sus datos.
- **Acceso y corrección:** Las personas tienen derecho a acceder a sus datos y a que se corrijan los datos incorrectos que les conciernen. Proporcionar a los sujetos la capacidad de acceder y actualizar sus datos según sea necesario garantiza la precisión de los datos y permite a las personas mantener el control sobre su información, lo cual es fundamental para la autonomía personal.
- **Seguridad:** La seguridad implica la implementación de medidas técnicas y organizativas adecuadas para proteger los datos personales contra la destrucción, pérdida, alteración, divulgación o acceso no autorizados accidentales o ilegales a los mismos. Esto incluye prácticas como el cifrado, el almacenamiento seguro de datos y el acceso físico y digital controlado.

Juntos, estos aspectos de la privacidad de los datos forman un marco sólido que las organizaciones pueden utilizar para proteger los datos personales y respetar los derechos de privacidad de las personas. Ayudan a construir una cultura consciente de la privacidad que se alinea con los estándares legales y las expectativas éticas en la era digital.

5. Comprender la conveniencia en la IA

En el contexto de la IA, la conveniencia se refiere a la facilidad y eficiencia con la que se pueden completar las tareas o mejorar las experiencias a través de las tecnologías de IA. La IA agiliza los procesos, reduce el esfuerzo manual y proporciona experiencias personalizadas adaptadas a las preferencias individuales.

La conveniencia, facilitada por las tecnologías de IA, abarca la integración perfecta de la automatización, la personalización y la eficiencia en varios dominios. Desde la atención médica y las finanzas hasta la vida cotidiana, las soluciones impulsadas por IA optimizan los procesos, potencian la toma de decisiones y enriquecen las experiencias, lo que en última instancia mejora la productividad y el bienestar general.

La conveniencia, en el ámbito de la IA, encarna la integración perfecta de la tecnología para agilizar las tareas y aumentar la eficiencia. Las tecnologías de IA aprovechan los algoritmos y el aprendizaje automático para automatizar procesos, reducir el esfuerzo manual y adaptar las experiencias a las preferencias individuales. Los sistemas de IA optimizan los flujos de trabajo mediante el análisis de grandes cantidades de datos y el aprendizaje de patrones, lo que ahorra tiempo y recursos y ofrece resultados mejorados. A continuación, se presentan algunos puntos en los que la IA contribuye a la conveniencia en varias áreas:

- **Automatización de tareas rutinarias: La** IA sobresale en la automatización de tareas repetitivas y mundanas que tradicionalmente requieren intervención humana. Por ejemplo, el software impulsado por IA puede encargarse de la programación de citas, la respuesta a las consultas de los clientes o la gestión automática de correos electrónicos. Esto acelera el proceso y libera recursos humanos para tareas más complejas y creativas, aumentando así la productividad y la eficiencia.
- **Personalización: los** sistemas de IA pueden analizar grandes cantidades de datos para adaptar los servicios y productos a las preferencias y necesidades individuales. En contextos de consumo, esto podría manifestarse como una IA que recomienda productos en función de compras anteriores o hábitos de visualización, como se ve en el comercio minorista en línea y los servicios de transmisión. En aplicaciones más personales, la IA puede ajustar los sistemas de calefacción del hogar en función del horario y las preferencias del propietario, lo que garantiza la conveniencia y optimiza el uso de energía.
- **Mejora la toma de decisiones:** La IA mejora la toma de decisiones al proporcionar a las personas y organizaciones información derivada del análisis de grandes conjuntos de datos que serían demasiado complejos para que los humanos los procesaran manualmente. En sectores como la atención médica, los algoritmos de IA ayudan a diagnosticar enfermedades mediante el análisis de datos de imágenes médicas más rápido y, a menudo, con mayor precisión que los profesionales humanos. En finanzas, los sistemas de IA pueden analizar datos de mercado para proporcionar información sobre inversiones o detectar transacciones fraudulentas con gran precisión y velocidad.
- **Eficiencia y gestión de recursos:** La IA mejora significativamente la eficiencia general y optimiza la gestión de recursos. Por ejemplo, los sistemas de gestión de la cadena de suministro y logística impulsados por IA pueden predecir rápidamente la demanda, optimizar las rutas de entrega y gestionar el inventario, reduciendo los residuos y los costes. Del mismo modo, las redes inteligentes impulsadas por IA pueden gestionar la distribución de electricidad en una ciudad de forma más eficiente, adaptándose a los cambios en la demanda en tiempo real.
- **Accesibilidad y facilidad de uso: Las** tecnologías de IA a menudo hacen que la tecnología sea más accesible y fácil para una gama más amplia de personas, incluidas aquellas con discapacidades. Los dispositivos asistidos

por voz impulsados por IA ayudan a las personas con discapacidades visuales o físicas a interactuar con la tecnología y administrar su entorno de manera más efectiva. La IA también puede derribar las barreras lingüísticas con servicios de traducción en tiempo real, haciendo que la información y la comunicación sean accesibles para los hablantes no nativos.

5.1. Ejemplos de conveniencia de la vida real

Cuidado de la salud: En la atención médica, las tecnologías de IA ofrecen conveniencia al automatizar los procesos de diagnóstico, analizar imágenes médicas y detectar patrones en los datos de los pacientes para ayudar a los profesionales de la salud a diagnosticar enfermedades de manera más rápida y precisa. Esto ahorra tiempo a los proveedores de atención médica y a los pacientes, lo que conduce a una prestación de atención médica más eficiente y mejores resultados para los pacientes.

FUENTE DE LA IMAGEN | Generado por DALL-E

Conveniencia en la atención médica a través de la IA

- **Procesos de diagnóstico automatizados**: La IA acelera el análisis de datos médicos, como imágenes y pruebas de diagnóstico, lo que permite diagnósticos más rápidos.
- **Precisión diagnóstica mejorada**: los algoritmos de IA detectan patrones sutiles en los datos que los humanos podrían pasar por alto, lo que aumenta la precisión y la fiabilidad de los diagnósticos.
- **Análisis predictivo**: La IA utiliza datos de varias fuentes, incluida la tecnología portátil, para predecir posibles problemas de salud antes de que se agraven, lo que facilita la atención médica preventiva.
- **Gestión optimizada de** pacientes: La IA optimiza los flujos de trabajo de los hospitales mediante la gestión de los horarios, la admisión de pacientes y el enrutamiento a los entornos de atención adecuados, lo que reduce los tiempos de espera y mejora las experiencias de los pacientes.
- **Planes de tratamiento personalizados**: Los algoritmos analizan los datos individuales de los pacientes para adaptar los planes de tratamiento que tienen más probabilidades de ser eficaces en función del perfil de salud único del paciente.

Finanzas: Las herramientas financieras impulsadas por IA brindan conveniencia al ofrecer asesoramiento de inversión personalizado basado en objetivos financieros individuales, tolerancia al riesgo y tendencias del mercado. Estas herramientas utilizan algoritmos para analizar grandes cantidades de datos financieros y proporcionar recomendaciones para estrategias de inversión, lo que ayuda a las personas a tomar decisiones informadas y optimizar sus carteras financieras.

FUENTE DE LA IMAGEN | Generado por DALL-E

Conveniencia en las finanzas a través de la IA

- **Asesoramiento de inversión personalizado**: Las herramientas de IA analizan los objetivos financieros, la tolerancia al riesgo y las situaciones económicas personales de una persona para proporcionar recomendaciones de inversión personalizadas.
- **Análisis avanzado de datos**: Estos sistemas utilizan algoritmos para examinar grandes volúmenes de datos financieros, incluidas las tendencias del mercado y el rendimiento histórico, para identificar las mejores oportunidades de inversión.
- **Gestión optimizada de la cartera**: La IA ayuda a ajustar dinámicamente las carteras de inversión en función de los cambios del mercado en tiempo real y los objetivos financieros individuales, mejorando los rendimientos potenciales al tiempo que gestiona la exposición al riesgo.
- **Información eficiente del mercado**: Las herramientas de IA proporcionan información rápida sobre las condiciones del mercado, lo que ayuda a los inversores a comprender la compleja dinámica del mercado y a tomar decisiones oportunas.
- **Evaluación de riesgos**: Los algoritmos evalúan el riesgo asociado con varias opciones de inversión, lo que permite a los usuarios tomar decisiones informadas basadas en la tolerancia al riesgo.
- **Automatización de transacciones**: La IA puede automatizar las transacciones comerciales y de inversión, agilizando el proceso de ejecución y reduciendo la probabilidad de error humano.
- **Seguridad mejorada**: La IA emplea medidas de seguridad avanzadas, como la detección de anomalías, para identificar y mitigar las posibles amenazas a las cuentas de inversión.

Vida cotidiana: Los dispositivos domésticos inteligentes habilitados para IA, como los asistentes de voz, los termostatos inteligentes y los sistemas automatizados de seguridad para el hogar, ofrecen conveniencia al automatizar las tareas domésticas y mejorar la conveniencia y la seguridad. Estos dispositivos aprenden las preferencias del usuario con el tiempo y ajustan la configuración en consecuencia, proporcionando experiencias fluidas y personalizadas que simplifican las rutinas diarias y mejoran la calidad de vida.

FUENTE DE LA IMAGEN | Generado por DALL-E

Conveniencia en la vida cotidiana a través de la IA

- **Automatización de las tareas domésticas**: Los dispositivos habilitados para IA, como las aspiradoras inteligentes y los electrodomésticos de cocina automatizados, se encargan de las tareas rutinarias, liberando el tiempo de los usuarios.
- **Configuraciones personalizadas**: Los dispositivos como los termostatos inteligentes y los sistemas de iluminación aprenden y se adaptan a las preferencias del usuario, ajustando automáticamente la configuración para una conveniencia óptima.
- **Tecnología asistida por voz**: Los asistentes de voz como *Alexa y Google Assistant* permiten el control manos libres sobre los dispositivos domésticos, lo que facilita el acceso a la información y la gestión de tareas.
- **Seguridad mejorada en el hogar**: Los sistemas de seguridad inteligentes utilizan la IA para monitorear los entornos domésticos, reconocer rostros familiares, detectar actividades inusuales y alertar a los propietarios de viviendas sobre posibles amenazas.
- **Eficiencia energética**: Los dispositivos impulsados por IA optimizan el uso de energía en el hogar al aprender las horas pico de uso y ajustar las operaciones para reducir el desperdicio, lo que reduce las facturas de servicios públicos.
- **Integración perfecta**: Los ecosistemas de hogar inteligente conectan varios dispositivos, lo que les permite trabajar juntos sin problemas, mejorando la conveniencia del usuario a través de controles e interacciones integrados.
- **Monitoreo y control remotos**: Los propietarios pueden monitorear y controlar dispositivos domésticos inteligentes de forma remota a través de teléfonos inteligentes, lo que brinda tranquilidad cuando está fuera de casa.

Si bien la IA mejora significativamente la conveniencia, también plantea consideraciones importantes, como las preocupaciones sobre la privacidad, la necesidad de supervisión humana y la posibilidad de desplazamiento laboral en sectores específicos. Garantizar que los sistemas de IA se utilicen de forma ética y responsable es crucial para maximizar sus beneficios y minimizar los posibles daños. En conclusión, la conveniencia en la IA consiste en aprovechar el poder de la inteligencia artificial para hacer la vida más fácil, más agradable y eficiente. A medida que las tecnologías de IA evolucionan y se integran en diversos aspectos de la vida, es probable que su potencial para impulsar la conveniencia se amplíe, lo que promete impactos transformadores en todas las facetas de la sociedad.

CONSEJOS Y RECOMENDACIONES PARA LOS

Presente escenarios hipotéticos que desafíen a los participantes a decidir entre priorizar la privacidad o la conveniencia.

Ejemplo

Escenario hipotético: Aplicación de seguimiento de salud

Antecedentes: Está considerando una nueva aplicación de seguimiento de la salud, *HealthMate.* Esta aplicación ofrece planes personalizados de acondicionamiento físico y dieta mediante el análisis de sus actividades diarias, ingesta de alimentos y métricas de salud.

Desafío del escenario: *HealthMate* puede sincronizarse con sus redes sociales para recopilar más datos sobre el estilo de vida para mejorar la personalización y compartir información con socios que podrían enviarle anuncios dirigidos en función de sus objetivos de salud.

Preguntas para la reflexión:

- ¿Optaría por la personalización mejorada, sabiendo que sus datos podrían usarse para publicidad dirigida?
- ¿Cómo sopesa los beneficios de un plan de salud personalizado frente a los riesgos de compartir sus datos?

Este escenario incita a los participantes a evaluar el equilibrio entre la conveniencia de las recomendaciones de salud personalizadas y los riesgos de privacidad del intercambio de datos.

6. La interacción de la privacidad y la conveniencia

La interacción dinámica entre la privacidad y la conveniencia en la tecnología moderna presenta oportunidades y desafíos. A medida que la tecnología se integra cada vez más en nuestra vida diaria, comprender cómo equilibrar estos aspectos se vuelve crucial. A continuación, se presentan las implicaciones de las compensaciones, los dilemas éticos y las consecuencias en el mundo real de equilibrar la privacidad con la conveniencia.

Compensaciones entre privacidad y conveniencia:

Equilibrar la privacidad y la conveniencia requiere una navegación cuidadosa de las compensaciones inherentes involucradas:

- Restricciones en la recopilación de datos: La implementación de protecciones de privacidad a menudo significa limitar la cantidad y el tipo de datos recopilados y utilizados. Esto puede restringir la funcionalidad de los sistemas de IA que dependen de grandes conjuntos de datos para mejorar la prestación de servicios y la experiencia del usuario.
- Sacrificar la privacidad por la conveniencia: Por otro lado, mejorar la conveniencia a menudo implica recopilar y analizar una gran cantidad de datos personales. Esto puede incluir el seguimiento de los comportamientos, preferencias y ubicaciones de los usuarios, que podrían invadir la privacidad personal.

Dilemas éticos que surgen de estas disyuntivas:

El equilibrio entre privacidad y conveniencia pone en primer plano varios dilemas éticos:

- Compromiso de la privacidad: ¿Cuánta privacidad están dispuestas a renunciar las personas para mejorar la conveniencia? La respuesta a menudo varía según las percepciones individuales y el contexto de uso de la tecnología.
- Prácticas de recopilación de datos: Las organizaciones también son éticamente responsables de recopilar, usar y compartir información personal. Las consideraciones éticas deben guiar las decisiones sobre qué datos se recopilan, cómo se utilizan y cuánto se informa a los usuarios sobre estas prácticas.

Implicaciones en el mundo real de elegir la conveniencia sobre la privacidad o viceversa:

Las decisiones tomadas con respecto al equilibrio entre privacidad y conveniencia tienen efectos tangibles:

- Consecuencias de priorizar la conveniencia: Elegir la conveniencia a menudo aumenta riesgos como violaciones de datos, robo de identidad e invasiones significativas de la privacidad. Esto puede provocar daños a largo plazo en la confianza de los usuarios y posibles repercusiones legales para las empresas.
- Efectos de priorizar la privacidad: Por el contrario, priorizar en gran medida la privacidad podría limitar la efectividad de soluciones tecnológicas específicas. Esto podría conducir a servicios menos personalizados y potencialmente ralentizar la innovación y el avance tecnológico que podría beneficiar a la sociedad.

Navegar por el equilibrio entre la privacidad y la conveniencia requiere un enfoque matizado que tenga en cuenta las implicaciones éticas, sociales y personales de la tecnología. Las organizaciones y las personas deben esforzarse por encontrar un punto medio que respete la privacidad del usuario y, al mismo tiempo, aproveche los beneficios que puede aportar la tecnología. Este equilibrio no es estático y necesita una evaluación continua a medida que la tecnología evoluciona y los valores sociales cambian. Comprender estas dinámicas ayuda a las partes interesadas a tomar decisiones informadas que protejan los derechos individuales y, al mismo tiempo, adopten los avances de la era digital.

7. Importancia de la privacidad y la conveniencia

La privacidad y la conveniencia en la IA son fundamentales para dar forma a la confianza de los usuarios y la adopción de la tecnología. A medida que la IA se integra en nuestra vida diaria, es crucial lograr un equilibrio entre la protección de los datos personales y la mejora de las experiencias de los usuarios. La privacidad garantiza la protección de la información personal, mientras que la conveniencia mejora la eficiencia y la personalización en el uso de la tecnología. El reto consiste en armonizar estos aspectos para fomentar el despliegue ético de la IA y su aceptación por parte de la sociedad, garantizando que los avances en materia de IA aporten beneficios sin comprometer los derechos individuales.

Mejorar la confianza de los usuarios y la adopción de tecnología

Dar prioridad a la privacidad y la conveniencia en las tecnologías de IA es esencial para generar confianza en los usuarios y fomentar una adopción más amplia. Cuando los sistemas de IA gestionan la privacidad de forma transparente y proporcionan experiencias cómodas y sin problemas, los usuarios las aceptan más fácilmente. Así es como priorizar estos aspectos puede influir en la confianza de los usuarios y la adopción de la tecnología:

- **Generar confianza**: Los usuarios tienden a confiar en la tecnología que respeta su privacidad y mejora su vida diaria sin intrusiones significativas. Las soluciones de IA que protegen eficazmente los datos de los usuarios al tiempo que ofrecen experiencias personalizadas y eficientes tienden a construir una reputación positiva entre los usuarios.

- **Fomentar la adopción**: Cuando los usuarios perciben que una tecnología de IA mejora su productividad o conveniencia sin comprometer su privacidad, es más probable que la adopten y la recomienden. Demostrar el valor de la IA a través de la conveniencia, junto con sólidas salvaguardas de privacidad, motiva a los usuarios a integrar estas tecnologías en sus rutinas diarias.

Consideraciones legales y éticas en la implementación de la IA

La implementación de tecnologías de IA requiere una cuidadosa consideración de las implicaciones legales y éticas para garantizar el uso responsable y el cumplimiento de las regulaciones:

- **Cumplimiento normativo**: Leyes como el Reglamento General de Protección de Datos (RGPD) proporcionan directrices estrictas sobre la privacidad de los datos que deben seguir las tecnologías de IA que operan en la Unión Europea o se dirigen a usuarios de esta Unión. El cumplimiento de estas regulaciones no se trata solo de una necesidad legal, sino también de demostrar un compromiso con la privacidad del usuario.

- **Marcos éticos**: Más allá de los requisitos legales, los marcos éticos guían el desarrollo de la IA para garantizar que las tecnologías se utilicen en beneficio de la sociedad. Estos marcos ayudan a los desarrolladores y a las empresas a navegar por problemas complejos como el sesgo, la equidad, la transparencia y la responsabilidad en los sistemas de IA.

Impacto en las normas sociales y los comportamientos individuales

El papel de la IA en el equilibrio de la privacidad y la conveniencia tiene efectos significativos en las normas sociales y los comportamientos individuales:

- **Dar forma a las expectativas y preferencias de privacidad**: A medida que las tecnologías de IA se integran más en la vida cotidiana, influyen en la forma en que las personas perciben y valoran su privacidad. Esto puede llevar a cambios en las expectativas, ya que algunos usuarios se vuelven más conscientes de la privacidad, mientras que otros pueden volverse insensibles a compartir información personal.

- **Influir en las rutinas diarias y la toma de decisiones**: Las tecnologías de IA que ofrecen conveniencia pueden alterar significativamente la vida diaria. Los dispositivos domésticos inteligentes, las experiencias de compra personalizadas y la gestión automatizada de las finanzas personales pueden remodelar las actividades cotidianas, facilitando la vida y creando dependencias de la tecnología para la toma de decisiones.

Al comprender y abordar la compleja interacción entre la privacidad y la conveniencia, los desarrolladores e implementadores de IA pueden mejorar la aceptación de la tecnología y garantizar que contribuya positivamente al desarrollo de la sociedad y al bienestar individual. Estas consideraciones son cruciales para el crecimiento sostenible de las tecnologías de IA en una sociedad cada vez más consciente de la privacidad y los estándares éticos.

8. Navegar por los riesgos y beneficios

Los riesgos potenciales asociados con las tecnologías de IA en materia de privacidad y conveniencia incluyen:

- **Violaciones de datos:** los sistemas de IA a menudo manejan grandes cantidades de datos confidenciales, lo que los convierte en objetivos de ataques maliciosos. Las violaciones de datos pueden resultar en el acceso no autorizado a la información personal, lo que provoca robo de identidad, pérdidas financieras y daños a la reputación.

- **Pérdida de privacidad:** las tecnologías de IA pueden recopilar y analizar datos personales sin el consentimiento o el conocimiento de las personas, lo que compromete los derechos de privacidad. Esto puede erosionar la

confianza entre los usuarios y los sistemas de IA, lo que genera preocupaciones sobre el uso indebido y la explotación de datos.

- **Sesgos en los algoritmos de IA:** Los algoritmos de IA pueden perpetuar los sesgos en los datos de entrenamiento, lo que conduce a resultados discriminatorios. Los algoritmos de IA sesgados pueden dar lugar a un trato o una toma de decisiones injustos, exacerbando las desigualdades sociales y socavando la confianza en los sistemas de IA.

Consecuencias de estos riesgos:

Estos riesgos pueden socavar la confianza en los sistemas de IA y tener consecuencias negativas para las personas y la sociedad:

- **Erosión de la confianza:** Las filtraciones de datos y las violaciones de la privacidad pueden erosionar la confianza entre los usuarios y los sistemas de IA, lo que provoca una disminución de la adopción y el uso. Las personas pueden volverse reacias a compartir sus datos o interactuar con tecnologías de IA, lo que obstaculiza sus beneficios potenciales.
- **Impactos sociales negativos:** Los algoritmos de IA sesgados pueden perpetuar la discriminación y la desigualdad, lo que provoca un trato injusto y disturbios sociales. Esto puede dañar la cohesión social y la confianza en las instituciones, exacerbando las divisiones sociales.
- **Repercusiones legales y regulatorias:** Las violaciones de datos y las violaciones de la privacidad pueden tener repercusiones legales y regulatorias para las organizaciones responsables de los sistemas de IA. Las multas, las demandas y el daño a la reputación pueden tener importantes implicaciones financieras y operativas.

Beneficios potenciales:

A pesar de los riesgos, las tecnologías de IA ofrecen numerosos beneficios potenciales:

- **Mayor eficiencia:** La IA agiliza los procesos y automatiza las tareas, aumentando la eficiencia y la productividad. Al manejar tareas repetitivas, la IA libera tiempo para que los empleados se concentren en actividades de mayor valor.
- **Mejores decisiones basadas en datos:** Los algoritmos de IA analizan grandes cantidades de datos para descubrir información y patrones, lo que permite a las organizaciones tomar decisiones mejor informadas. La información impulsada por IA puede optimizar las operaciones, mejorar las experiencias de los clientes e impulsar la innovación.
- **Experiencias de usuario mejoradas:** La personalización de la IA mejora las experiencias de los usuarios al ofrecer recomendaciones y servicios personalizados. Desde recomendaciones personalizadas de productos hasta mantenimiento predictivo en la fabricación, la IA mejora la satisfacción y el compromiso de los usuarios.

Mitigación de riesgos:

Las organizaciones deben implementar medidas de seguridad sólidas para mitigar estos riesgos, garantizar la transparencia y la responsabilidad en los sistemas de IA y priorizar las consideraciones éticas a lo largo del ciclo de vida del desarrollo de la IA. Al abordar de forma proactiva estos riesgos, las organizaciones pueden generar confianza en los sistemas de IA y minimizar el daño potencial a las personas y a la sociedad.

Para navegar eficazmente por los riesgos y beneficios de las tecnologías de IA, las organizaciones pueden adoptar las siguientes estrategias:

- **Implementar medidas de seguridad sólidas:** Las organizaciones deben implementar medidas de seguridad sólidas para protegerse contra las violaciones de datos y el acceso no autorizado. Esto incluye cifrado, controles de acceso y auditorías de seguridad periódicas.
- **Garantizar la transparencia y la rendición de cuentas:** La transparencia y la rendición de cuentas son esenciales para generar confianza en los sistemas de IA. Las organizaciones deben ser transparentes sobre cómo se utilizan las tecnologías de IA y garantizar la rendición de cuentas de sus acciones.
- **Priorizar las consideraciones éticas:** Las consideraciones éticas deben guiar el desarrollo y la implementación de la IA. Las organizaciones deben priorizar la equidad, la transparencia y la rendición de cuentas en los sistemas de IA para mitigar los sesgos y garantizar el uso ético de los datos.

Al adoptar una gestión de riesgos proactiva y prácticas responsables de implementación de IA, las organizaciones pueden maximizar los beneficios de las tecnologías de IA y, al mismo tiempo, minimizar el daño potencial a las personas y la sociedad.

9. Estudio de caso: La IA en la vigilancia

Escenario

Las autoridades locales implementan sistemas de vigilancia impulsados por IA en una ciudad bulliciosa para mejorar la seguridad pública. Estos sistemas utilizan tecnologías avanzadas como el reconocimiento facial, el análisis del comportamiento y la detección de objetos para monitorear espacios públicos, identificar amenazas potenciales y responder a emergencias en tiempo real.

FUENTE DE LA IMAGEN | Generado por DALL-E

Mejoras de seguridad frente a violaciones de privacidad:

Mejoras de seguridad	Violaciones de la privacidad
Los sistemas de vigilancia de IA refuerzan las medidas de seguridad al permitir que los organismos encargados de hacer cumplir la ley detecten y respondan rápidamente a actividades sospechosas o posibles amenazas de seguridad. La tecnología de reconocimiento facial ayuda a identificar a las personas en las listas de vigilancia o a las	A pesar de los beneficios de seguridad, surgen preocupaciones con respecto a las posibles violaciones de la privacidad asociadas con la vigilancia de IA. El monitoreo continuo y la recopilación de datos en espacios públicos plantean preguntas sobre el derecho de las personas a la privacidad y a no ser objeto de vigilancia constante. Existe el riesgo de identificación errónea y falsos positivos, lo que conduce a acusaciones erróneas e invasiones de la privacidad de personas inocentes marcadas

Importancia de la implementación reflexiva de la IA:

El caso de la IA en la vigilancia pone de manifiesto la importancia de una implementación reflexiva de la IA para equilibrar las necesidades de seguridad con las consideraciones de privacidad:

- **Políticas y regulaciones claras:** La implementación reflexiva de la IA implica establecer políticas y regulaciones claras que rijan las tecnologías de vigilancia para garantizar la transparencia, la responsabilidad y el cumplimiento de las leyes de privacidad.
- **Uso ético de los datos:** Las organizaciones deben priorizar el uso ético de los datos recopilados a través de sistemas de vigilancia, respetar los derechos de las personas a la privacidad y minimizar el riesgo de violaciones de la privacidad o uso indebido de la información personal.
- **Transparencia algorítmica y rendición de cuentas:** La implementación de algoritmos y mecanismos transparentes de IA para la rendición de cuentas garantiza que los sistemas de vigilancia sean justos, precisos y responsables de sus acciones, mitigando el riesgo de sesgos o errores que podrían conducir a violaciones de la privacidad.

- **Participación pública y supervisión:** Involucrar al público en los debates sobre la vigilancia de la IA e involucrar a las partes interesadas en la toma de decisiones promueve la transparencia, la confianza y la rendición de cuentas, fomentando el despliegue y el uso responsables de las tecnologías de vigilancia.

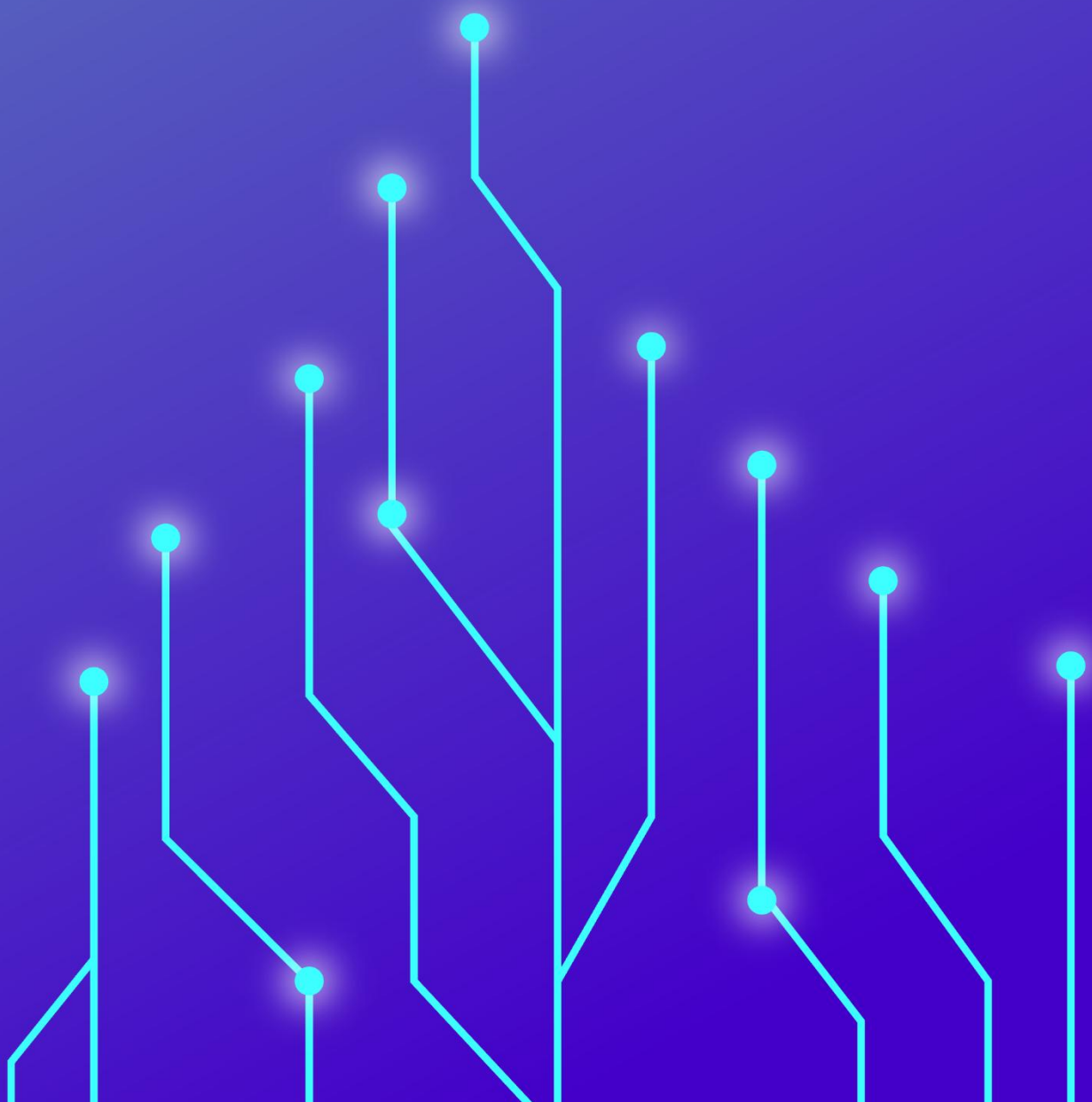
En conclusión, el caso de la IA en la vigilancia subraya la necesidad de una implementación reflexiva de la IA que equilibre las mejoras de seguridad con las consideraciones de privacidad. Las organizaciones pueden implementar sistemas de vigilancia que mejoren la seguridad al tiempo que priorizan los principios éticos, la transparencia y la responsabilidad, respetando los derechos de privacidad de las personas y fomentando la confianza pública.

10. Conclusión

Explorar la privacidad y la conveniencia dentro de la inteligencia artificial (IA) destaca un equilibrio crítico necesario para fomentar la confianza y la adopción más amplia de la tecnología. A medida que la IA impregna diversos aspectos de la atención médica, las finanzas y la vida diaria, aporta enormes beneficios como una mayor eficiencia y servicios personalizados, acompañados de la necesidad de sólidas protecciones de privacidad y consideraciones éticas. Adherirse a marcos legales como el GDPR y abordar los dilemas éticos es esencial para garantizar una implementación responsable de la IA.

A medida que la IA da forma a las normas sociales y a los comportamientos individuales, es crucial mantener un equilibrio entre la conveniencia del usuario y la privacidad de los datos. Los desarrolladores, los usuarios y los responsables políticos deben colaborar para garantizar que las tecnologías de IA mejoren la vida de las personas sin comprometer los derechos fundamentales. Hacer hincapié en la transparencia y la rendición de cuentas será clave para aprovechar el potencial de la IA de forma responsable, garantizando que siga siendo una tecnología beneficiosa y fiable en nuestro mundo cada vez más digital.

CU3 | Los algoritmos y sus limitaciones



Índice de contenidos

1. Introducción 69

2. Comprender los conceptos fundamentales de los algoritmos, sus modelos y limitaciones 73

2.1. Complejidad algorítmica 73

2.2. Caracterización del algoritmo 74

2.3. Estructura del algoritmo 75

2.4. Eficiencia del algoritmo 78

2.5. Limitaciones 80

2.6. Limitaciones de los algoritmos de aprendizaje automático 82

3. Reconocer el papel de los algoritmos en diversos sectores y los posibles escollos que se derivan de su uso 85

3.1. Papel de los algoritmos en diversos sectores 86

3.2. Posibles inconvenientes del uso de algoritmos 96

4. Comprender la complejidad algorítmica, la optimización y las limitaciones de la toma de decisiones algorítmicas 100

4.1. Complejidad algorítmica 100

4..2. Complejidad algorítmica del tiempo 101

4.3. Complejidad algorítmica de la memoria 103

4.4. Optimización 104

4.5. Limitaciones de la toma de decisiones algorítmicas 106

4.6. ¿Cómo identificar las posibles fuentes de sesgo y errores de los sistemas algorítmicos? Una cuestión crucial para garantizar la equidad, la transparencia y la rendición de cuentas 107

4.7. Abordar y mitigar los riesgos y sesgos algorítmicos 108

5. Referencias 111

1. Introducción[1]

Este manual está diseñado para ser una herramienta clave para los instructores que imparten el curso de capacitación del Proyecto Charlie. No solo describirá los contenidos básicos del curso, sino que también proporcionará instrucciones detalladas, consejos y recomendaciones a los capacitadores sobre cómo impartir el curso de manera efectiva. Además, el manual propondrá una variedad de actividades destinadas a involucrar a los alumnos y mejorar su experiencia de aprendizaje. También incluye directrices sobre cómo evaluar los resultados de aprendizaje de los alumnos, garantizando que se cumplan los objetivos del curso y que tanto los formadores como los alumnos tengan una comprensión clara de las competencias que se espera que se desarrollen a lo largo de la formación. Este enfoque integral ayudará a los formadores a facilitar un entorno de aprendizaje dinámico y eficaz[2].

Esta unidad presentará los conceptos fundamentales de los algoritmos, sus modelos y limitaciones. Los participantes comprenderán el papel que desempeñan los algoritmos en diversos sectores y los posibles escollos que se derivan de su uso. Los temas tratados incluirán la complejidad algorítmica, la optimización y las limitaciones de la toma de decisiones algorítmicas.

En la era digital actual, los algoritmos desempeñan un papel cada vez más importante en la configuración de varios aspectos de nuestras vidas, desde el contenido que consumimos en línea hasta las decisiones tomadas por los sistemas automatizados. Comprender las complejidades de la complejidad algorítmica, las estrategias de optimización y las limitaciones inherentes a los procesos de toma de decisiones algorítmicas es crucial para cualquiera que navegue por este panorama.

Comprender la **complejidad algorítmica** implica examinar la eficiencia y el rendimiento de los algoritmos a medida que abordan diversas tareas computacionales. Implica comprender los recursos, como el tiempo y la memoria, que los algoritmos necesitan para resolver problemas y cómo estos requisitos se amplían con tamaños de entrada más grandes. Esta comprensión nos permite

[1] Nota: En la preparación de este documento, los autores emplearon modelos de lenguaje amplio (LLM) para ayudar con la revisión del texto, la estructuración del contenido y la identificación de ejemplos pertinentes incluidos en el material. Después de utilizar esta herramienta de IA, los autores revisaron y refinaron meticulosamente el contenido según fuera necesario. Los autores asumen toda la responsabilidad de la integridad y exactitud del contenido en la publicación final.

[2] Para el profesor/formador: Este manual tiene como objetivo guiarlo a través del material de la unidad del curso. Los PPT de soporte contienen información similar a la que cubrió el E-learning, pero en forma más detallada. Los PPT son extensos y no están destinados a ser cubiertos en su totalidad durante las sesiones interactivas. En su lugar, puede preguntar al comienzo de las sesiones interactivas qué temas fueron más difíciles de comprender para los estudiantes y usar solo esas diapositivas de apoyo durante la sesión. Puedes preguntar esto usando herramientas como Mentimeter o Kahoot como ejemplo, agregando los temas del curso del contenido de la unidad de competencia descritos en viñetas al final de la sección de introducción. Además, se pretende que el estudio de caso basado en CU también se maneje durante la sesión interactiva. Cualquier sugerencia adicional en el manual es solo una guía; Siéntase libre de usarlo como mejor le parezca.

evaluar la viabilidad y aplicabilidad del empleo de algoritmos específicos en contextos del mundo real (Cormen et al., 2009).

Las metodologías de optimización constituyen otro aspecto fundamental de la comprensión de los algoritmos. Estos métodos tienen como objetivo refinar la eficiencia y eficacia de los algoritmos ajustando sus parámetros o reestructurando su lógica fundamental. Ya sea que implique reducir el tiempo de cálculo, maximizar la utilización de recursos u optimizar para objetivos particulares, el empleo de estrategias de optimización adecuadas puede mejorar significativamente el rendimiento del algoritmo en varios dominios (Cormen et al., 2009).

Sin embargo, junto con los beneficios potenciales de la toma de decisiones algorítmica, existen limitaciones inherentes que exigen una consideración cuidadosa. Los algoritmos operan dentro de marcos predefinidos y dependen de la calidad de los datos de entrada y de la precisión de las suposiciones subyacentes. Los sesgos en los datos, las suposiciones de modelado defectuosas o los casos extremos inesperados pueden dar lugar a resultados erróneos o consecuencias no deseadas, lo que pone de manifiesto la fragilidad de los sistemas algorítmicos de toma de decisiones.

Profundizar en la comprensión de la complejidad algorítmica, las estrategias de optimización y las limitaciones inherentes a los procesos de toma de decisiones algorítmicas puede fomentar una comprensión más sólida de los matices y desafíos algorítmicos. Equipa a las personas con las habilidades de pensamiento crítico necesarias para evaluar críticamente las soluciones algorítmicas, identificar posibles escollos y abogar por prácticas algorítmicas responsables y éticas en un mundo cada vez más impulsado por algoritmos.

CONSEJOS Y RECOMENDACIONES PARA LOS

Consejos y recomendaciones para que los profesores se involucren con los conceptos iniciales del algoritmo

1. Aclarar los conceptos clave
Antes de profundizar en las discusiones, asegúrese de que todos los participantes tengan una comprensión clara de la terminología básica, como "complejidad algorítmica", "optimización" y "limitaciones de los algoritmos". Esto se puede lograr a través de una breve presentación o un folleto de glosario al comienzo de la sesión.

2. Relacionarse con aplicaciones del mundo real
Conecte conceptos abstractos con aplicaciones del mundo real en diversos sectores, como las finanzas, la atención médica o el comercio electrónico. Esto ayuda a los participantes a ver la relevancia y los impactos de los algoritmos en

escenarios cotidianos, lo que hace que el contenido sea más atractivo y comprensible.

3. Usa ayudas visuales
Los algoritmos pueden ser complejos y abstractos. Utilice diagramas, diagramas de flujo y animaciones para visualizar conceptos como el flujo algorítmico, la complejidad y los procesos de optimización. Esto puede ayudar a los participantes a comprender y recordar más fácilmente la información.

4. Fomenta las preguntas
Cree un ambiente abierto donde los participantes se sientan cómodos haciendo preguntas. Esto se puede facilitar a través de sesiones regulares de preguntas y respuestas después de cubrir cada tema principal, asegurándose de que los participantes comprendan completamente el material antes de continuar.

5. Incorporar estudios de casos
Discuta casos reales en los que la toma de decisiones algorítmica ha tenido éxito o se ha enfrentado a desafíos. El análisis de estos casos puede generar un debate sobre las implicaciones éticas, la eficiencia y las consideraciones prácticas del uso de algoritmos.

6. Promover el pensamiento crítico
Anime a los participantes a pensar críticamente sobre las limitaciones y los posibles sesgos de los modelos algorítmicos. Esto podría ser a través de discusiones grupales o debates sobre cómo mitigar estos problemas en el diseño e implementación de algoritmos.

Actividades de calentamiento recomendadas

1. Juego de roles de algoritmos
Una actividad de juego de roles en la que los participantes simulan un algoritmo en acción ya sea clasificando elementos manualmente, siguiendo un conjunto simple de instrucciones para lograr un objetivo o representando los pasos que podría tomar un algoritmo en los procesos de toma de decisiones. Este ejercicio ayuda a ilustrar el concepto de procesamiento paso a paso en algoritmos.

2. Rompehielos de complejidad algorítmica
Comience con una tarea simple, como organizar los libros por color, y amplíela para organizarlos por género, luego por autor dentro de cada género. Discuta cómo la complejidad aumenta con cada capa agregada, trazando paralelismos con la complejidad algorítmica y cómo los recursos se utilizan más a medida que las tareas se vuelven más complejas.

3. Desafío de optimización
Asigne a los grupos pequeños una tarea fija, como organizar sillas en una habitación para que quepan tantas personas como sea posible, pero con la menor cantidad de filas. Permítales intentarlo, discuta sus estrategias y luego revele cómo las estrategias de optimización podrían mejorar sus soluciones.

4. Debate ético
Organizar un debate sobre las implicaciones éticas de los algoritmos en la toma de decisiones, centrándose en los posibles sesgos y las consecuencias de depender en gran medida de los sistemas automatizados. Esto puede hacer que

los participantes se familiaricen con las complejidades y responsabilidades de trabajar con algoritmos.

5. Juego de romper el algoritmo
Proporcione a los participantes un algoritmo simple o un conjunto de reglas y desafíelos a encontrar los casos extremos o escenarios en los que falla el algoritmo. Esta actividad puede poner de manifiesto las limitaciones y los posibles fallos de la lógica algorítmica.
Estas actividades y estrategias no solo harán que el proceso de aprendizaje sea atractivo, sino que también empoderarán a los participantes con una comprensión y apreciación más profundas de las complejidades involucradas en la toma de decisiones algorítmicas.

2. Comprender los conceptos fundamentales de los algoritmos, sus modelos y limitaciones

2.1. Complejidad algorítmica

Los algoritmos son hoy en día omnipresentes. Guían a las computadoras para procesar información, resolver problemas y tomar decisiones importantes. Pueden resolver problemas, desde organizar datos hasta encontrar las mejores rutas de viaje o sugerir películas. En una era dominada por el *big data* y las redes sociales, saber cómo funcionan los algoritmos es clave para entender sus limitaciones y posibles sesgos. No solo mejoran la forma en que las computadoras realizan tareas; También influyen en nuestras experiencias cotidianas, por lo que son esenciales para la informática moderna.

Un algoritmo es una secuencia de instrucciones inequívocas para resolver un problema, debe ser correcto, siempre da una solución correcta y debe ser finito, debe terminar. Es un procedimiento paso a paso que nos dice qué hacer para lograr un determinado objetivo. Podemos pensar en ello como cocinar: seguimos una receta que nos dice exactamente qué ingredientes usar, qué cantidad de cada uno y qué pasos seguir. De manera similar, un algoritmo guía a una computadora o a una persona a través de una serie de acciones para lograr un resultado específico, ya sea ordenar una lista de números, encontrar la ruta más corta en un mapa o recomendar películas según sus preferencias.

El concepto de algoritmos existe desde hace mucho tiempo, desde las civilizaciones antiguas. Uno de los primeros ejemplos es el algoritmo euclidiano, atribuido al antiguo matemático griego Euclides, que se desarrolló alrededor del año 300 a.C. para encontrar el máximo común divisor de dos números.

A lo largo de la historia, varias culturas y civilizaciones desarrollaron sus propios algoritmos para resolver problemas matemáticos, navegar por terrenos geográficos o realizar tareas de manera eficiente. Estos algoritmos a menudo se transmitían oralmente o a través de textos escritos.

El término "algoritmo" proviene del nombre del matemático persa Muhammad ibn Musa al-Khwarizmi, que vivió durante la Edad de Oro islámica en el siglo IX. Al-Khwarizmi escribió un libro llamado "Al-Kitab al-Mukhtasar fi Hisab al-Jabr wal-Muqabala" (El libro compendioso sobre el cálculo por finalización y equilibrio), que introdujo métodos sistemáticos para resolver ecuaciones matemáticas. La palabra "algoritmo" se deriva de la versión latinizada de su nombre, "Algoritmi".

Desde entonces, los algoritmos han evolucionado y se han vuelto fundamentales en varios campos, especialmente con la llegada de la informática moderna. Hoy en día, los algoritmos se utilizan no solo en matemáticas, sino también en informática, ingeniería, biología, economía y muchas otras disciplinas para resolver problemas complejos y automatizar tareas.

2.2. Caracterización del algoritmo

Las caracterizaciones de algoritmos son intentos de formalizar la palabra algoritmo. El término algoritmo no tiene una definición formal generalmente aceptada. Knuth (1997) definió una lista de cinco propiedades que son ampliamente aceptadas como requisitos para un algoritmo:

- **Finitud:** "Un algoritmo siempre debe terminar después de un número finito de pasos... un número muy finito, un número razonable". La finitud garantiza que los algoritmos terminen después de un número finito de pasos, evitando bucles infinitos que podrían consumir recursos computacionales indefinidamente. Es similar a completar un rompecabezas en el que, finalmente, la pieza final cae en su lugar, señalando la conclusión del algoritmo. Esta característica de finitud proporciona previsibilidad y control sobre los procesos computacionales.
- **Definición:** "Cada paso de un algoritmo debe estar definido con precisión; Las actuaciones a llevar a cabo deben especificarse de forma rigurosa e inequívoca para cada caso". Hace hincapié en la necesidad de adoptar medidas inequívocas que no dejen lugar a la interpretación. Un algoritmo definido describe claramente cada acción que se debe realizar en cada etapa de la resolución de problemas, lo que garantiza que no haya ambigüedad ni confusión sobre lo que se debe hacer. Esta característica es crucial porque permite a cualquier persona, independientemente de su formación o experiencia, comprender e implementar el algoritmo correctamente.
- **Entrada:** "... cantidades que se le dan inicialmente antes de que comience el algoritmo. Estas entradas se toman de conjuntos específicos de objetos". La entrada se refiere a los datos o información sobre los que opera el algoritmo para producir un resultado. Esta entrada puede ser cualquier forma de datos, como números, texto, imágenes o incluso otros algoritmos. La entrada es esencial porque proporciona la materia prima que el algoritmo procesa y manipula para resolver un problema o realizar una tarea.
- **Producción:** "... cantidades que tienen una relación especificada con los insumos". La salida se refiere al resultado o solución producida por el algoritmo después de procesar la entrada. Este resultado también puede tomar varias formas dependiendo de la naturaleza del problema que se está resolviendo, como un solo valor, un conjunto de valores, una decisión, una representación visual o cualquier otra forma de información que represente el resultado deseado.
- **Efectividad:** "... Todas las operaciones que se deben realizar en el algoritmo deben ser lo suficientemente básicas como para que, en principio, puedan ser realizadas exactamente y en un período de tiempo finito por un hombre utilizando papel y lápiz". La efectividad enfatiza la practicidad y la viabilidad de los algoritmos. Al igual que una receta con ingredientes oscuros o inalcanzables se vuelve ineficaz, los algoritmos deben utilizar

recursos accesibles para realizar sus tareas. Al garantizar que los algoritmos sean prácticos y factibles, la eficacia garantiza que las soluciones sean alcanzables dentro de las limitaciones de los recursos y el tiempo disponibles.

Como podemos observar, la descripción anterior de las características de un algoritmo puede ser intuitivamente clara, carece de rigor formal, ya que no está exactamente claro qué significa "definido con precisión", o "especificado rigurosa e inequívocamente", o "suficientemente básico", y así sucesivamente. Consideramos que es suficiente para nuestros propósitos.

2.3. Estructura del algoritmo

Uno de los conceptos principales detrás de los algoritmos es la abstracción. La abstracción en este contexto se refiere a la idea de ocultar detalles complejos y centrarse solo en los aspectos esenciales necesarios para comprender y resolver un problema. Es como alejarse para ver el panorama general sin perderse en los detalles finos. En el diseño de algoritmos, la abstracción ayuda a simplificar el proceso de resolución de problemas al dividirlo en partes más pequeñas y manejables. Esto nos permite abordar cada parte por separado, sin necesidad de comprender cada detalle intrincado de una sola vez.

Los algoritmos suelen implicar varias instrucciones que trabajan juntas, en lugar de un solo paso. Al crear programas complejos, ejecutamos una serie de estas instrucciones en secuencia, las repetimos o tomamos decisiones basadas en condiciones específicas.

Saber cómo funcionan los algoritmos implica comprender las diferentes formas en que se pueden combinar las instrucciones. Estos esquemas proporcionan un enfoque estructurado para organizar y secuenciar las instrucciones en un algoritmo. Ayudan a descomponer problemas complejos en tareas manejables, lo que hace que el algoritmo sea más comprensible y fácil de implementar. Los algoritmos suelen estar compuestos por estructuras simples que se organizan de forma jerárquica. Estas estructuras controlan el flujo de instrucciones dentro del algoritmo, dictando la secuencia en la que se ejecutan. Esto a menudo se conoce como la estructura de control del algoritmo.

Una estructura de control dirige esencialmente el orden de ejecución de las instrucciones en un algoritmo. Determina el camino que toma el proceso de ejecución, en función de las condiciones definidas dentro de la estructura. Las estructuras de control son un aspecto fundamental de cualquier algoritmo, ya que garantizan que las instrucciones se ejecuten en el orden correcto y que se logre el resultado deseado.

Una característica común de todas las estructuras de control es que tienen un único punto de entrada y un único punto de salida. El punto de entrada único

marca el comienzo de la estructura de control, donde el proceso de ejecución entra en la estructura. Por otro lado, el punto de salida único significa el final de la estructura de control, donde el proceso de ejecución abandona la estructura y pasa a la siguiente parte del algoritmo. Esta característica garantiza que el proceso de ejecución siga una ruta definida dentro de la estructura de control, manteniendo la integridad y corrección del algoritmo.

Hay tres esquemas de composición de instrucción:

· **Secuencial:** En este esquema, las instrucciones se ejecutan una tras otra, siguiendo un orden lineal. Está muy cerca de seguir una receta paso a paso. Los algoritmos secuenciales son frecuentes en las tareas cotidianas, como hacer un sándwich o calcular la suma de números. Garantizan que las acciones ocurran en una secuencia específica, sin ninguna ramificación o toma de decisiones.

Ejemplo: Un algoritmo para hacer un pastel.
1. Precalienta el horno.
2. Engrase y enharine un molde para pasteles.
3. En un tazón, mezcle la mantequilla y el azúcar.
4. Batir los huevos, uno a la vez.
5. Agregue el extracto de vainilla.
6. En un recipiente aparte, combine la harina, el polvo de hornear y la sal.
7. Agregue gradualmente los ingredientes secos a la mezcla húmeda, alternando con leche.
8. Vierta la masa en el molde para pasteles preparado.
9. Coloque el molde para pasteles en el horno precalentado.
10. Hornee durante 25-30 minutos, o hasta que al insertar un palillo en el centro, éste salga limpio
11. Retira el bizcocho del horno y déjalo enfriar en la sartén durante 10 minutos.
15. Sirve el bizcocho.

· **Condicional o Alternativa:** Estas instrucciones introducen puntos de decisión. En función de ciertas condiciones, el programa toma diferentes caminos. Se debe pensar en ello como elegir entre múltiples opciones. Las declaraciones condicionales (como "if", "else" y "switch") nos permiten manejar varios escenarios.

Ejemplo: Algoritmo que simula el proceso de decidir si llevar un paraguas o una chaqueta al salir de casa en base a la previsión meteorológica:

1. Consulta el pronóstico del tiempo para ver si llueve.
2. Si el pronóstico predice lluvia, entonces:
3. Lleva un paraguas.
4. De lo contrario, si el pronóstico no predice lluvia y la temperatura está por debajo de los 15 grados centígrados, entonces:
5. Llévate una chaqueta.
6. Sal de casa.

Hay que tener en cuenta que la sangría de las instrucciones 3 y 5 tiene un papel específico en la definición del algoritmo. Estas instrucciones solo se ejecutan si la instrucción anterior se evalúa positivamente.

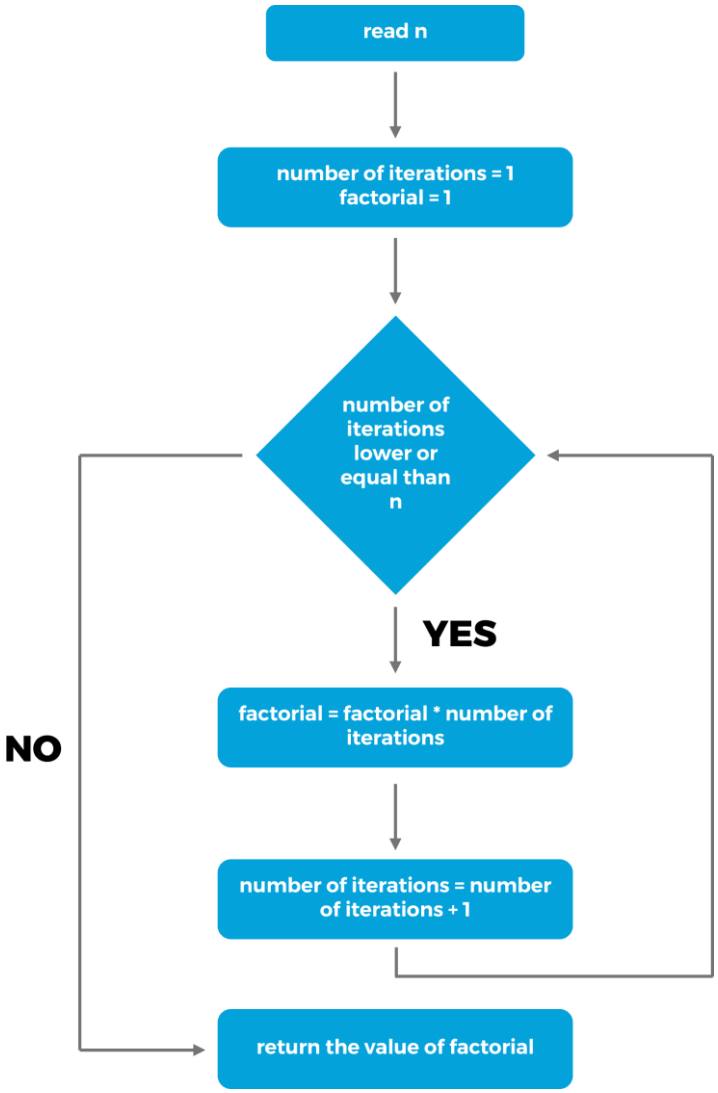
· **Repetitivo o Iterativo:** La repetición es el corazón de estos algoritmos. Ejecutan un conjunto de instrucciones repetidamente hasta que se cumple una condición específica. Los bucles (como "for" y "while") entran en esta categoría. Son esenciales para tareas como el procesamiento de grandes conjuntos de datos, la simulación de eventos o la resolución de problemas de optimización.

Ejemplo: Un programa que calcula el factorial de un número usando un bucle. El factorial de un número es una operación matemática que multiplica un número dado, n, por cada número entero menor que n hasta 1.

1. Lee n, ese es el número para el que queremos calcular el factorial.
2. Establezca el valor factorial en 1.
3. Bucle de 1 a n:
4. Multiplique factorial por el valor del bucle.
5. Mostrar el valor del factorial.

Tenga en cuenta que la sangría de la instrucción número 4 también tiene un papel específico en la definición del algoritmo. Esta instrucción se ejecuta n veces una vez en cada iteración.

Gráfica 1. Ejemplo de bucle



Fuente: | Elaboración propia

Podemos construir algoritmos cada vez más complejos combinando estas tres técnicas. Veamos un ejemplo de un algoritmo que utiliza todos los esquemas de composición para encontrar el valor máximo de un número entero en una lista:

Lista con 10 números enteros. El número 10 es el primer elemento de la lista.

1. El elemento actual es el primer elemento de la lista.
2. El valor máximo es el elemento actual.
3. Si bien no evaluamos todos los elementos de la lista, bucle:
4. Si el artículo actual es mayor que el valor máximo, entonces:
5. El valor máximo es el artículo actual.
6. El elemento actual es el siguiente elemento de la lista.
7. Devuelve el valor máximo.

2.4. Eficiencia del algoritmo

En el campo de la informática, los algoritmos se representan como programas. Un programa puede definirse como la descripción de un algoritmo en el repertorio finito de instrucciones de una máquina. El conjunto de programas que tiene un equipo de cómputo se denomina software. Para conseguir que un ordenador lleve a cabo una tarea, primero debemos pensar y diseñar un algoritmo para resolverla, luego debemos programarlo en el ordenador. El objetivo es transformar el algoritmo conceptual en un conjunto claro de instrucciones en un lenguaje de programación específico, basado en diferentes enfoques del proceso de programación (paradigmas de programación).

A medida que los problemas se vuelven más complejos, también lo hacen los algoritmos que los resuelven. Es en este punto donde se hace necesario poder evaluar el coste que tiene un algoritmo. La eficiencia algorítmica se refiere a la eficiencia con la que un algoritmo utiliza los recursos computacionales. Esta propiedad es esencial para comprender el rendimiento de un algoritmo e implica analizar su consumo de recursos, incluidos el tiempo y el espacio. Lograr la máxima eficiencia implica minimizar el uso de recursos. Sin embargo, comparar la eficiencia de los algoritmos puede ser complejo porque los diferentes recursos no se pueden comparar directamente.

Podemos considerar que un algoritmo es eficiente cuando su consumo de recursos, denominado coste computacional, se mantiene en o por debajo de un umbral aceptable. Este umbral se determina en función de la capacidad del algoritmo para funcionar dentro de un período de tiempo o espacio razonable en una computadora estándar, a menudo en relación con el tamaño de la entrada. En esencia, "aceptable" implica que el algoritmo puede ejecutarse dentro de las limitaciones prácticas, lo que garantiza que siga siendo viable para su uso en el mundo real.

Hay dos medidas principales de la eficiencia algorítmica:

Complejidad temporal: Es la complejidad computacional que describe la cantidad de tiempo que tarda un algoritmo en ejecutarse en función del tamaño de la entrada al programa.

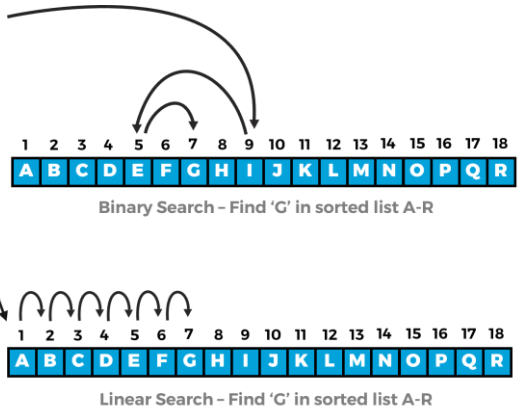
Complejidad del espacio: Esta es la cantidad de memoria que un algoritmo necesita para ejecutarse hasta su finalización.

El objetivo es minimizar los requisitos de tiempo y espacio. Sin embargo, a menudo hay una compensación entre estos dos. Por ejemplo, un algoritmo que se ejecuta más rápido puede requerir más memoria, y un algoritmo que usa menos memoria puede ser más lento. La elección del algoritmo depende de los requisitos específicos de la tarea. Los algoritmos eficientes logran un equilibrio entre estas consideraciones, con el objetivo de obtener un rendimiento óptimo en escenarios del mundo real.

Veamos un ejemplo simple de dos algoritmos diferentes que se pueden usar para buscar un elemento en una lista, pero que tienen la misma complejidad de espacio, pero diferente complejidad de tiempo. A continuación, analizaremos cuándo utilizar cada uno:

La búsqueda lineal es un algoritmo que se utiliza para encontrar un valor dentro de una lista, es muy similar al algoritmo para encontrar el valor máximo de la lista que hemos desarrollado antes. Evalúa cada elemento de la lista, empezando desde el principio hasta que se encuentra el valor objetivo o se alcanza el final de la lista. Durante cada iteración, el algoritmo compara el elemento actual con el valor objetivo. Si se encuentra una coincidencia, el algoritmo finaliza y devuelve el índice del valor de destino.

La búsqueda binaria es un algoritmo eficiente que se utiliza para encontrar un valor objetivo dentro de una lista ordenada dividiendo repetidamente el intervalo de búsqueda por la mitad. Comienza examinando el elemento central de la lista y lo compara con el valor objetivo. Si el elemento central coincide con el objetivo, la búsqueda se realiza correctamente. De lo contrario, si el destino es menor que el elemento central, la búsqueda continúa en la mitad inferior de la lista; Si el objetivo es mayor, la búsqueda se desplaza a la mitad superior. Este proceso de reducción a la mitad del intervalo de búsqueda se repite hasta que se encuentra el valor de destino o el intervalo de búsqueda se vacía. La búsqueda binaria funciona según el principio de divide y vencerás, reduciendo significativamente el espacio de búsqueda con cada iteración.



Fuente: | Tablero Infinity

Con estos dos ejemplos podemos entender fácilmente que existen múltiples algoritmos para realizar una misma tarea. Cada algoritmo tiene sus propias condiciones, fortalezas y debilidades; La búsqueda lineal es fácil de implementar y adecuada para listas pequeñas y datos sin clasificar. Sin embargo, la eficiencia del algoritmo disminuye a medida que crece el tamaño de la lista, lo que lo hace menos útil para grandes conjuntos de datos, ya que, si el elemento que estamos buscando está cerca del final de la lista, debemos evaluar la mayoría de sus elementos. La búsqueda binaria es muy eficiente, por lo que es ideal para buscar conjuntos de datos grandes y ordenados. Sin embargo, requiere que los datos se ordenen inicialmente, lo que puede ser un requisito previo para su aplicación.

2.5. Limitaciones

En los últimos años, hemos observado un aumento notable en las tecnologías de aprendizaje automático e inteligencia artificial, las cuales se apoyan en gran medida en algoritmos para procesar información de conjuntos de datos masivos. Este aumento ha impulsado avances notables en varios dominios, incluido el reconocimiento y la generación de imágenes y videos, la robótica, el análisis médico, la toma de decisiones, la clonación de voz o el procesamiento del lenguaje natural. Estos avances ilustran el potencial ilimitado que ofrecen los algoritmos, impulsándonos a una era marcada por una innovación sin precedentes, pero es importante entender que incluso con estos importantes avances, los algoritmos tienen sus propias limitaciones. Son problemas que no pueden ser resueltos por un algoritmo y hay problemas que no estamos preparados para resolver.

En primer lugar, vamos a describir los problemas que un algoritmo no puede resolver. Hay dos categorías: los problemas indecidibles son aquellos que son teóricamente imposibles de resolver por cualquier algoritmo y los problemas intratables, aquellos problemas que no tienen soluciones en un tiempo razonable.

Un ejemplo clásico de problema indecidible es el problema de la detención, un problema de decisión con una respuesta binaria (sí o no) que es indecidible. Plantea el reto de determinar si un programa informático se detendrá o se ejecutará indefinidamente. El enunciado de este problema es como: "¿Es posible escribir un programa que pueda decir, dado cualquier programa y sus entradas, y sin ejecutar el programa, si se detendrá?"

Por otro lado, están los problemas intratables. Algunos problemas intratables se pueden resolver en un tiempo razonable solo para entradas pequeñas, pero el algoritmo no se escala a entradas más grandes. Algunos problemas intratables se resuelven todos los días, aunque sean intratables. Tal tiene algoritmos de enrutamiento, horarios, programación de recursos. Estos tienden a usar una solución adivinada a un problema intratable que luego se puede verificar rápidamente. A esto se le llama heurística, utiliza el conocimiento y la experiencia para hacer conjeturas inteligentes, el enfoque heurístico también puede buscar encontrar una solución adecuada y no necesariamente la perfecta o ideal.

Uno de los problemas intratables más famosos es el Problema del **Viajante de Ventas** (TSP). En este desafío, la tarea consiste en encontrar la ruta más corta que permita a un vendedor visitar cada ciudad en un mapa exactamente una vez antes de regresar a la ciudad de partida. El objetivo es minimizar la distancia total recorrida, lo que supone un gran reto computacional debido al aumento exponencial de posibilidades a medida que crece el número de ciudades.

Hagamos algunos números para observar cómo crece. Para calcular el número de carreteras entre ciudades debemos calcular el factorial del número de ciudades menos uno. El número de rutas que debemos explorar es la mitad del número de conexiones entre ciudades (solo debemos usar cada carretera una vez).

Observando la tabla, podemos observar cómo el número de posibilidades crece muy rápido[3]:

Number of cities	Number of connections between cities	Number of possible routes
5	4! = 24	12
10	9! = 362.880	181.440
15	14! = 87.178.291.200	43.589.145.600

[3] Ejemplo tomado de https://runestone.academy/ns/books/published/mobilecsp/Unit5-Algorithms-Procedural-Abstraction/Limits-of-Algorithms.html

2.6. Limitaciones de los algoritmos de aprendizaje automático

En la era del machine learning (deep), contamos con una enorme colección de algoritmos diseñados para resolver múltiples tareas, por ejemplo: Árboles de Decisión, Bosque Aleatorio, Máquinas de Vectores de Soporte o Redes Neuronales (en todas sus versiones)... Todos ellos se pueden adaptar a diferentes problemáticas mediante un proceso de formación donde les mostramos datos y ellos mismos se adaptan. Mediante este proceso podemos obtener soluciones adecuadas. Es importante tener en cuenta que también tienen sus propias limitaciones.

- **Calidad y cantidad de datos:** Los algoritmos de aprendizaje automático dependen en gran medida de la calidad y la cantidad de datos disponibles para el entrenamiento. Los datos limitados o sesgados pueden dar lugar a predicciones de modelos inexactas o sesgadas. Además, la insuficiencia de datos puede obstaculizar la capacidad del algoritmo para generalizar bien a datos no vistos.

- **Sobreajuste y subajuste:** Los modelos de aprendizaje automático pueden sufrir un sobreajuste, en el que aprenden a capturar ruido o patrones irrelevantes en los datos de entrenamiento, lo que lleva a un rendimiento deficiente en los datos nuevos. Por otro lado, el ajuste insuficiente se produce cuando el modelo es demasiado simplista, lo que resulta en un rendimiento subóptimo.

- **Interpretabilidad:** Muchos modelos de aprendizaje automático, especialmente los complejos como las redes neuronales, carecen de interpretabilidad, lo que dificulta la comprensión y la confianza en sus predicciones. Esto puede ser problemático en ámbitos en los que la interpretabilidad es crucial, como la sanidad y las finanzas.

- **Recursos computacionales:** El entrenamiento de modelos complejos de aprendizaje automático a menudo requiere recursos computacionales significativos, incluido hardware de alto rendimiento e infraestructura de procesamiento de datos a gran escala. Los recursos computacionales limitados pueden limitar la escalabilidad y la practicidad de las soluciones de aprendizaje automático.

- **Experiencia específica del dominio:** El desarrollo de soluciones efectivas de aprendizaje automático a menudo requiere experiencia específica del dominio para preprocesar correctamente los datos, diseñar características relevantes e interpretar los resultados del modelo.

- **Aprendizaje continuo y adaptación:** Los algoritmos de aprendizaje automático pueden tener dificultades para adaptarse a entornos cambiantes o a la evolución de las distribuciones de datos a lo largo del tiempo. Es necesario un seguimiento y un reentrenamiento continuos de los modelos para garantizar que su rendimiento siga siendo sólido y esté actualizado.

Consejos y recomendaciones para que los profesores se involucren con los conceptos de complejidad de los algoritmos

1. Construir una base sólida
Comience con una descripción general básica de lo que son los algoritmos y su desarrollo histórico para proporcionar contexto. Esto ayuda a construir una narrativa que hace que el concepto abstracto sea más tangible y relacionable.

2. Usa analogías y metáforas
Compara los algoritmos con actividades cotidianas como cocinar o navegar por una ciudad. Esto puede ayudar a desmitificar el concepto y hacerlo accesible a los alumnos que pueden no tener una sólida formación técnica.

3. Demostraciones interactivas
Muestre algoritmos en acción a través de sesiones de codificación en vivo o herramientas web interactivas que permiten a los participantes ingresar ejemplos y ver cómo los algoritmos los procesan. Esto puede ser particularmente atractivo e informativo.

4. Fomentar un aula interactiva
Fomente las preguntas y las discusiones para permitir que los participantes expresen su comprensión o confusión sobre los temas que se están tratando. Esto también ayuda a medir su compromiso y comprensión en tiempo real.

5. Aprendizaje incremental
Divida los temas complejos en partes más pequeñas y manejables. Comience con conceptos básicos e introduzca gradualmente más complejidad. Este enfoque de andamiaje ayuda a prevenir la sobrecarga cognitiva.

Actividades de calentamiento

1. Creación sencilla de algoritmos
Pida a los participantes que escriban un algoritmo sencillo para una tarea cotidiana, como preparar té o prepararse para el trabajo. Esta actividad les hace pensar en las secuencias lógicas paso a paso que son fundamentales para el diseño de algoritmos.

2. Exploración histórica de algoritmos
Discuta el algoritmo euclidiano como grupo, explorando sus pasos y cómo se usó históricamente para resolver problemas. Esto puede conducir a una apreciación más profunda de cómo los conceptos fundamentales de los algoritmos tienen raíces de larga data.

3. Juego de roles de la estructura de control
Organice un juego de roles en el que cada participante actúe como parte de una estructura de control en un algoritmo, como un condicional o un bucle. Esto les

ayudará a visualizar y comprender cómo interactúan las diferentes partes de un algoritmo.

4. Desafío de depuración de algoritmos
Proporcione un algoritmo simple con errores intencionales y pida a los participantes que lo "depuren". Esta puede ser una forma divertida y atractiva de profundizar su comprensión de cómo se estructuran y funcionan los algoritmos.

Consejos y recomendaciones de evaluación

1. Retroalimentación continua
Proporcionar retroalimentación inmediata y continua a lo largo del curso. Esto ayuda a los participantes a corregir malentendidos temprano y refuerza su aprendizaje a medida que avanzan.

2. Tareas prácticas
Diseñe tareas que requieran que los participantes apliquen lo que han aprendido en escenarios prácticos. Por ejemplo, podrían optimizar un algoritmo existente o identificar ineficiencias en un algoritmo determinado.

3. Revisión por pares
Incorporar sesiones de revisión por pares en las que los participantes evalúen el trabajo de los demás. Esto no solo proporciona múltiples puntos de retroalimentación, sino que también fomenta el aprendizaje colaborativo y el pensamiento crítico.

4. Usa rúbricas
Desarrollar rúbricas claras y basadas en criterios para evaluar las tareas. Esto garantiza la coherencia en la calificación y expectativas claras para los participantes.

5. Autoevaluación
Anime a los participantes a evaluar su propio trabajo en función de los criterios dados. Esto promueve la autorreflexión y el aprendizaje más profundo.

Estas estrategias ayudarán a los formadores a impartir eficazmente contenidos complejos sobre algoritmos, al tiempo que garantizan que los participantes participen activamente y sean capaces de aplicar sus conocimientos en la práctica.

3. Reconocer el papel de los algoritmos en diversos sectores y los posibles escollos que se derivan de su uso

Obtener una comprensión profunda **del papel de los algoritmos en varios sectores**, al tiempo que se reconocen los posibles escollos, es fundamental para fomentar una visión integral y holística de las implicaciones y responsabilidades vinculadas a la implementación de algoritmos. Esta noción enfatiza la necesidad de una comprensión profunda de cómo los algoritmos pueden influir en diferentes aspectos de la sociedad, la economía y la gobernanza. Las diversas aplicaciones de los algoritmos, desde la mejora de la eficiencia operativa hasta la personalización de las experiencias de los usuarios, demuestran su importante impacto en la vida moderna (Mourtzis et al., 2022). Sin embargo, esta influencia también conlleva riesgos sustanciales, como la posibilidad **de sesgos**, **violaciones de la privacidad y la amplificación de las** desigualdades sociales existentes **(Patel, 2024).**

El despliegue de algoritmos debe guiarse por consideraciones éticas y **marcos regulatorios** para mitigar estos riesgos (Tsamados et al., 2021). Es crucial desarrollar sistemas sólidos de rendición de cuentas que garanticen que los algoritmos no perpetúen ni exacerben las prácticas dañinas o las disparidades. Esto incluye rigurosos procesos de prueba y validación que evalúan la equidad y precisión de los algoritmos , especialmente en áreas sensibles como la atención médica, la justicia penal y los servicios financieros. Las partes interesadas, incluidos los tecnólogos, los responsables políticos y el público, deben entablar diálogos continuos para examinar las implicaciones de las decisiones algorítmicas y garantizar que estas herramientas sirvan a los intereses más amplios de la sociedad (Donovan et al., 2018).

Además, la **transparencia** en las operaciones algorítmicas y los procesos de toma de decisiones es esencial. El público necesita garantías de que las decisiones algorítmicas se toman teniendo en cuenta las consideraciones éticas y están abiertas al escrutinio. Esta transparencia es crucial no solo para generar confianza, sino también para permitir que los expertos y los reguladores realicen auditorías y evaluaciones de manera efectiva (Felzmann et al., 2019). También son vitales las iniciativas educativas que mejoren la alfabetización algorítmica entre el público y los responsables políticos (Reisdorf y Blank, 2021). Tales esfuerzos empoderarán a más personas para que participen en las discusiones y tomen decisiones informadas sobre el uso y la regulación de estas tecnologías.

La investigación y el desarrollo en el campo de la equidad algorítmica y la ética deben seguir evolucionando. A medida que la tecnología avance, surgirán nuevos desafíos, que requerirán soluciones innovadoras que aborden cuestiones técnicas y éticas. La colaboración entre el mundo académico, la industria y el gobierno puede facilitar el intercambio de mejores prácticas y promover el desarrollo de algoritmos más equitativos y responsables.

La conversación sobre los algoritmos también debe incluir el potencial de un cambio positivo. Cuando se diseñan e implementan de manera responsable, los algoritmos tienen el poder de mejorar la accesibilidad, la eficiencia y la equidad. Por ejemplo, en la educación, los algoritmos pueden ayudar a crear experiencias de aprendizaje personalizadas que se adapten a las necesidades individuales de los estudiantes, lo que podría cerrar las brechas en el logro educativo. En la atención médica, los algoritmos predictivos pueden mejorar los diagnósticos y los resultados de los pacientes al identificar los riesgos y recomendar planes de tratamiento personalizados.

En conclusión, la integración de algoritmos en diversos sectores presenta tanto oportunidades como desafíos (Dwivedi et al., 2021). Al fomentar un entorno que fomente las prácticas éticas, la transparencia y la inclusión, la sociedad puede aprovechar los beneficios de los algoritmos y minimizar sus riesgos. Como sugiere Williamson (2018), una comprensión matizada de estas dinámicas es esencial para aprovechar todo el potencial de las aplicaciones algorítmicas de una manera beneficiosa y justa para todos los miembros de la sociedad.

3.1. Papel de los algoritmos en diversos sectores

Como destacan Parker et al. (2016), "los algoritmos han permeado en todos los sectores, revolucionando los procesos y la toma de decisiones". En esta sección, exploramos algunos sectores destacados donde la influencia algorítmica es notablemente significativa:

a. Finanzas: La IA en las finanzas ha evolucionado significativamente desde sus inicios rudimentarios hasta convertirse en una herramienta indispensable en la industria. El uso de la IA en las finanzas se remonta a la década de 1980, cuando se adaptaron programas básicos para automatizar tareas sencillas como la entrada de datos y la contabilidad. Sin embargo, la verdadera fase de transformación comenzó a finales de la década de 1990 y principios de la de 2000 con la llegada de algoritmos de aprendizaje automático más sofisticados y la explosión de la disponibilidad de datos (Aksoy y Gurol, 2021). Las instituciones financieras reconocieron rápidamente el potencial de la IA para proporcionar evaluaciones de riesgos más precisas, un procesamiento de transacciones más rápido y un mejor servicio al cliente. Esto se vio impulsado por el aumento de la potencia computacional y el desarrollo de técnicas avanzadas de análisis de datos. En la década de 2010, la IA había comenzado a remodelar el comercio, la suscripción, la detección de fraudes y la gestión de las relaciones con los clientes de manera profunda (Davenport y Mittal, 2023).

Estos son algunos ejemplos de uso de la IA en el sector financiero:

- Trading algorítmico: Uno de los ejemplos más famosos de trading algorítmico exitoso es Renaissance Technologies, un fondo de cobertura conocido por su Medallion Fund. Este fondo utiliza modelos matemáticos complejos para analizar y ejecutar operaciones a altas velocidades. Mediante el empleo de algoritmos avanzados, Renaissance ha sido capaz de lograr rendimientos sobresalientes, superando con creces el rendimiento del mercado (Qin, 2012).
- Gestión de riesgos: JPMorgan Chase utiliza algoritmos para gestionar el riesgo crediticio. Su sistema evalúa el riesgo de impago de préstamos en función de numerosos factores, como las condiciones económicas, el rendimiento del sector y el historial crediticio individual. Este enfoque algorítmico les permite adaptar sus ofertas de préstamos y minimizar las posibles pérdidas por deudas incobrables.
- Detección de fraudes: PayPal emplea algoritmos de aprendizaje automático para detectar transacciones fraudulentas. Estos algoritmos analizan miles de atributos de transacciones en tiempo real, incluido el monto, el dispositivo utilizado y el historial de transacciones del usuario. Al identificar patrones que se desvían de la norma, los sistemas de PayPal pueden marcar transacciones potencialmente fraudulentas con alta precisión, evitando así pérdidas financieras y protegiendo a los usuarios.
- Servicio al cliente: Bank of America utiliza un asistente virtual impulsado por IA llamado Erica. Esta herramienta algorítmica ayuda a los clientes a administrar sus cuentas, realizar un seguimiento de los gastos y realizar pagos. Erica también puede proporcionar actualizaciones en tiempo real sobre los cambios en las cuentas y sugerir formas de ahorrar dinero en función de los patrones de gasto analizados por algoritmos.

b. Atención médica: La IA en la atención médica ha experimentado una evolución notable, desde las primeras aplicaciones experimentales hasta convertirse en un componente crítico de la práctica médica moderna. Esta transformación ha sido impulsada en gran medida por los avances en el aprendizaje automático, el análisis de big data y la creciente digitalización de los registros sanitarios. A medida que profundicemos, exploraremos las diversas facetas de la implementación de la IA en la atención médica, incluida la asistencia diagnóstica, el tratamiento personalizado, la predicción de enfermedades, la gestión de pacientes y las cirugías robóticas (Bohr y Memarzadeh, 2020). He aquí algunos ejemplos notables:

- **Asistencia diagnóstica:** La capacidad de la IA para analizar grandes conjuntos de datos ha revolucionado los procesos de diagnóstico en la atención sanitaria. Al integrar la IA con las tecnologías de imagen, los profesionales médicos pueden detectar enfermedades con mayor precisión y velocidad que los métodos tradicionales. IBM Watson Health demuestra el poder de la IA para mejorar la precisión del diagnóstico. Los

algoritmos avanzados de IA de Watson pueden analizar el significado y el contexto de los datos estructurados y no estructurados en notas e informes clínicos. Por ejemplo, se ha utilizado en oncología para identificar opciones de tratamiento para pacientes con cáncer mediante el cruce de millones de notas clínicas de oncología con los registros médicos de los pacientes y la literatura existente.

- **Tratamiento personalizado: La** IA facilita la medicina personalizada al considerar los perfiles genéticos individuales, los factores ambientales y las opciones de estilo de vida para adaptar los tratamientos. Este enfoque mejora significativamente los resultados de los pacientes al dirigirse a las terapias que tienen más probabilidades de funcionar para pacientes específicos. **Tempus** utiliza la IA para recopilar y analizar grandes cantidades de datos genómicos y clínicos con el fin de ayudar a los médicos a crear planes de tratamiento personalizados para los pacientes con cáncer. Su plataforma utiliza algoritmos de aprendizaje automático para comprender los datos moleculares y terapéuticos y predecir qué tratamientos pueden ser más eficaces para cada paciente.
- **Predicción y tratamiento de enfermedades**: Los modelos de IA se utilizan cada vez más para predecir la probabilidad de desarrollo, progresión y posibles complicaciones de la enfermedad. Este enfoque proactivo permite intervenciones más tempranas, lo que podría salvar vidas y reducir los costos de atención médica. DeepMind de Google ha desarrollado sistemas de inteligencia artificial que pueden predecir una lesión renal aguda hasta 48 horas antes de que ocurra con una precisión notable. El modelo de IA analiza una amplia gama de datos de salud en tiempo real, lo que permite intervenciones oportunas que pueden prevenir el deterioro y mejorar los resultados.
- **Gestión y monitorización de pacientes**: La IA mejora la gestión de pacientes mediante la monitorización continua de los datos sanitarios a través de la tecnología portátil y los dispositivos IoT. Estas herramientas alertan a los proveedores de atención médica sobre cambios en la condición de un paciente, lo que permite una respuesta inmediata y ajustes continuos de la atención. Virta Health emplea herramientas impulsadas por IA para gestionar enfermedades crónicas como la diabetes tipo 2. Su plataforma monitorea los datos del paciente en tiempo real y ajusta los planes de tratamiento de manera dinámica, lo que ayuda a mantener niveles óptimos de glucosa en sangre y reduce la dependencia de los medicamentos.
- **Cirugía robótica: La** cirugía robótica, asistida por IA, ofrece alta precisión, menos trauma y tiempos de recuperación más rápidos. Los algoritmos de IA proporcionan datos en tiempo real a los cirujanos e incluso pueden automatizar ciertos aspectos de la cirugía para mejorar la seguridad y los resultados. El sistema quirúrgico robótico da Vinci de Intuitive Surgical utiliza la IA para mejorar la precisión quirúrgica. El sistema proporciona a los cirujanos vistas 3D de alta definición del sitio quirúrgico y traduce los movimientos de la mano del cirujano en movimientos más pequeños y precisos de instrumentos diminutos dentro del cuerpo del paciente.

A medida que la IA continúa avanzando, su potencial para transformar la atención médica crece exponencialmente. Estos ejemplos subrayan la capacidad de la IA para mejorar la precisión de los diagnósticos, adaptar los tratamientos a cada paciente, predecir el curso de las enfermedades, gestionar las enfermedades crónicas de forma eficaz y ejecutar procedimientos quirúrgicos con una precisión sin precedentes. La integración de la IA en la atención médica no solo optimiza los resultados de los pacientes, sino que también agiliza el trabajo de los proveedores de atención médica, allanando el camino para enfoques más innovadores de la medicina y la cirugía. El desarrollo y la implementación continuos de tecnologías de IA son muy prometedores para abordar algunos de los problemas más desafiantes de la atención médica en la actualidad.

c. Comercio electrónico: La IA mejora significativamente las experiencias tanto de los clientes como de **las empresas**, especialmente en los sectores minorista y de comercio electrónico. A través de la acumulación de grandes cantidades de datos de los consumidores y su integración en algoritmos de aprendizaje automático, los minoristas pueden desarrollar funcionalidades avanzadas de personalización, recomendación y automatización. Estas funciones impulsadas por IA ahora son estándar en todas las plataformas de compra y sirven para mejorar la participación del cliente y agilizar las operaciones, lo que aumenta las ganancias de la empresa y la eficiencia de los recursos. Estos son algunos ejemplos del uso de la IA en el comercio electrónico y el comercio minorista:

- **Estrategias publicitarias mejoradas**
  Smartly.io: Utiliza la IA para automatizar y optimizar las campañas publicitarias en las redes sociales, lo que reduce significativamente el esfuerzo manual y mejora el rendimiento de los anuncios y la participación en todas las plataformas.
  eBay: emplea la IA para ofrecer recomendaciones de compra personalizadas y consejos al cliente, lo que aumenta la satisfacción y la retención de los compradores.
- **Optimización de ventas automotrices**
  Cox Automotive: Implementa IA a través de su herramienta Esntial para optimizar los procesos de compra de automóviles en línea, incluida la estimación de pagos y la evaluación de riesgos, mejorando el recorrido digital del cliente.
- **Sistemas de entrega eficientes**
  Gopuff: Integra la IA para optimizar las rutas de entrega, garantizando una entrega rápida y eficiente de los productos esenciales del día a día a través de sus centros de microcumplimiento.
- **Desarrollo de productos innovadores**
  Mondelez International: Aprovecha la IA en su sector de investigación y desarrollo para acelerar la innovación y mejorar la rentabilidad en la producción de aperitivos.
- **Experiencias de compra personalizadas**

Mirakl: Utiliza la IA para adaptar las recomendaciones de productos, mejorando la participación del cliente y las ventas en los mercados digitales.
Rue Gilt Groupe: Aplica la IA para refinar las recomendaciones de productos, mejorando la experiencia de compra en sus sitios web de productos de lujo.

- **Gestión avanzada de inventario y cadena de suministro**
  IBM Watson: Apoya a las empresas minoristas en la optimización de las operaciones y la personalización de las interacciones con los clientes a través del análisis de datos en tiempo real.
  Anaplan: Utiliza inteligencia predictiva para pronosticar los resultados comerciales y optimizar las estrategias minoristas, impactando todo, desde la gestión de inventario hasta la adquisición de clientes.
- **Detección de falsificaciones y moderación de contenido**
  3PM Solutions: Emplea IA para detectar productos falsificados y garantizar la precisión de las calificaciones de los vendedores en los principales mercados en línea.
- **IA conversacional y compromiso con el cliente**
  Valyant AI: Desarrolla herramientas conversacionales impulsadas por IA para restaurantes de servicio rápido, mejorando las experiencias de pedido de los clientes a través de tecnologías basadas en voz.

d. Transporte: No hay duda de que la IA está remodelando la industria del transporte al mejorar la eficiencia, la seguridad y las experiencias de los usuarios. Esta sección profundiza en las diversas formas en que la IA se está integrando en los sistemas de transporte, desde el transporte público y la logística hasta la movilidad personal y la gestión del tráfico.

- **Vehículos autónomos**
  Coches autónomos: La IA impulsa las funcionalidades básicas de los vehículos autónomos, permitiéndoles navegar, evitar obstáculos y tomar decisiones en tiempo real. Empresas como Tesla y Waymo lideran los avances en esta área, mejorando significativamente la seguridad vial y la eficiencia de los vehículos.
  Drones para entrega: Empresas como Amazon están experimentando con drones que utilizan la IA para entregar paquetes de forma autónoma, lo que promete reducir los tiempos de entrega y los costes de la logística de última milla.
- **Gestión del Tráfico y Ciudades Inteligentes**
  Sistemas de tráfico inteligentes: La IA se utiliza para analizar los patrones de tráfico y optimizar las secuencias de los semáforos, reduciendo la congestión y mejorando la seguridad vial. Ciudades como Barcelona y Singapur utilizan sistemas de IA para mantener un flujo de tráfico fluido en intersecciones concurridas.
  Mantenimiento predictivo: La IA ayuda a predecir cuándo los vehículos de transporte público y las infraestructuras (como puentes y carreteras)

necesitarán mantenimiento, evitando averías y prolongando la vida útil de los activos públicos.

- **Optimización del transporte público**
  Planificación de rutas: Los algoritmos de IA optimizan los horarios de autobuses y trenes en función de datos en tiempo real y patrones de tráfico históricos, maximizando la eficiencia operativa y minimizando los tiempos de espera de los pasajeros.
  Gestión de la capacidad: Durante las horas punta, **los sistemas de IA pueden predecir el flujo de pasajeros y ajustar la frecuencia de los servicios de tránsito en consecuencia, mejorando la conveniencia de los viajeros.**
- **Carga y Logística**
  Almacenamiento automatizado: Empresas como Amazon utilizan robots impulsados por IA para agilizar las operaciones de almacén, aumentando la velocidad y la precisión de los procesos de recogida y embalaje.
  Enrutamiento optimizado: Las herramientas de IA analizan numerosas variables, como el clima, las condiciones del tráfico y las ventanas de entrega, para sugerir las rutas más eficientes para el envío, ahorrando tiempo y costos de combustible.
- **Características de seguridad mejoradas**
  Sistemas para evitar colisiones: Los sensores impulsados por IA y los sistemas a bordo de los vehículos pueden identificar peligros potenciales y tomar medidas inmediatas para evitar accidentes.
  Sistemas de monitoreo del conductor: Estos sistemas utilizan la IA para monitorear el estado de alerta y la salud general de los conductores para prevenir accidentes causados por la fatiga del conductor o emergencias médicas.
- **Servicio al cliente y soporte**
  Chatbots de IA: Las empresas de transporte emplean chatbots de IA para proporcionar asistencia en tiempo real a los viajeros, gestionando las consultas sobre horarios, tarifas e interrupciones sin intervención humana.
  Recomendaciones de viaje personalizadas: los algoritmos de IA analizan las preferencias del usuario y su comportamiento anterior para ofrecer sugerencias de viaje personalizadas, mejorando la experiencia del cliente.

e. Educación: La IA está transformando el panorama educativo al permitir experiencias de aprendizaje personalizadas, automatizar tareas administrativas y facilitar entornos educativos inmersivos. Esta sección explora las diversas aplicaciones de la IA en la educación, destacando cómo apoya a los docentes, mejora el aprendizaje de los estudiantes y optimiza la gestión educativa.

- **Aprendizaje Personalizado**
  Plataformas de aprendizaje adaptativo: Los sistemas de IA como DreamBox Learning y Knewton brindan experiencias de aprendizaje personalizadas al adaptarse al ritmo y estilo de aprendizaje de cada estudiante. Estas plataformas evalúan los niveles de conocimiento de los estudiantes y ajustan automáticamente el contenido para que se adapte a sus necesidades de aprendizaje.

Tutores y asistentes de IA: Los sistemas de tutoría impulsados por IA, como Carnegie Learning, ofrecen retroalimentación y asistencia en tiempo real a los estudiantes, ayudándolos a comprender temas complejos a través de la instrucción y la práctica personalizadas.

- **Desarrollo de Contenidos Educativos**
  Generación automatizada de contenidos: Las herramientas de IA pueden ayudar a crear y personalizar contenidos educativos. Herramientas como Quillionz utilizan la IA para generar cuestionarios y resúmenes a partir de libros de texto y artículos, mejorando los materiales de estudio sin una amplia intervención humana.
  Aplicaciones de aprendizaje de idiomas: Aplicaciones como Duolingo utilizan la IA para adaptar los cursos de aprendizaje de idiomas a la competencia y el progreso del usuario, lo que hace que el aprendizaje de idiomas sea más accesible y eficaz.
- **Evaluación y retroalimentación**
  Sistemas de calificación automatizados: La IA automatiza la calificación de las pruebas estandarizadas e incluso la redacción de ensayos, lo que reduce la carga administrativa de los educadores y les permite centrarse más en la enseñanza y menos en la calificación.
  Análisis predictivo: Los sistemas de IA analizan los datos de los estudiantes para predecir los riesgos y resultados académicos, lo que permite a los educadores intervenir temprano con los estudiantes en riesgo para mejorar sus resultados de aprendizaje.
- **Automatización Administrativa**
  Procesos de matrícula y admisión: La IA agiliza las tareas administrativas, como las admisiones y la matrícula, automatizando el procesamiento de datos y la toma de decisiones, reduciendo así el papeleo y mejorando la precisión.
  Gestión de recursos: Las herramientas de IA ayudan a gestionar los recursos escolares, la programación y la asistencia de los estudiantes, optimizando las operaciones y reduciendo los costes generales.
- **Accesibilidad mejorada**
  Tecnologías de asistencia: Las herramientas impulsadas por la IA, como el reconocimiento de voz y los convertidores de texto a voz, hacen que los materiales de aprendizaje sean más accesibles para los estudiantes con discapacidades, lo que garantiza la inclusión y la igualdad de oportunidades para todos los alumnos.
  Ayudas visuales y auditivas para el aprendizaje: Las aplicaciones impulsadas por IA, como el Lector inmersivo de Microsoft, ayudan a los estudiantes con dislexia al ofrecer asistencia en la lectura, lo que demuestra que la tecnología puede cerrar las brechas de aprendizaje.
- **Entornos de aprendizaje inmersivos**
  Realidad Virtual (RV) y Realidad Aumentada (RA): La IA integrada con la RV y la RA crea experiencias de aprendizaje inmersivas que simulan escenarios del mundo real, lo que hace que temas complejos como la ciencia y la historia sean más atractivos y comprensibles.

Juegos educativos y simulaciones: Los juegos educativos mejorados con IA se adaptan a las respuestas de los estudiantes, proporcionando un entorno de aprendizaje dinámico que estimula la participación y mejora el aprendizaje.

f. Sector militar y de seguridad: La IA es cada vez más fundamental para mejorar las medidas de seguridad y las operaciones militares. En esta sección se describe cómo las tecnologías de IA se integran en las estrategias de defensa, los sistemas de vigilancia y los protocolos de seguridad, mejorando la eficiencia y la eficacia al tiempo que se abordan desafíos complejos en la seguridad nacional y mundial.

- **Vigilancia y monitoreo mejorados**
  Sistemas de vigilancia automatizados: Los sistemas impulsados por IA se emplean para monitorear áreas sensibles, utilizando análisis de datos en tiempo real para detectar actividades o amenazas inusuales. Estos sistemas pueden diferenciar entre comportamientos normales y sospechosos, reduciendo significativamente los errores humanos y los tiempos de respuesta.
  Vigilancia con drones: Los drones impulsados por IA se utilizan para la vigilancia aérea, proporcionando un reconocimiento exhaustivo sobre zonas de difícil acceso o zonas de conflicto sin poner en riesgo vidas humanas.
- **Ciberseguridad y Défense**
  Detección y respuesta a amenazas: Los sistemas de IA analizan los patrones en el tráfico de red para identificar posibles amenazas cibernéticas, como ataques de malware y ransomware, mucho más rápido de lo que podrían hacerlo los operadores humanos. Una vez que se detectan las amenazas, la IA también puede automatizar las respuestas o recomendar estrategias de mitigación.
  Cifrado de datos: La IA mejora las técnicas de cifrado, lo que dificulta que entidades no autorizadas descifren información confidencial, asegurando así las transmisiones de datos en redes militares y de seguridad.
- **Sistemas de Defensa Autónomos**
  Unidades robóticas de combate: La IA se utiliza para operar unidades robóticas autónomas o semiautónomas que pueden realizar diversas tareas militares, desde funciones de reconocimiento hasta de combate, lo que reduce la necesidad de soldados humanos en operaciones peligrosas.
  Missile Défense Systems: Los algoritmos de IA ayudan a predecir e interceptar amenazas entrantes con alta precisión, como misiles o vehículos aéreos no tripulados (UAV), lo que garantiza posturas de defensa proactivas.
- **Análisis de inteligencia y apoyo a la toma de decisiones**
  Integración y análisis de datos: Los sistemas de IA integran y analizan grandes cantidades de inteligencia de múltiples fuentes, proporcionando una visión general completa que ayuda a los estrategas militares a tomar decisiones informadas de forma rápida y precisa.
  Simulación y entrenamiento: Las simulaciones impulsadas por IA y los entornos de realidad virtual (VR) brindan capacitación realista para el

personal de seguridad y los soldados, preparándolos para una variedad de escenarios sin los riesgos del combate en vivo.

- **Mantenimiento y Logística**
  Mantenimiento predictivo: Las herramientas de IA predicen cuándo es necesario el mantenimiento de los equipos militares, lo que ayuda a prevenir fallos de funcionamiento y prolonga la vida útil de los activos valiosos.
  Optimización de la cadena de suministro: La IA optimiza la logística y la gestión de la cadena de suministro, garantizando que los recursos se asignen y entreguen de manera eficiente donde sea necesario, especialmente en operaciones militares complejas.

g. Medios de comunicación y ocio: La IA se ha convertido en una fuerza transformadora en las industrias de los medios de comunicación y el ocio, mejorando la creación de contenidos, personalizando las experiencias de los usuarios y optimizando las operaciones. Esta sección explora cómo las tecnologías de IA se están integrando en diversos aspectos de la producción, distribución y actividades de ocio de los medios de comunicación.

- **Creación y gestión de contenidos**
  Periodismo automatizado: Las herramientas de IA, como el software de generación de lenguaje natural (NLG), se utilizan para crear artículos e informes de noticias, especialmente para contenido basado en datos, como resultados deportivos y actualizaciones financieras, lo que permite una publicación más rápida y libera a los periodistas humanos para investigaciones en profundidad.
  Producción de vídeo y música: Los algoritmos de IA ayudan en la edición, desde la corrección de color en los vídeos hasta la mezcla de sonido en la música, lo que agiliza los procesos de producción y permite una experimentación más creativa.
- **Sistemas de Personalización y Recomendación**
  Recomendaciones de medios: Las plataformas de streaming como Netflix y Spotify utilizan la IA para analizar los hábitos de visualización y escucha, respectivamente, ofreciendo recomendaciones de contenido personalizadas para mantener a los usuarios interesados y mejorar la satisfacción.
  Publicidad dirigida: El análisis de los datos de los usuarios impulsado por la IA ayuda a las empresas de medios a crear y ofrecer anuncios dirigidos, lo que aumenta la eficiencia y la eficacia de las campañas de marketing.
- **Participación de la audiencia y medios interactivos**
  Chatbots y asistentes virtuales: Los chatbots impulsados por IA brindan interacción en tiempo real con los usuarios, ofreciendo atención al cliente, recomendaciones de contenido y experiencias interactivas en entornos de juegos y realidad virtual (VR).
  Realidad aumentada (RA) y RV: La IA es fundamental para desarrollar experiencias inmersivas de RA y RV, mejorando el realismo y la interactividad de los entornos virtuales utilizados en juegos, visitas virtuales y contenidos educativos.

- **Eficiencia operativa**
  Publicidad programática: La IA automatiza la compra y venta de espacios publicitarios, optimizando la ubicación y los precios en tiempo real para maximizar el retorno de la inversión para los anunciantes y los medios de comunicación.
  Distribución de contenidos: Las herramientas de IA ayudan a optimizar las estrategias de distribución, analizando los patrones de consumo para determinar los mejores canales y momentos para lanzar contenidos con el fin de maximizar el alcance y la interacción.
- **Mejora de los procesos creativos**
  Escritura de guiones y desarrollo de la trama: Las herramientas de IA ayudan en la escritura de guiones al sugerir desarrollos de la trama, el diálogo y las interacciones de los personajes en función de las convenciones del género y las preferencias de la audiencia.
  Diseño artístico y gráfico: Las herramientas impulsadas por IA ayudan a los artistas y diseñadores sugiriendo mejoras, generando ideas y automatizando tareas repetitivas, lo que permite una mayor libertad creativa y experimentación.
- **Gestión y Planificación de Eventos**
  Gestión de multitudes: La IA se utiliza en la planificación y gestión de grandes eventos, analizando los datos de los asistentes para optimizar el diseño del lugar, la programación y las medidas de seguridad.
  Experiencias personalizadas: En parques de ocio y eventos, los sistemas de IA ofrecen itinerarios y experiencias personalizadas basadas en las preferencias y el comportamiento anterior de los visitantes, lo que mejora la satisfacción y la retención de los clientes.

3.2. Posibles inconvenientes del uso de algoritmos

A pesar de sus numerosas ventajas, los algoritmos pueden precipitar escollos significativos, como el sesgo y la discriminación, los problemas de privacidad, la falta de transparencia y responsabilidad, y una dependencia excesiva de la automatización (O'Neil, 2016). Cada uno de estos escollos no solo plantea riesgos éticos y operativos, sino que también podría socavar la confianza pública en las tecnologías de IA. Aquí, exploramos estas categorías con más detalle, proporcionando ejemplos de varios sectores para ilustrar los peligros potenciales y los errores en la implementación de la IA.

a. Sesgo y discriminación: Los sistemas de IA pueden perpetuar y amplificar inadvertidamente los sesgos si se entrenan con conjuntos de datos que no son representativos o contienen errores perjudiciales. Este problema es especialmente importante en sectores como el financiero y el sanitario, donde estos sesgos pueden dar lugar a un trato injusto de las personas.

- Finanzas: Los sistemas de IA utilizados en la calificación crediticia pueden reflejar los sesgos raciales o de género existentes en los datos históricos. Por ejemplo, si un modelo se entrena con datos de aprobación de préstamos anteriores y esos datos reflejan sesgos históricos contra un grupo demográfico en particular, el modelo también puede discriminar contra ese grupo.
- Atención médica: Ha habido casos en los que las herramientas de diagnóstico de IA han mostrado discrepancias en las recomendaciones de tratamiento basadas en diferencias raciales o de género. Un ejemplo notable es un sistema de IA que diagnosticó erróneamente ciertas enfermedades de la piel en personas de piel más oscura debido a un conjunto de datos que consistía predominantemente en tonos de piel más claros.

b. Preocupaciones de privacidad: La gran cantidad de datos necesarios para entrenar y operar los sistemas de IA plantea importantes preocupaciones de privacidad, particularmente sobre cómo se recopilan, almacenan y utilizan los datos. Esto es evidente en sectores como el comercio electrónico y la educación, donde los datos personales se utilizan ampliamente para mejorar las experiencias de los clientes y estudiantes.

- Comercio electrónico: Los minoristas que utilizan la IA para personalizar las experiencias de compra pueden recopilar grandes cantidades de información personal, que va desde el historial de compras hasta los datos de ubicación en tiempo real. Esto genera preocupaciones sobre la posibilidad de violaciones de datos y el uso no autorizado de información personal.

- Educación: En el sector educativo, los sistemas de IA utilizados para realizar un seguimiento del progreso de los estudiantes pueden recopilar información confidencial sobre dificultades de aprendizaje, datos de salud e incluso patrones de comportamiento. Esto plantea riesgos en cuanto a quién tiene acceso a estos datos y cómo podrían utilizarse más allá del contexto educativo.

c. Falta de transparencia y rendición de cuentas: Los sistemas de IA a menudo operan como **"cajas negras",** con procesos de toma de decisiones que son oscuros u ocultos para los usuarios y los reguladores. Esta falta de transparencia puede dar lugar a problemas de rendición de cuentas, especialmente cuando las decisiones tienen consecuencias significativas en la vida de las personas.

- Seguridad y militar: En aplicaciones militares, la toma de decisiones impulsada por la IA en armas autónomas puede conducir a resultados de vida o muerte. La falta de transparencia en la forma en que se toman estas decisiones complica las implicaciones éticas y plantea importantes preocupaciones sobre la rendición de cuentas.
- Atención médica: Es posible que un sistema de IA utilizado en el diagnóstico de pacientes no proporcione información sobre su proceso de toma de decisiones, lo que dificulta que los médicos entiendan cómo se llegó a un determinado diagnóstico. Esto no solo hace que sea más difícil confiar en el juicio de la IA, sino que también complica la responsabilidad en casos de diagnóstico erróneo.

d. Dependencia excesiva de la automatización: Depender en gran medida de la IA puede conducir a una degradación de la experiencia humana, ya que las habilidades se atrofian cuando los sistemas de IA se hacen cargo de los procesos de toma de decisiones. Esta dependencia excesiva puede ser particularmente problemática en entornos de alto riesgo como el transporte y las finanzas.

- Transporte: La industria de la aviación ha visto incidentes en los que la dependencia excesiva de los sistemas automatizados ha contribuido a los accidentes. Los pilotos a veces dependen demasiado del piloto automático y pueden carecer de suficiente experiencia en vuelo manual para tomar el control de manera efectiva en caso de falla del sistema.
- Finanzas: **El mercado de valores ha experimentado varias caídas repentinas, atribuidas en parte a los algoritmos de negociación automatizados**. La dependencia excesiva de estos sistemas sin suficiente supervisión humana puede provocar caídas repentinas del mercado basadas en errores algorítmicos.

Consejos y recomendaciones para que los profesores se involucren con el papel de los algoritmos en varios sectores

1. Casos de estudio de uso
Comience las sesiones examinando estudios de casos del mundo real en los que los algoritmos han desempeñado un papel crucial en varios sectores. Esto proporciona una comprensión práctica de cómo se aplican los conceptos abstractos en diferentes industrias como las finanzas, la atención médica y el comercio electrónico.

2. Resalte las consideraciones éticas
Enfatice las implicaciones éticas del uso de algoritmos, discutiendo los posibles sesgos, las preocupaciones sobre la privacidad y la responsabilidad. Esto fomentará el pensamiento crítico y la conciencia ética entre los participantes.

3. Fomente las discusiones interactivas
Facilitar discusiones grupales sobre las implicaciones de la implementación de algoritmos. Plantee preguntas como "¿Qué podría salir mal con este algoritmo en una aplicación del mundo real?" o "¿Cómo podemos mitigar los riesgos potenciales?"

4. Utilizar recursos multimedia
Incorpora vídeos, infografías y podcasts que ilustren el impacto de los algoritmos. Esta variedad puede adaptarse a diferentes estilos de aprendizaje y mantener el contenido atractivo.

5. Simular el impacto del algoritmo
Utilice simulaciones o herramientas de software que permitan a los participantes ver los efectos de los cambios en los algoritmos en entornos simulados. Este enfoque práctico ayuda a solidificar la comprensión a través de la aplicación práctica.

Actividades de calentamiento

1. Mapeo de roles de algoritmos
Asigna roles a los participantes donde cada persona representa un componente de un algoritmo en un sector específico. Deben explicar el impacto de su papel y los posibles escollos, promoviendo una comprensión inmersiva de las funciones de los algoritmos y las implicaciones éticas.

2. Debate sobre la ética de los algoritmos
Organizar un debate sobre un uso controvertido de algoritmos, como el reconocimiento facial o la vigilancia predictiva. Esto anima a los participantes a explorar y articular diferentes perspectivas sobre los desafíos éticos.

3. Análisis de escenarios
Presente a pequeños grupos escenarios que impliquen el uso de algoritmos en varios sectores, pidiéndoles que identifiquen los posibles beneficios, riesgos y las decisiones éticas que los responsables deberían tener en cuenta.

Consejos y recomendaciones de evaluación

1. Análisis del estudio de caso
Asigne a los participantes que analicen un estudio de caso en el que se utilicen algoritmos en un sector específico. Evaluar su capacidad para identificar tanto los beneficios como las consideraciones éticas descritas en el caso.

2. Presentaciones grupales
Hacer que los participantes expongan diferentes aspectos de las funciones y riesgos de los algoritmos en diversos sectores. Evalúe su comprensión en función de la profundidad del análisis y la capacidad de comprometerse con consideraciones éticas.

3. Ensayos reflexivos
Pida a los participantes que escriban ensayos en los que reflexionen sobre cómo la comprensión de los algoritmos puede influir en sus responsabilidades profesionales y consideraciones éticas. Evaluar su capacidad para integrar los contenidos del curso con escenarios profesionales personales o hipotéticos.

4. Cuestionarios y pruebas
Utilice los cuestionarios para evaluar la comprensión de conceptos clave como el sesgo algorítmico y la transparencia. Incluya preguntas basadas en escenarios que requieran pensamiento crítico más allá de la memorización.

5. Comentarios de los compañeros
Incorporar la evaluación entre pares como parte del proceso de evaluación. Esto no solo proporciona múltiples perspectivas sobre la comprensión del participante, sino que también fomenta el compromiso con el material de aprendizaje desde el punto de vista de un revisor.

Con el uso de estas estrategias, los formadores pueden educar eficazmente a los participantes sobre el impacto significativo de los algoritmos en todos los sectores, asegurándose de que comprendan tanto los aspectos tecnológicos como las implicaciones éticas más amplias.

4. Comprender la complejidad algorítmica, la optimización y las limitaciones de la toma de decisiones algorítmicas

4.1. Complejidad algorítmica

En ciencias de la computación, los algoritmos pueden hacer o deshacer el éxito de los sistemas y las aplicaciones porque necesitan funcionar de manera eficiente. La complejidad algorítmica es como una luz guía para desarrolladores e ingenieros. Se trata de averiguar cuánta potencia de cálculo necesitan los diferentes algoritmos, y utilizamos la **notación Big O** para hacerlo. Pero no se trata solo de hacer que las cosas funcionen más rápido; Comprender la complejidad algorítmica nos ayuda a saber qué tan bien funcionarán nuestros algoritmos a medida que hagamos nuestros sistemas más grandes, y cómo administrar los recursos que necesitan.

La complejidad algorítmica se refiere básicamente a la rapidez y el espacio que necesita un algoritmo para realizar un trabajo. La complejidad del tiempo nos dice cuánto tiempo tarda un algoritmo en resolver un problema en función de la magnitud del problema. La complejidad del espacio, por otro lado, tiene que ver con la cantidad de memoria que utiliza el algoritmo mientras está funcionando. Estas mediciones nos ayudan a averiguar qué tan bien funciona un algoritmo y si puede manejar grandes tareas sin ralentizarse demasiado.

En el mundo de la informática, existe esta cosa llamada notación Big O, que es la forma matemática de hablar de lo rápido o lento que se ejecutan los algoritmos. Nos ayuda a comprender cuánto tiempo o espacio necesita un algoritmo a medida que le damos más cosas con las que trabajar. Entonces, si decimos que un algoritmo es O(n), significa que a medida que le damos más entrada (llamémoslo "n"), tarda linealmente más tiempo en terminar. Pero si es O(n^2), entonces tomará mucho más tiempo a medida que aumente 'n', porque está creciendo cuadráticamente. Por lo tanto, la notación Big O es una forma rápida de comprender qué tan eficiente será un algoritmo cuando se trata de una gran cantidad de datos.

La importancia del análisis de la complejidad algorítmica abarca varios ámbitos:

- **Predicción del rendimiento:** Cuando analizamos la complejidad de los algoritmos, podemos anticipar cómo se comportarán con diferentes entradas. Esta capacidad de predecir el rendimiento es crucial para elegir el algoritmo más apropiado para un problema en particular y para optimizar el rendimiento del sistema.

- **Evaluación de la escalabilidad:** En entornos donde los tamaños de entrada pueden cambiar, comprender la complejidad algorítmica nos ayuda a evaluar qué tan bien pueden escalar los algoritmos. Los algoritmos con

características de complejidad favorables funcionan de forma fiable en una amplia gama de tamaños de entrada, lo que garantiza la escalabilidad y la capacidad de respuesta.

- **Gestión de recursos:** El uso eficiente de los recursos es fundamental en entornos donde los recursos son limitados, como los sistemas integrados y la computación en la nube. El análisis de la complejidad algorítmica informa las decisiones sobre cómo asignar los recursos, lo que garantiza un uso óptimo y minimiza el desperdicio.

- **Diseño de algoritmos:** La comprensión de la complejidad algorítmica influye en el proceso de diseño, guiando a los desarrolladores hacia algoritmos que equilibren la eficiencia y la funcionalidad. Al seleccionar algoritmos con características de complejidad favorables, los desarrolladores pueden crear sistemas que funcionen de manera óptima sin sacrificar la funcionalidad.

4..2. Complejidad algorítmica del tiempo

Consideremos el problema omnipresente de la clasificación. Los algoritmos como el ordenamiento por burbujas, que tienen una complejidad algorítmica denotada como O(n2), tienden a tener un rendimiento deficiente cuando se trata de conjuntos de datos extensos en comparación con alternativas más eficientes como el ordenamiento por fusión o el ordenamiento rápido, que cuentan con una complejidad algorítmica de O(n log n). Comprender esta diferencia es fundamental para tomar decisiones bien informadas a la hora de seleccionar algoritmos. Esto, a su vez, conduce a mejoras en el rendimiento del sistema y, posteriormente, mejora la experiencia del usuario (Karp, 2010).

Para una ilustración más clara, consideremos un escenario en el que necesitamos ordenar un conjunto de datos que contiene mil millones de entradas (n = 1.000.000.000), lo que podría ser similar a ordenar a los usuarios de una gran red social.

- **Clasificación de burbujas:** Con una complejidad algorítmica de O(n2), cuando se aplica a este conjunto de datos masivo, el tiempo de ejecución sería astronómico. Para dar una estimación, el tiempo de ejecución sería proporcional a aproximadamente 1018, lo que resultaría en un número asombroso. Si cada entrada requiere 1 nanosegundo para su clasificación, el tiempo total requerido por la ordenación por burbujas para procesar mil millones de entradas ascendería a 1018 nanosegundos, lo que equivale a 11574,07 días o aproximadamente 31,6881 años. Es probable que este proceso ineficiente provoque retrasos significativos en la clasificación de los datos, lo que provocaría una mala experiencia del usuario.

- **Clasificación rápida:** Por otro lado, con una complejidad temporal de O(n log n), la ordenación rápida demuestra ser significativamente más eficiente. Al ordenar un conjunto de datos de mil millones de entradas, el tiempo de ejecución mediante la ordenación rápida sería proporcional a aproximadamente 1.000.000.000 * log2(1.000.000.000). Si cada entrada requiere 1 nanosegundo para

ordenar, el tiempo de ejecución necesario para que *quicksort* ordene un conjunto de datos que contiene mil millones de entradas sería de aproximadamente 29.897.352.853,986263 nanosegundos, lo que equivale a 29,8974 segundos o aproximadamente 0,00034603 días. En consecuencia, los usuarios experimentarían tiempos de respuesta más rápidos y una experiencia de usuario más fluida en general.

La implementación de algoritmos de clasificación más eficientes, como *quicksort*, ilustra el potencial transformador de la optimización en los procesos computacionales (Bentley, 1999)

¿Por qué ponemos el ejemplo de la clasificación? La organización de los datos es crucial para que los algoritmos funcionen de manera eficiente, lo que equivale a descubrir una gran cantidad de beneficios, como un procesamiento más rápido y unas interacciones más fluidas con el usuario. Imagínese buscando un artículo específico en una habitación desordenada; Lleva mucho tiempo y es laborioso. Del mismo modo, los datos no ordenados en los algoritmos requieren una búsqueda lineal, en la que se inspecciona cada elemento hasta que se encuentra el deseado. Este proceso, con una complejidad de O(n) (que representa el tamaño del conjunto de datos), se asemeja a escanear un libro página por página para localizar una palabra, un método funcional pero ineficiente, especialmente para grandes conjuntos de datos.

Ahora, imagínate la misma habitación ordenada, con los artículos cuidadosamente dispuestos. Encontrar lo que necesitas se vuelve fácil porque sabes dónde buscar. Del mismo modo, los datos ordenados permiten a los algoritmos utilizar la búsqueda binaria, un enfoque significativamente más eficiente. La búsqueda binaria divide el intervalo de búsqueda por la mitad repetidamente hasta encontrar el elemento deseado, con una complejidad de tiempo de O(log n), escalando mejor con conjuntos de datos más grandes. Es como localizar una palabra en un diccionario hojeando las páginas, reduciendo a la mitad el espacio de búsqueda con cada vuelta.

En esencia, la clasificación de datos permite que los algoritmos trabajen de manera más inteligente, no más difícil. Convierte una laboriosa búsqueda lineal en una búsqueda binaria ultrarrápida, lo que reduce los tiempos de procesamiento y mejora el rendimiento general. Por lo tanto, la próxima vez que encuentre datos sin clasificar, recuerde el impacto de la clasificación y su capacidad para desbloquear la eficiencia algorítmica.

4.3. Complejidad algorítmica de la memoria

Al profundizar en la complejidad algorítmica relativa a la memoria, uno se encuentra con un aspecto crítico del análisis de algoritmos que complementa el examen de la complejidad del tiempo. La memoria, un recurso finito en los entornos informáticos, influye profundamente en el rendimiento del algoritmo y la eficiencia del sistema. Explorar el concepto de complejidad algorítmica relacionada con la memoria puede sonar complejo, pero vamos a desglosarlo en términos más simples.

Cuando hablamos de complejidad algorítmica y memoria, básicamente nos referimos a cuánto espacio necesita un algoritmo para resolver un problema. Piense en ello como organizar su bolso para diferentes tareas: algunas tareas pueden necesitar más espacio, mientras que otras necesitan menos.

Imagina que estás clasificando un montón de tarjetas con números escritos en ellas. Una forma de hacerlo se denomina ordenación por combinación. La ordenación combinada divide las tarjetas en grupos más pequeños, ordena cada grupo y, a continuación, las vuelve a juntar. Si bien la ordenación por fusión es realmente buena para ordenar rápidamente, también necesita algo de espacio adicional para trabajar. Este espacio extra es como tener mesas extra para organizar tus cartas mientras las clasificas.

Otro ejemplo es cuando estamos calculando los números de Fibonacci, como los que se encuentran en la naturaleza (como el patrón de los pétalos de una flor o la espiral de una concha marina). Hay una manera de calcular estos números usando un método llamado programación dinámica. Este método nos ayuda a encontrar los números más rápido, pero también necesita algo de espacio adicional para recordar los números que ya ha calculado. Es como tener un cuaderno en el que anotas los números a medida que los descubres.

Ahora, hablemos de cómo almacenamos la información en los programas informáticos. Tenemos diferentes herramientas llamadas estructuras de datos, como matrices y listas, que nos ayudan a mantener las cosas organizadas. Cada una de estas herramientas ocupa una cierta cantidad de espacio en la memoria de la computadora.

Por ejemplo, imagina que estás haciendo una lista con los nombres de tus amigos. Si usas una lista simple, como escribir sus nombres en una hoja de papel, no ocupa mucho espacio. Pero si usas algo más elegante, como una tabla con muchas columnas, puede que ocupe más espacio porque estás almacenando información adicional sobre cada amigo.

Por lo tanto, cuando analizamos la complejidad algorítmica relacionada con la memoria, básicamente estamos pensando en cuánto espacio necesitan nuestros algoritmos y estructuras de datos para hacer su trabajo de manera eficiente. Al

comprender esto, podemos tomar mejores decisiones en la programación para asegurarnos de que nuestros programas funcionen sin problemas y no usen demasiada memoria. Es como organizar tu bolso de forma inteligente para que puedas llevar todo lo que necesitas sin que te pese demasiado pesado.

En resumen, pensar en la memoria y la complejidad algorítmica nos ayuda a escribir mejores programas que funcionen bien y no acaparen todos los recursos del ordenador. ¡Es como ser un organizador inteligente para tu computadora!

La complejidad algorítmica es un concepto crucial en la informática moderna. Ofrece una forma estructurada de evaluar y mejorar la eficiencia de los algoritmos. Mediante el uso de la notación Big O, podemos comprender los rasgos básicos de los algoritmos y cómo se desempeñan con diferentes cantidades de datos. Comprender la complejidad algorítmica ayuda a los desarrolladores e ingenieros a crear sistemas que puedan manejar las demandas en constante cambio de la computación. A medida que avanza la tecnología, conocer la complejidad algorítmica seguirá siendo vital para impulsar la innovación y mejorar lo que pueden hacer los ordenadores.

4.4. Optimización

La optimización se erige como un faro de innovación y eficiencia, guiando a los algoritmos para que funcionen con una velocidad y precisión incomparables. En esencia, la optimización es el arte de refinar los algoritmos para minimizar la complejidad y maximizar el rendimiento, liberando todo su potencial para abordar problemas complejos con delicadeza.

Imagine a los algoritmos como los motores que impulsan nuestro mundo digital, impulsando todo, desde los motores de búsqueda hasta los sistemas logísticos. Al igual que un motor bien ajustado, un algoritmo optimizado funciona sin problemas, navegando rápidamente a través de vastos conjuntos de datos y cálculos intrincados. Pero ¿cómo exactamente la optimización logra esta hazaña?

La optimización abarca una gran cantidad de estrategias destinadas a optimizar los algoritmos. Implica identificar cuellos de botella, eliminar pasos redundantes y ajustar los procesos para garantizar la máxima eficiencia. En esencia, la optimización consiste en encontrar el equilibrio óptimo entre la utilización de los recursos y la calidad de los resultados.

Uno de los objetivos fundamentales de la optimización es minimizar la complejidad. Los algoritmos a menudo enfrentan desafíos cuando se les asigna la tarea de resolver problemas complejos que requieren amplios recursos computacionales. Al optimizar los algoritmos, nuestro objetivo es simplificar su estructura y reducir la carga computacional, haciéndolos más manejables y escalables.

Además, la optimización busca mejorar el rendimiento aprovechando enfoques innovadores. Esto implica explorar técnicas de vanguardia como la computación

paralela, los algoritmos heurísticos y el aprendizaje automático para aumentar la eficiencia. Al adoptar estos avances, la optimización permite a los algoritmos ofrecer resultados más rápidos sin consumir menos recursos. Pensemos, por ejemplo, en la optimización de los algoritmos de búsqueda utilizados por los navegadores de Internet. A través de un ajuste meticuloso y mejoras algorítmicas, los motores de búsqueda pueden examinar rápidamente grandes cantidades de datos para ofrecer resultados relevantes en milisegundos. Esta optimización no solo mejora la experiencia del usuario, sino que también reduce la carga del servidor y el consumo de energía.

Centrándonos en la computación paralela, en el ámbito de la informática, las Unidades de Procesamiento Gráfico (GPU) han surgido como componentes indispensables, especialmente en el contexto de la Inteligencia Artificial (IA) y el aprendizaje automático. A pesar de su uso generalizado, muchos todavía no están familiarizados con el funcionamiento interno de las GPU y su papel fundamental en el avance de las tecnologías de IA.

En esencia, una GPU es un circuito electrónico especializado diseñado para manipular y alterar rápidamente la memoria para acelerar la creación de imágenes en un búfer de fotogramas destinado a la salida a un dispositivo de visualización. Originalmente desarrolladas para renderizar gráficos, las GPU han evolucionado hasta convertirse en procesadores altamente paralelizados capaces de ejecutar miles de tareas computacionales simultáneamente.

A diferencia de las unidades centrales de procesamiento (CPU) tradicionales, que sobresalen en tareas de procesamiento secuencial, las GPU están optimizadas para el procesamiento en paralelo. Consisten en numerosas unidades de procesamiento más pequeñas llamadas "núcleos", cada una capaz de ejecutar tareas de forma independiente. Esta arquitectura paralela permite a las GPU abordar tareas computacionalmente intensivas con una eficiencia notable. La GPU se diseñó para paralelizar tareas que implican realizar la misma operación en varios datos simultáneamente. Este paralelismo permite a las GPU manejar tareas computacionalmente intensivas con una eficiencia notable (renderizado de gráficos y procesamiento de imágenes y videos que implica la realización de numerosos cálculos para cada píxel, simulación y modelado, minería de criptomonedas y operaciones de matrices paralelizadas, entre otras).

El paralelismo es particularmente ventajoso en las aplicaciones de IA, donde las tareas a menudo implican el procesamiento simultáneo de grandes cantidades de datos. Tareas como el entrenamiento de redes neuronales profundas, el reconocimiento de imágenes, el procesamiento del lenguaje natural y el análisis de datos se benefician enormemente de la destreza del procesamiento paralelo de las GPU.

La IA depende en gran medida de operaciones matemáticas complejas, como las multiplicaciones de matrices y las convoluciones, que son fundamentales para entrenar e implementar modelos de aprendizaje automático. Estas operaciones implican la manipulación de matrices grandes y la realización de numerosos

cálculos al mismo tiempo, lo que las convierte en candidatas ideales para la ejecución paralela en GPU.

En este punto, debemos introducir el consumo de energía de los algoritmos de IA. A medida que estos sistemas se han vuelto omnipresentes en varios dominios, es necesario centrarse en la eficiencia energética para mitigar el impacto ambiental y los costos operativos. Las técnicas de optimización y la complejidad algorítmica desempeñan un papel fundamental para lograr el ahorro de energía dentro de los marcos de IA. La optimización implica refinar los algoritmos para obtener el máximo rendimiento de las tareas con recursos computacionales mínimos. El consumo de energía en los marcos de IA surge de tareas como la manipulación de datos y el entrenamiento de modelos. La baja complejidad algorítmica es crucial para reducir el consumo de energía, ya que requiere menos operaciones computacionales.

4.5. Limitaciones de la toma de decisiones algorítmicas

La toma de decisiones algorítmica se ha convertido en una herramienta para automatizar y optimizar procesos en diversos ámbitos. Sin embargo, aunque es prometedora, la toma de decisiones algorítmica alberga limitaciones inherentes que justifican un escrutinio meticuloso. Desde las finanzas hasta la atención médica, desde la educación hasta la justicia penal, se han implementado algoritmos que prometen una eficiencia, precisión y objetividad sin precedentes. De todos modos, existe una compleja red de restricciones que dificultan la efectividad irrestricta de la toma de decisiones algorítmicas (Yeung, 2018).

Sesgo y discriminación: Una de las limitaciones predominantes y perjudiciales que afectan a la toma de decisiones algorítmicas es la inclinación hacia el sesgo y la discriminación. A pesar de los esfuerzos por lograr la imparcialidad, los algoritmos suelen heredar sesgos presentes en los datos de entrenamiento que utilizan. La presencia de injusticias históricas y sesgos sociales en estos datos puede mantener e intensificar las disparidades, lo que da lugar a resultados discriminatorios. Además, la opacidad de las metodologías algorítmicas presenta dificultades para identificar y mitigar los sesgos, lo que aumenta las aprensiones con respecto a la justicia y la igualdad en los procesos de toma de decisiones. Esta puede ser una oportunidad positiva para detectar sesgos y discriminaciones que antes pasaban desapercibidos para los observadores humanos. A través del análisis de grandes conjuntos de datos utilizando algoritmos avanzados, se pueden identificar patrones sutiles y disparidades, lo que permite tomar medidas correctivas para rectificar los sesgos sistémicos. Las características de transparencia y trazabilidad de ciertos algoritmos permiten un escrutinio minucioso del proceso de toma de decisiones, lo que permite a las partes interesadas abordar los sesgos centrales.

Falta de comprensión contextual: La toma de decisiones algorítmica opera dentro de un dominio contextualmente ingenuo, que depende únicamente de métricas cuantitativas y reglas predefinidas. Esta restricción es especialmente pronunciada en entornos intrincados y dinámicos donde las sutilezas contextuales son parte integral de la toma de decisiones. El juicio humano, aumentado por la conciencia contextual y la competencia en el dominio, con frecuencia supera las capacidades algorítmicas para discernir complejidades matizadas y tomar decisiones informadas. Como resultado, los sistemas algorítmicos pueden manifestar rigidez e insuficiencia para adaptarse a contextos en evolución, comprometiendo así la eficacia y aplicabilidad de sus decisiones.

Consecuencias imprevistas y dilemas éticos: La naturaleza determinista de los algoritmos los hace susceptibles a ramificaciones imprevistas y dilemas éticos, derivados de la simplificación excesiva y el reduccionismo inherentes a los procesos de toma de decisiones algorítmicos. Pueden surgir resultados imprevistos, efectos en cascada y repercusiones no deseadas debido a la incapacidad de los algoritmos para adaptarse a la complejidad y complejidad de los escenarios del mundo real. Además, los dilemas éticos relacionados con la violación de la privacidad, la erosión de la autonomía y la responsabilidad moral subrayan la necesidad de una deliberación y supervisión meticulosas en la implementación de sistemas algorítmicos.

4.6. ¿Cómo identificar las posibles fuentes de sesgo y errores de los sistemas algorítmicos? Una cuestión crucial para garantizar la equidad, la transparencia y la rendición de cuentas

Una de las principales fuentes de sesgo en la toma de decisiones algorítmicas se deriva de los datos utilizados para entrenar y probar los algoritmos. Los sesgos presentes en los datos pueden propagarse a lo largo del proceso de toma de decisiones, lo que da lugar a resultados sesgados. Para identificar posibles sesgos en los datos, es necesario realizar un análisis exhaustivo del proceso de recopilación de datos para evaluar cualquier sesgo sistemático o subrepresentación de ciertos grupos, examinar las características demográficas del conjunto de datos para identificar disparidades o desequilibrios que puedan conducir a resultados sesgados y evaluar las etapas de preprocesamiento de datos. como la limpieza de datos y la selección de características, para identificar cualquier transformación o distorsión no deseada de los datos.

Para identificar posibles sesgos y errores en el diseño y la implementación de algoritmos, se deben examinar los modelos algorítmicos y las metodologías que evalúan los supuestos y restricciones subyacentes del algoritmo para determinar si se alinean con las normas éticas y sociales, examinando el proceso de entrenamiento para identificar las posibles fuentes de sobreajuste o subajuste, que pueden conducir a predicciones inexactas o injustas. Investigar la elección

de las características y variables utilizadas por el algoritmo para garantizar que sean relevantes, no discriminatorias y representativas de la población. El sobreajuste y el subajuste hacen referencia a dos problemas comunes que surgen al entrenar modelos de aprendizaje automático. El sobreajuste se produce cuando un modelo aprende a capturar ruido o fluctuaciones aleatorias en los datos de entrenamiento, en lugar de los patrones o relaciones subyacentes. Como resultado, el modelo funciona bien en los datos de entrenamiento, pero no se generaliza a datos nuevos e invisibles. Por otro lado, el ajuste insuficiente se produce cuando un modelo es demasiado simplista para capturar la estructura subyacente de los datos, lo que da lugar a un rendimiento deficiente tanto en los datos de entrenamiento como en los datos nuevos.

Con el fin de identificar posibles sesgos y errores en la evaluación y validación, se deben ensayar metodologías para medir las métricas de rendimiento del algoritmo en diferentes grupos demográficos con el fin de detectar disparidades o inequidades, realizar análisis de sensibilidad para evaluar la solidez del algoritmo a las variaciones en los datos de entrada o los parámetros del modelo, y solicitar comentarios de las partes interesadas. incluyendo a las comunidades afectadas y a los expertos en la materia, para identificar posibles sesgos o preocupaciones éticas que se hayan pasado por alto durante el desarrollo.

Además, es necesario un seguimiento y una auditoría continuos para detectar y mitigar sesgos, errores o consecuencias no deseadas.

4.7. Abordar y mitigar los riesgos y sesgos algorítmicos

Una vez identificados, los riesgos y sesgos algorítmicos deben abordarse mediante un enfoque multifacético que abarque consideraciones técnicas, procedimentales y éticas. Las intervenciones técnicas pueden implicar el refinamiento de los algoritmos a través de técnicas como la eliminación de sesgos de los algoritmos, la diversificación de los datos de entrenamiento y la incorporación de métricas de equidad en la evaluación del modelo. Las intervenciones procedimentales implican la implementación de protocolos de validación sólidos, medidas de transparencia y mecanismos de rendición de cuentas para garantizar que las decisiones algorítmicas se examinen y validen de manera efectiva. Las intervenciones éticas giran en torno al fomento de una cultura de conciencia ética y responsabilidad entre los desarrolladores, los responsables políticos y los usuarios finales, haciendo hincapié en las implicaciones éticas de la toma de decisiones algorítmicas y promoviendo directrices y normas éticas.

La mitigación de los riesgos y sesgos algorítmicos requiere un seguimiento, una evaluación y una adaptación continuos a los retos y contextos en evolución. El monitoreo continuo del desempeño y el impacto algorítmico permite a las partes interesadas detectar y abordar los sesgos y riesgos emergentes de manera

oportuna. Los mecanismos de evaluación, como las auditorías de sesgos, los análisis de sensibilidad y las evaluaciones de impacto, proporcionan información sobre la eficacia de las estrategias de mitigación de riesgos e informan sobre las mejoras iterativas. Además, fomentar la colaboración y el compromiso interdisciplinarios facilita el intercambio de conocimientos y mejores prácticas entre dominios, enriqueciendo el esfuerzo colectivo para minimizar los riesgos y sesgos algorítmicos.

CONSEJOS Y RECOMENDACIONES PARA LOS

Consejos y recomendaciones para que los profesores se involucren con la complejidad algorítmica

1. Introducción conceptual
Comience con una explicación directa de lo que es la complejidad algorítmica, utilizando analogías relacionadas con actividades cotidianas que requieren planificación y recursos, como organizar un evento o planificar un viaje, para ilustrar los conceptos de eficiencia y gestión de recursos.

2. Demostraciones visuales
Utilice ayudas visuales, como tablas y gráficos, para demostrar el rendimiento de los diferentes algoritmos en diversas condiciones. Esto ayuda a los participantes a comprender visualmente cómo la complejidad afecta el rendimiento.

3. Actividades prácticas
Incorporar ejercicios prácticos donde los participantes puedan experimentar con diferentes algoritmos para resolver problemas específicos. Esto podría implicar ejercicios de codificación simples o el uso de herramientas de software que simulen el rendimiento del algoritmo.

4. Discutir aplicaciones del mundo real
Vincular el debate a aplicaciones del mundo real en las que la complejidad algorítmica juega un papel crítico, como el procesamiento de datos en las grandes empresas tecnológicas o la búsqueda de rutas en los sistemas de navegación GPS, para destacar su importancia.

5. Fomentar el aprendizaje colaborativo
Facilitar discusiones grupales y sesiones de resolución de problemas donde los alumnos puedan debatir los mejores algoritmos para usar en diferentes escenarios, promoviendo una comprensión más profunda a través de la colaboración.

Actividades de calentamiento

1. Desafío de clasificación
Organice una actividad práctica en la que los participantes clasifiquen manualmente objetos (como libros o tarjetas) utilizando diferentes métodos para imitar los algoritmos de clasificación. Analice cómo cada método se relaciona con la complejidad temporal y espacial de un algoritmo.

2. Juego de combinación de complejidad de algoritmos
Cree un juego de cartas en el que los participantes hagan coincidir diferentes algoritmos con sus notaciones de Big O y ejemplos de casos de uso óptimos. Este enfoque interactivo ayuda a solidificar la comprensión de los conceptos teóricos a través del juego.

3. Lluvia de ideas sobre eficiencia
Pida a los participantes que hagan una lista de las tareas diarias que realizan y que podrían optimizarse con algoritmos (como clasificar correos electrónicos u organizar archivos). Analice cómo la comprensión de la complejidad algorítmica puede conducir a mejoras en estas tareas.

Consejos y recomendaciones de evaluación

1. Comprobaciones de concepto
Utilice cuestionarios cortos y frecuentes para evaluar la comprensión de conceptos clave como la notación Big O y las diferencias entre la complejidad del tiempo y el espacio. Esto ayuda a garantizar que los participantes comprendan los elementos fundamentales antes de pasar a temas más complejos.

2. Pruebas prácticas de codificación
Si corresponde, incluya pruebas prácticas de codificación en las que los participantes deben escribir o identificar los algoritmos más eficientes para problemas dados. Esto pone a prueba su capacidad para aplicar los conocimientos teóricos en escenarios prácticos.

3. Evaluación basada en proyectos
Asigna un pequeño proyecto en el que los participantes analicen el rendimiento de un sistema o software y propongan mejoras basadas en la complejidad algorítmica. Esto evalúa su capacidad para aplicar su aprendizaje a los sistemas del mundo real.

4. Presentaciones grupales
Pida a los participantes que trabajen en grupos para preparar presentaciones sobre cómo la complejidad algorítmica afecta a diferentes sectores, como las finanzas, la atención médica o la tecnología. Esto evalúa su comprensión y capacidad para comunicar información compleja.

5. Ensayos reflexivos
Pida a los participantes que escriban ensayos reflexivos sobre cómo la comprensión de la complejidad algorítmica puede afectar su trabajo o estudios. Esto ayuda a evaluar su capacidad de integración y reflexión sobre los conocimientos adquiridos.

Con estas estrategias, los formadores pueden impartir eficazmente contenidos complejos sobre la complejidad algorítmica y, al mismo tiempo, garantizar que los participantes participen, comprendan el material a fondo y puedan aplicar sus conocimientos de forma práctica.

5. Referencias

Aksoy, T., & Gurol, B. (2021). Inteligencia artificial en técnicas y tecnologías de auditoría asistida por ordenador (CAATTs) y propuesta de aplicación para auditores. En *Ecosistema de Auditoría y Contabilidad Estratégica en la Era Digital: Enfoques Globales y Nuevas Oportunidades* (pp. 361-384). Cham: Springer International Publishing.

Bentley, P. (1999). *Diseño evolutivo por ordenadores*. Morgan Kaufmann.

Bohr, A., & Memarzadeh, K. (2020). El auge de la inteligencia artificial en las aplicaciones sanitarias. En *Inteligencia Artificial en salud* (pp. 25-60). Editorial Académica.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introducción a los algoritmos*. Prensa del MIT.

Davenport, T. H., & Mittal, N. (2023). *Todo en IA: cómo las empresas inteligentes ganan a lo grande con la inteligencia artificial.* Prensa de negocios de Harvard.

Yeung, K. (2018). Regulación algorítmica: una interrogación crítica. *Regulación y gobernanza, 12*(4), 505-523. https://doi.org/10.1111/rego.12158

Donovan, J., Caplan, R., Matthews, J., & Hanson, L. (2018). *Responsabilidad algorítmica: Un manual*. https://apo.org.au/node/142131

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... y Williams, M. D. (2021). Inteligencia Artificial (IA): Perspectivas multidisciplinarias sobre los desafíos, oportunidades y agenda emergentes para la investigación, la práctica y la política. *Revista Internacional de Gestión de la Información, 57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparencia en la que puede confiar: Requisitos de transparencia para la inteligencia artificial entre las normas legales y las preocupaciones contextuales. *Big Data y Sociedad, 6*(1), 2053951719860542. https://doi.org/10.1177/2053951719860542

Karp, R.M. (2010). Reducibilidad entre problemas combinatorios. En: Jünger, M., et al. *50 años de programación entera 1958-2008*. Springer, Berlín, Heidelberg. https://doi.org/10.1007/978-3-540-68279-0_8

Knuth, D. E. (1997). *El Arte de la Programación de Computadoras: Algoritmos Fundamentales, volumen 1.* Profesional de Addison-Wesley.

Mourtzis, D., Angelopoulos, J., & Panopoulos, N. (2022). Una revisión bibliográfica de los retos y oportunidades de la transición de la Industria 4.0 a la Sociedad 5.0. Energías, 15(17), 6276. https://doi.org/10.3390/en15176276

O'Neil, C. (2016). *Armas de destrucción matemática: cómo el big data aumenta la desigualdad y amenaza la democracia.* Corona.

Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). *Revolución de las plataformas: cómo los mercados interconectados están transformando la economía y cómo hacer que trabajen para usted.* WW Norton & Company.

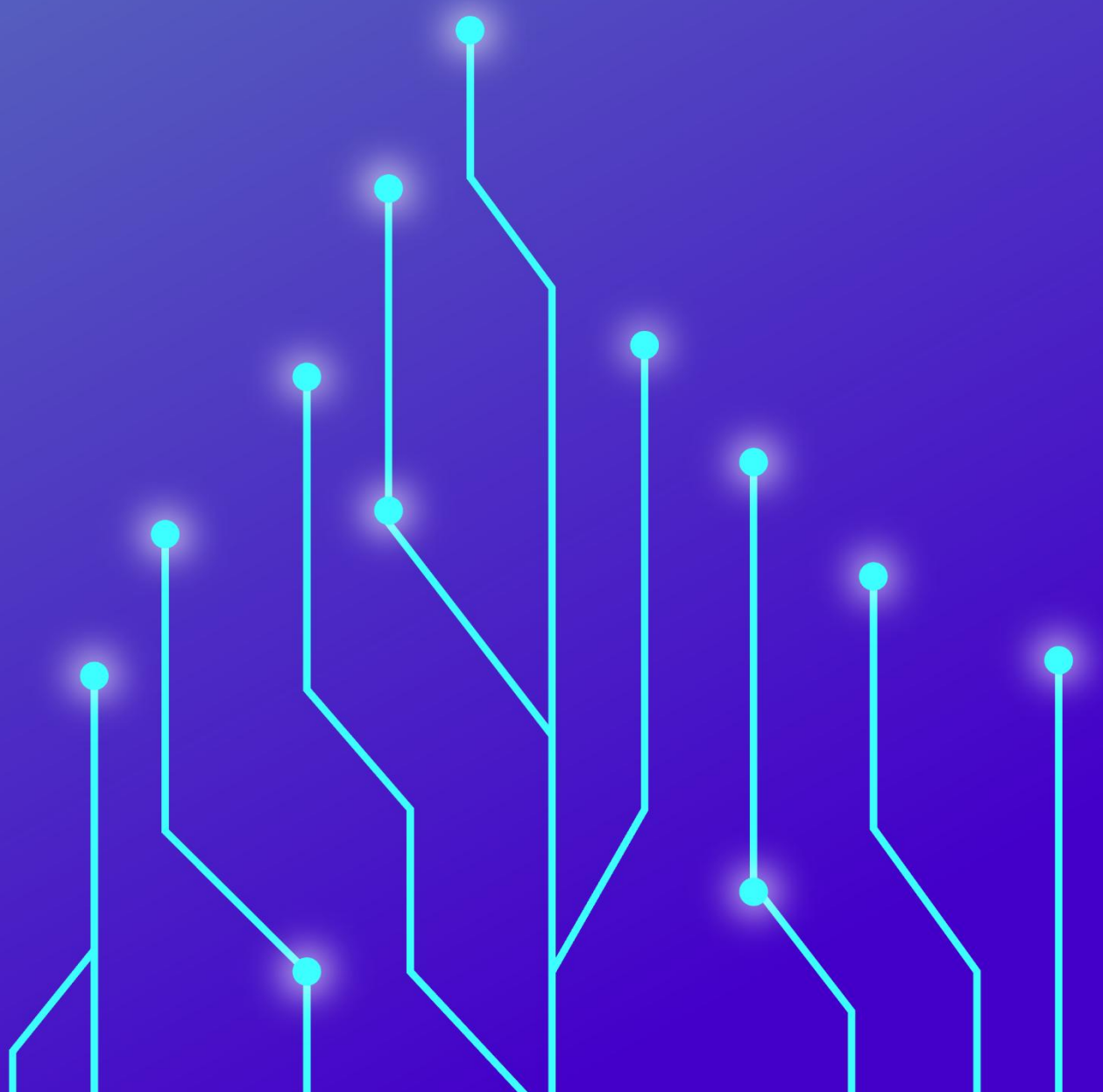
Patel, K. (2024). Reflexiones éticas sobre la IA centrada en los datos: equilibrio entre beneficios y riesgos. *Revista Internacional de Investigación y Desarrollo de Inteligencia Artificial, 2*(1), 1-17. https://orcid.org/0009-0005-9197-2765

Qin, X. (2012). Haciendo uso del big data: la próxima generación de trading de algoritmos. En *Inteligencia Artificial e Inteligencia Computacional: 4ª Conferencia Internacional*, AICI 2012, Chengdu, China, 26-28 de octubre de 2012. Actas 4 (pp. 34-41). Springer, Berlín, Heidelberg.

Reisdorf, B. C., & Blank, G. (2021). Alfabetización algorítmica y confianza en la plataforma. En *Manual de desigualdad digital* (pp. 341-357). Editorial Edward Elgar.

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2021). La ética de los algoritmos: problemas clave y soluciones. En: Floridi, L. (eds) Ética, gobernanza y políticas en inteligencia artificial. Serie Estudios Filosóficos, vol 144. Springer, Cham. https://doi.org/10.1007/978-3-030-81907-1_8

CU4 | Imparcialidad y sesgo de los datos



Índice de contenidos

1. Introducción 115
2. Equidad y sesgo: la analogía de los dulces 116
3. Comprender el panorama: equidad y sesgo de la IA 117
4. Sesgo en los sistemas de IA 118
   4.1. Aparición de sesgos en los sistemas de IA 118
   4.2. Consecuencias del sesgo en la IA 119
   4.3. Abordar el sesgo de la IA 119
   4.4. Descripción general de los tipos de sesgo 119
5. Métodos para identificar y medir los sesgos en la IA 121
6. Estrategias para abordar y mitigar los sesgos 123
7. Comprender las métricas de equidad 124
8. Equidad en la recopilación y el preprocesamiento de datos 125
   8.1. Técnicas de preprocesamiento 126
9. Tendencias emergentes en la equidad de la IA 128
10. Conclusión 130

1. Introducción

Los sistemas de Inteligencia Artificial (IA) no son solo conceptos abstractos, sino que son cada vez más parte integral de nuestra vida cotidiana. Influyen en las decisiones en sectores que van desde la sanidad y la educación hasta las finanzas y la seguridad. Sin embargo, a medida que crece la influencia de la IA, también lo hace el potencial de estos sistemas para perpetuar los sesgos sociales existentes o introducir otros nuevos. Comprender los conceptos de equidad y sesgo en la IA no es solo un ejercicio académico, sino que es crucial para desarrollar sistemas justos y equitativos que impacten en nuestras vidas.

Este CU sobre la equidad y el sesgo de la IA se trata de comprender el problema y empoderarlo para que sea parte de la solución. Su objetivo es arrojar luz sobre cómo pueden surgir inadvertidamente los sesgos en los algoritmos de IA, los posibles impactos de estos sesgos y las estrategias para mitigarlos. El curso también explora los desafíos éticos, sociales y técnicos de la creación de sistemas de IA justos, enfatizando su papel en las prácticas éticas de desarrollo de IA. Al completar esta CU, podrá:

Identificar y comprender los tipos de sesgo

- Reconocer los diferentes tipos de sesgos que pueden ocurrir en los sistemas de IA, incluyendo el sesgo de datos, el sesgo algorítmico y el sesgo de resultados, y comprender los mecanismos por los cuales estos sesgos se manifiestan.

Medir y cuantificar el sesgo

- Comprender las herramientas y metodologías para evaluar y medir el impacto de los sesgos en los sistemas de IA, utilizando una variedad de métricas de equidad y técnicas analíticas.

Desarrollar estrategias de mitigación

- Adquirir conocimientos prácticos sobre cómo implementar estrategias para abordar y reducir el sesgo en los sistemas de IA, incluyendo prácticas diversas de recolección de datos, ajustes algorítmicos y monitoreo de la implementación.

Evaluar la equidad del sistema de IA

- evaluar críticamente los sistemas de IA en términos de equidad en múltiples dimensiones, asegurando que estos sistemas no perpetúen ni exacerben las desigualdades sociales.

Prepararse para desafíos futuros

- anticiparse y responder a los desafíos emergentes en la equidad de la IA a medida que la tecnología evoluciona, asegurando que se mantengan adaptables y con visión de futuro en sus enfoques.

Sugerencia: comience con una breve actividad en la que los participantes compartan sus percepciones o experiencias personales con la IA en su vida diaria. Esto puede ayudar a los participantes a conectar el concepto abstracto de IA con sus implicaciones prácticas.

2. Equidad y sesgo: la analogía de los dulces

FUENTE DE LA IMAGEN | Generado por DALL-E

Imagínate a ti mismo en un salón de clases, mirando un tazón de tus dulces favoritos en el escritorio del maestro. Cada estudiante puede elegir una pieza, pero hay un giro: los estudiantes que usan camisas rojas obtienen dos piezas, mientras que todos los demás obtienen una. Eso no me parece justo. A esto se le llama "sesgo", cuando alguien (o algo, como una IA) da ventajas injustas a personas específicas sin una buena razón.

Ahora, imagínese si, en cambio, el profesor les pide a todos que resuelvan un rompecabezas, y quien lo haga bien, independientemente del color de su camisa, recibe un caramelo extra. Esto se siente justo porque todos tienen la misma oportunidad: se trata del rompecabezas, no de la camiseta. Esto es 'equidad', donde todos tienen las mismas oportunidades y nadie es favorecido por razones arbitrarias.

Al igual que en este escenario de dulces, los sistemas de IA en el mundo real deberían dar a todos una oportunidad justa, ya sea recomendando películas, aprobando préstamos o tocando música. Cuando la IA es justa, todos resuelven el rompecabezas y obtienen los dulces.

Cuando es sesgado, es como si algunas personas estuvieran recibiendo más dulces solo por el color de su camisa. ¡Y queremos aspirar a la equidad!

Sugerencia: emplee esta analogía de forma interactiva haciendo que los participantes simulen el escenario con dulces reales o una actividad virtual. Este enfoque práctico podría ayudar a solidificar los conceptos de equidad y sesgo.

3. Comprender el panorama: equidad y sesgo de la IA

La equidad de datos en inteligencia artificial se refiere al principio de que los sistemas de IA deben tratar a todas las personas de manera equitativa, sin resultados injustos o prejuiciosos basados en características inherentes o adquiridas. La equidad en la IA garantiza que las decisiones tomadas por estos sistemas no favorezcan ni perjudiquen a ningún grupo particular de usuarios sobre otros. Esto es particularmente importante en aplicaciones que afectan la vida de las personas, como la contratación, las aprobaciones de préstamos, la aplicación de la ley y los diagnósticos de atención médica.

El sesgo en la IA, por su parte, se refiere a errores sistemáticos en los datos o algoritmos que conducen a resultados injustos para grupos específicos. Los sesgos pueden ser explícitos, como los incorporados intencionalmente en un sistema, o más a menudo, implícitos, que surgen de datos de entrenamiento no representativos o incompletos o de un diseño algorítmico defectuoso que inadvertidamente beneficia o perjudica a poblaciones específicas.

FUENTE DE LA IMAGEN | Generado por DALL-E

No se puede exagerar el problema de la equidad y el sesgo en el desarrollo de la IA. Los sistemas de IA que operan con sesgos pueden perpetuar y amplificar las desigualdades sociales existentes, lo que conduce a ciclos de desventaja difíciles de romper. Además, los sistemas sesgados erosionan la confianza en la tecnología, lo que podría estancar la adopción y la innovación. Garantizar la equidad en la IA es una necesidad técnica y un imperativo moral para mantener la confianza de la sociedad y prevenir daños.

CONSEJOS Y RECOMENDACIONES PARA LOS

Sugerencia: utilice estudios de casos o artículos de noticias que destaquen los problemas recientes de sesgo en los sistemas de IA, seguidos de discusiones grupales. Esto puede contextualizar la importancia de la equidad en la IA.

Sesgo de género en la IA de contratación laboral

En 2018, se reveló que un sistema de IA utilizado por Amazon para el reclutamiento estaba sesgado contra las mujeres. La IA se había enseñado a sí misma que los candidatos masculinos eran preferibles al observar patrones en las presentaciones de currículums durante un período de 10 años, predominantemente de hombres, debido al desequilibrio de género de la industria tecnológica. El sistema degradaba los currículos que incluían la palabra "femenino", como en "capitana del club de ajedrez femenino" o "universidad femenina".

4. Sesgo en los sistemas de IA

El sesgo en la IA se refiere a errores sistemáticos que dan lugar a resultados injustos para ciertos grupos, favoreciendo normalmente a un grupo demográfico sobre otros. Estos sesgos pueden surgir de diversas fuentes, como los datos utilizados para entrenar los sistemas de IA, los algoritmos que procesan estos datos o los sesgos interpretativos de quienes diseñan e implementan estos sistemas.

4.1. Aparición de sesgos en los sistemas de IA

- **Sesgo de datos**: Este sesgo se produce cuando los conjuntos de datos utilizados para entrenar algoritmos de IA no representan a la población en general. Esto puede suceder debido a desigualdades históricas o simplemente porque los métodos de recopilación de datos capturan a un grupo demográfico más minuciosamente que otros.

- **Sesgo del algoritmo**: A veces, los algoritmos pueden estar diseñados de manera que favorezcan inherentemente resultados específicos. Esto puede deberse a la forma en que los algoritmos interpretan los datos o a los objetivos establecidos durante el desarrollo de la IA.

- **Bucles de retroalimentación**: El sesgo también puede perpetuarse y amplificarse mediante bucles de retroalimentación en los que las decisiones sesgadas iniciales conducen a una mayor recopilación de datos que refuerzan estos sesgos, creando un ciclo de discriminación.

4.2. Consecuencias del sesgo en la IA

- **Impacto social**: El sesgo de la IA puede reforzar las desigualdades sociales y los estereotipos existentes, lo que conduce a una sociedad en la que las tecnologías digitales perjudican sistemáticamente a ciertos grupos. Esta erosión de la confianza en las tecnologías de IA puede tener implicaciones más amplias para la adopción y la eficacia de la IA en diversos sectores.

- **Impacto individual**: A nivel personal, el sesgo en la IA puede conducir a un trato injusto en áreas críticas como la contratación, la aplicación de la ley, la atención médica y los servicios financieros. Las personas pueden enfrentarse a resultados injustos debido a procesos de toma de decisiones defectuosos influenciados por sistemas de IA sesgados.

4.3. Abordar el sesgo de la IA

Es esencial adoptar estrategias integrales a lo largo del ciclo de vida del desarrollo de la IA para mitigar el sesgo de la IA. Entre ellas se encuentran:

- Diversificar las fuentes de datos para garantizar una representación más amplia en los conjuntos de datos de entrenamiento.
- Desarrollar e implementar auditorías algorítmicas para identificar y corregir sesgos.
- Establecer sistemas de monitoreo continuo para detectar y abordar los sesgos emergentes en los sistemas de IA implementados.

4.4. Descripción general de los tipos de sesgo

Comprender los diferentes tipos de sesgos que pueden manifestarse en los sistemas de IA es esencial para los desarrolladores, los responsables políticos y los usuarios que pretenden crear e interactuar con tecnologías justas y equitativas. Esta sección explorará las diversas formas de sesgo, cada una acompañada de ejemplos del mundo real que ilustran su impacto.

1. Sesgo de selección: El sesgo de selección se produce cuando los datos utilizados para entrenar un sistema de IA no representan la población o el escenario objetivo, lo que da lugar a resultados sesgados. Por ejemplo, es posible que un modelo de IA entrenado para reconocer rasgos faciales utilizando un conjunto de datos compuesto principalmente por personas más jóvenes no identifique con precisión los rasgos de los adultos mayores, ya que el

entrenamiento del modelo no incluyó una muestra representativa de todos los grupos de edad.

2. Sesgo de confirmación: El sesgo de confirmación en la IA se manifiesta cuando los datos o su interpretación por parte de los diseñadores de modelos respaldan inadvertidamente creencias preexistentes, ignorando pruebas contradictorias. Una IA utilizada en la contratación entrenada en currículos predominantemente de un género puede seguir favoreciendo a los candidatos de ese género, reforzando el desequilibrio existente en la selección de empleos.

3. Sesgo de muestreo: El sesgo de muestreo es un tipo específico de sesgo de selección en el que los datos de muestra recopilados para entrenar la IA no representan con precisión a la población en general. Si un sistema de reconocimiento de voz se entrena principalmente con datos de hablantes con acento estadounidense, podría tener dificultades para comprender los acentos de otras regiones, como el británico o el australiano, debido a su entrenamiento en una muestra no representativa.

4. Sesgo algorítmico: El sesgo algorítmico se produce cuando los algoritmos que subyacen a los sistemas de IA producen resultados sesgados, a menudo debido a defectos inherentes en su diseño o en los objetivos establecidos durante su creación. Una IA de calificación crediticia que niega préstamos de manera desproporcionada a los solicitantes de ciertos vecindarios debido a datos históricos que reflejan sesgos crediticios pasados ejemplifica el sesgo algorítmico.

5. Sesgo de información: El sesgo de presentación de informes surge cuando los datos introducidos en un sistema de IA están sesgados por la tendencia a informar solo ciertos tipos de resultados. Por ejemplo, un sistema de monitorización de los efectos secundarios de los medicamentos por IA que recibe principalmente informes de casos graves puede sugerir erróneamente que la mayoría de los pacientes experimentan reacciones extremas, ignorando los efectos secundarios más leves pero más comunes.

6. Sesgo de medición: El sesgo de medición implica errores en la recopilación de datos que sesgan sistemáticamente los datos. Una IA de monitoreo ambiental que se basa en sensores menos efectivos en temperaturas más frías tendrá una visión sesgada de los niveles de contaminación, lo que podría informar menos durante los meses de invierno.

5. Métodos para identificar y medir los sesgos en la IA

Comprender y mitigar los sesgos en la inteligencia artificial es fundamental para construir sistemas justos y eficaces. En esta sección se detallan las metodologías para identificar y medir estos sesgos, proporcionando una vía clara para que los desarrolladores de IA y los científicos de datos garanticen que sus sistemas de IA funcionen sin sesgos injustos.

- **La auditoría de datos** es un proceso de revisión exhaustivo en el que los conjuntos de datos utilizados para entrenar modelos de IA se examinan en busca de posibles sesgos o anomalías. Para realizar una auditoría de datos, se debe:
    - **Evaluar la representatividad** de los datos: Evalúe si el conjunto de datos refleja con precisión la diversidad de la población a la que servirá el modelo. Esto incluye la comprobación de la representatividad demográfica y la cobertura de escenarios.
    - **Identifique los datos faltantes**: busque patrones en los datos faltantes: las brechas pueden introducir sesgos si subgrupos específicos están subrepresentados.
    - **Analice la coherencia** de las etiquetas: asegúrese de que el proceso de etiquetado de datos sea coherente e imparcial. Las discrepancias en el etiquetado pueden dar lugar a sesgos significativos en los resultados de los modelos.
    - **Análisis estadístico**: utilice herramientas estadísticas para detectar anomalías o distribuciones de datos sesgadas que puedan indicar datos sesgados.
- **El análisis de disparidades** implica comparar el rendimiento de los modelos en diferentes grupos demográficos u otros subgrupos relevantes para identificar discrepancias que podrían indicar sesgo. Los pasos para realizar un análisis de disparidad incluyen:
    - **Defina subgrupos relevantes:** en función de atributos como la edad, el sexo, el origen étnico u otras características pertinentes para la aplicación del modelo.
    - **Mida las métricas de rendimiento:** calcule las métricas de rendimiento, como la exactitud, la precisión, la recuperación y la puntuación F1 para cada subgrupo.
    - **Compare los resultados:** analice las diferencias en las métricas de rendimiento entre los grupos. Las discrepancias significativas pueden indicar la presencia de sesgo.
    - **Análisis de causa raíz:** Si se encuentran disparidades, investigue las causas subyacentes, que pueden estar relacionadas con la recopilación de datos, la arquitectura del modelo o la selección de características.

- **El análisis** de sensibilidad prueba la solidez de un modelo de IA alterando ligeramente los datos de entrada y observando cómo cambian las salidas. Esto es particularmente útil para identificar sesgos:
    - **Variar los datos** de entrada: introduzca pequeños cambios en los datos de entrada que reflejen variaciones plausibles en escenarios del mundo real.
    - **Supervise las fluctuaciones de salida**: observe si estos cambios conducen a cambios desproporcionados en las salidas del modelo.
    - **Evaluar el impacto en grupos específicos**: Concéntrese en cómo estos cambios afectan los resultados del modelo para diferentes grupos demográficos.
    - **Implementar ajustes**: utilice los hallazgos para refinar el preprocesamiento de datos, la ingeniería de características o el modelo para reducir las sensibilidades no deseadas.

6. Estrategias para abordar y mitigar los sesgos

La mitigación de los sesgos en los sistemas de IA implica un enfoque multifacético durante las fases de diseño, desarrollo e implementación. Las siguientes estrategias son esenciales para **reducir el riesgo de resultados sesgados**.

- **Representación diversa** Garantizar la diversidad en los datos y la composición del equipo es crucial para desarrollar sistemas de IA imparciales.
    - **Diversidad de datos**: Recopile datos de una amplia gama de fuentes para cubrir un amplio espectro de la población. Esto incluye diferentes datos demográficos, antecedentes socioeconómicos y otras características relevantes.
    - **Diversidad** de equipos: Los equipos diversos aportan diversas perspectivas que pueden identificar y mitigar posibles sesgos que podrían no ser evidentes para un grupo más homogéneo.
- **El desarrollo de algoritmos** que sean inherentemente conscientes de los sesgos y que puedan ajustarse a ellos es un enfoque proactivo para mitigarlos.
    - **Incorporar métricas de equidad**: Integre las consideraciones de equidad en la función objetivo del algoritmo, como minimizar la disparidad en las tasas de error entre los grupos.
    - **Entrenamiento de adversarios**: Implementar técnicas de entrenamiento de adversarios en las que los modelos se entrenan simultáneamente para predecir correctamente y minimizar el sesgo.
- **Seguimiento y evaluación periódicos** El seguimiento continuo y la evaluación periódica de los sistemas de IA son vitales para garantizar que permanezcan imparciales a lo largo del tiempo y se adapten a nuevos datos o contextos.
    - **Establecer marcos de supervisión**: Establezca sistemas que realicen un seguimiento continuo del rendimiento de los modelos de IA, centrándose principalmente en las métricas de equidad.
    - **Revisiones periódicas**: Revisar y recalibrar periódicamente los modelos en función de los resultados del seguimiento continuo, adaptándolos a los cambios en los patrones de datos o las normas sociales.
    - **Comentarios de las partes interesadas**: interactúe con las partes interesadas, incluidos los usuarios finales, para recopilar comentarios

sobre el rendimiento de la IA y la equidad percibida, lo que puede proporcionar información que no se captura en las evaluaciones cuantitativas.

7. Comprender las métricas de equidad

Las métricas de equidad son cruciales para evaluar qué tan bien los sistemas de IA tratan a personas o grupos de diferentes orígenes. Proporcionan un marco para medir y garantizar que los sistemas de IA no perpetúen ni exacerben los sesgos sociales existentes. A continuación, se presentan algunas de las **métricas críticas de equidad** que deben tenerse en cuenta:

- **Paridad estadística (paridad demográfica):** esta métrica evalúa si la proporción de resultados positivos es la misma en diferentes grupos demográficos (por ejemplo, género, raza). Es crucial para las aplicaciones en las que el objetivo es garantizar la igualdad de resultados, independientemente de las características de entrada relacionadas con la identidad de grupo.
- **Cuotas igualadas**: esta métrica requiere que la tasa de verdaderos positivos (sensibilidad) y la tasa de falsos positivos (especificidad) sean las mismas en todos los grupos. Se utiliza principalmente en entornos en los que las consecuencias de un error afectan a todos los grupos por igual, como en la justicia penal o la contratación.
- **Disparidad demográfica condicional**: esta métrica evalúa las diferencias en los resultados predictivos condicionados por atributos legítimos. Mide las disparidades no justificadas por atributos relevantes y es adecuado para escenarios en los que se esperan algunas diferencias entre grupos debido a esos atributos.
- **Equidad a través** de la conciencia: Este enfoque implica incorporar la conciencia de atributos sensibles (como la raza o el género) directamente en el proceso de toma de decisiones para garantizar decisiones equitativas. Esta métrica aboga por algoritmos que compensen las injusticias históricas o los sesgos sociales que podrían afectar a la equidad.
- **Equidad individual:** Esta métrica afirma que individuos similares deberían recibir predicciones similares independientemente de su pertenencia al grupo. Este concepto es especialmente relevante en servicios personalizados como las recomendaciones de empleo, donde la coherencia en el tratamiento entre perfiles de usuarios similares es esencial.
- **Equidad contra fáctica**: Según esta métrica, una decisión se considera justa si se hubiera tomado en un mundo contra fáctico en el que el individuo pertenecía a un grupo demográfico diferente, pero era idéntico.

Esta métrica es valiosa para evaluar la equidad a nivel individual y ayuda a comprender el impacto de atributos específicos en las decisiones.

La selección de las métricas de equidad adecuadas depende del contexto y los objetivos específicos de la aplicación de IA. Estas son algunas pautas a tener en cuenta:

1. **Requisitos específicos de la aplicación:** la elección de la métrica de equidad debe alinearse con los objetivos y los requisitos legales del dominio de la aplicación. Por ejemplo, las probabilidades igualadas pueden ser cruciales para las aplicaciones de justicia penal, mientras que la paridad estadística puede ser más apropiada para el marketing o la publicidad.

2. **Consideraciones regulatorias y éticas:** Algunas industrias pueden tener requisitos regulatorios que dictan métricas de equidad específicas. Las consideraciones éticas, como la necesidad de reparar las injusticias históricas, también pueden guiar la elección de las métricas de equidad.

3. **Evaluación de impacto:** Es esencial tener en cuenta el posible impacto social del sistema de IA y elegir métricas que minimicen los sesgos perjudiciales. Por ejemplo, la equidad contra fáctica puede ser crucial en dominios como la atención médica, donde es necesario un tratamiento individualizado.

4. **Viabilidad técnica:** Dependiendo de la complejidad del sistema de IA y de la disponibilidad de datos, algunas métricas pueden ser más difíciles de implementar que otras. Deben tenerse en cuenta las limitaciones técnicas para garantizar que las medidas de equidad elegidas puedan aplicarse de forma precisa y coherente.

8. Equidad en la recopilación y el preprocesamiento de datos

Garantizar que las prácticas de recopilación de datos sean diversas e imparciales es fundamental para crear sistemas de IA justos.

- **Muestreo representativo: Asegúrese** de que el proceso de recopilación de datos incluya muestras de todos los segmentos de población a los que servirá el sistema de IA. Esto puede implicar un muestreo estratificado para incluir grupos subrepresentados.

- **Recopilación continua de datos**: Actualice y amplíe periódicamente las recopilaciones de datos para reflejar las tendencias nuevas y emergentes dentro de la población, ya que los conjuntos de datos estáticos pueden quedar obsoletos rápidamente.

8.1. Técnicas de preprocesamiento

Garantizar la equidad en la recopilación y el preprocesamiento de datos es esencial para evitar que los sesgos se filtren en los sistemas de IA. A continuación, se presentan varias prácticas que pueden ayudar a lograr un sistema de IA más equilibrado y equitativo mediante una atención cuidadosa a las etapas iniciales del desarrollo de la IA.

- **Manejo de datos faltantes**: Analice los patrones de datos faltantes, ya que estos pueden introducir sesgos. Técnicas como la imputación deben aplicarse con cuidado para evitar introducir sesgos adicionales.
- **Selección de características**: evalúe críticamente qué características se incluyen en el proceso de entrenamiento del modelo. Las características correlacionadas con atributos confidenciales deben manejarse con cautela para evitar la discriminación de proxy.
- **Normalización y estandarización**: aplique la normalización o estandarización para garantizar que el modelo no pondere inherentemente ciertas características más simplemente debido a su escala numérica.
- **Diversas fuentes de datos:** Garantizar que la recopilación de datos abarque un amplio espectro de datos demográficos de la población es fundamental para evitar sesgos en los modelos de IA. El objetivo es recopilar datos que reflejen la variedad y diversidad del mundo real en el que operará el sistema de IA. Esto implica interactuar con varios grupos demográficos y utilizar múltiples puntos de recolección. Por ejemplo, al desarrollar una IA sanitaria, se deben recopilar datos de hospitales de diferentes regiones, incluidas las zonas rurales y urbanas, para captar diversos perfiles de salud e impactos medioambientales en las enfermedades.
- **Muestreo inclusivo**: El muestreo inclusivo tiene como objetivo garantizar que todos los segmentos de población estén representados en el conjunto de datos de entrenamiento. El objetivo es evitar la exclusión de grupos minoritarios o infrarrepresentados, cuya ausencia podría dar lugar a predicciones sesgadas de la IA. Se utilizan técnicas como el muestreo estratificado, en las que la población se divide en subgrupos (estratos) que reflejan características clave (por ejemplo, etnia, edad, género), y las muestras se seleccionan aleatoriamente de cada estrato para garantizar la inclusión en el conjunto de datos.
- **Limpieza de datos**: La limpieza de datos implica eliminar imprecisiones e inconsistencias que pueden sesgar los resultados del modelo de IA. El objetivo es garantizar que la calidad de los datos sea alta y esté libre de errores que puedan sesgar las decisiones del modelo. Este proceso incluye la corrección o eliminación de valores atípicos, el relleno de los valores faltantes mediante estrategias de imputación adecuadas que no sesguen

los datos y la garantía de que todas las entradas de datos sean coherentes y tengan el formato correcto.

- **Monitoreo regular**: El monitoreo de los datos y el rendimiento del modelo es crucial para detectar y abordar cualquier sesgo que pueda surgir con el tiempo. El objetivo es mantener una vigilancia continua para garantizar que el sistema de IA siga funcionando de manera justa a medida que se recopilan nuevos datos y cambian las condiciones. Establecer sistemas de monitoreo automatizados que evalúen continuamente los resultados de la IA para obtener métricas de equidad y alerten a los desarrolladores si se detectan sesgos. Este sistema debe evaluar periódicamente los datos que alimentan a la IA para comprobar si hay cambios en la calidad de los datos o en la representación.

- **Ingeniería de características justas:** La ingeniería de características justas implica seleccionar y transformar características para minimizar los sesgos y preservar la utilidad de los datos. El objetivo es garantizar que las aportaciones del modelo no perjudiquen inherentemente a ningún grupo. Las técnicas incluyen el análisis de la correlación de las características con los atributos sensibles y la modificación o eliminación de las características que podrían dar lugar a resultados sesgados. Las técnicas de reducción de dimensionalidad también pueden identificar y eliminar características redundantes o irrelevantes que pueden conllevar sesgos.

- **Transparencia y documentación: Mantener** la transparencia a través de la documentación detallada de los procesos de recopilación de datos, preprocesamiento y selección de características. El objetivo es proporcionar registros claros que puedan ser auditados para verificar el cumplimiento de los estándares de equidad. Mantener registros detallados de las decisiones de manejo de datos, las metodologías utilizadas para la limpieza de datos y la lógica detrás de la selección de características. Estos documentos deben estar a disposición de las partes interesadas y los organismos reguladores para examinar y evaluar la equidad del sistema de IA.

- **Detección de sesgos**: La detección de sesgos implica identificar posibles sesgos en el conjunto de datos antes de que se utilice para entrenar modelos de IA. El objetivo es corregir cualquier sesgo que afecte a la equidad del modelo de forma preventiva. Emplear herramientas de análisis estadístico y visualización para examinar las distribuciones de datos en diferentes grupos demográficos. Esto puede implicar software especializado o algoritmos diseñados para resaltar áreas donde los datos pueden no ser representativos o donde los resultados pueden afectar desproporcionadamente a ciertos grupos.

La equidad de un sistema de IA comienza con la forma en que se recopilan, procesan y preparan los datos incluso antes de que alcancen la etapa de entrenamiento algorítmico. Al implementar prácticas rigurosas en el manejo y preprocesamiento de datos, los desarrolladores pueden reducir

significativamente el riesgo de sesgos en los sistemas de IA. Estas prácticas no solo mejoran la equidad y la fiabilidad de las aplicaciones de IA, sino que también generan confianza entre los usuarios y las partes interesadas en entornos diversos y regulados.

9. Tendencias emergentes en la equidad de la IA

A medida que las tecnologías de IA evolucionan, se encuentran con conjuntos de datos complejos y entornos regulatorios dinámicos. Las nuevas tecnologías, como la IA explicable (XAI) y el aprendizaje federado, ofrecen oportunidades para mejorar la equidad de la IA al proporcionar una mayor transparencia en los procesos de toma de decisiones de la IA y permitir el procesamiento de datos descentralizado que puede ayudar a proteger la privacidad individual. Los desafíos futuros incluyen:

- **Interacciones de datos complejas**: Con la llegada de los grandes datos, comprender las interacciones y correlaciones dentro de conjuntos de datos grandes y complejos se vuelve más desafiante, pero esencial para garantizar la equidad.
- **Regulaciones adaptativas**: A medida que aumenta la conciencia pública sobre los sesgos de la IA, es probable que los organismos reguladores introduzcan pautas de equidad más estrictas, exigiendo que los sistemas de IA sean adaptables y transparentes sobre sus medidas de equidad.
- **Explicabilidad e interpretabilidad: La explicabilidad e** interpretabilidad en IA implica hacer que las operaciones y decisiones de los sistemas de IA sean transparentes y comprensibles para los humanos. El objetivo es garantizar que las partes interesadas comprendan cómo se toman las decisiones de IA, lo cual es fundamental para la confianza y la rendición de cuentas. Esta tendencia implica el desarrollo de métodos y herramientas que permitan la visualización y explicación de los procesos de decisión de la IA. Técnicas como las puntuaciones de importancia de las características, los árboles de decisión y los métodos independientes del modelo proporcionan información sobre el funcionamiento de los modelos complejos, como las redes neuronales profundas.
- **Ataques y defensas adversarias**: Los ataques adversarios implican la manipulación de las entradas de la IA para engañar a los sistemas de IA y hacer que tomen decisiones incorrectas. La defensa contra este tipo de ataques tiene como objetivo hacer que los sistemas de IA sean más robustos y seguros, y protegerlos de entradas maliciosas que podrían explotar sesgos o debilidades. La implementación de defensas contra ataques adversarios incluye el uso de técnicas como el entrenamiento adversario, en el que el modelo se expone a ejemplos adversarios durante el entrenamiento para aprender a resistirlos. Otros enfoques implican

evaluaciones periódicas de seguridad y la actualización de modelos para reconocer y contrarrestar nuevas estrategias de ataque.

- **Equidad en múltiples dimensiones:** La equidad en múltiples dimensiones se refiere a abordar los sesgos que se cruzan con varios factores demográficos y socioeconómicos, incluidos, entre otros, la raza, el género, la edad y la discapacidad. El objetivo es garantizar una equidad integral que tenga en cuenta las identidades humanas complejas y sus interacciones. Esto implica el desarrollo de métricas de equidad multidimensionales y el uso de técnicas de análisis de datos interseccionales para evaluar y mitigar los sesgos en la intersección de múltiples atributos.

- **Equidad en las tecnologías emergentes**: A medida que la IA se integra en tecnologías emergentes como los vehículos autónomos, la atención médica inteligente y los dispositivos IoT, garantizar la equidad en estas aplicaciones se vuelve crucial. El objetivo es implementar la equidad desde cero en estas nuevas tecnologías para evitar sesgos que podrían tener impactos sociales generalizados. Esto implica la realización de evaluaciones de impacto y auditorías de sesgo para las nuevas tecnologías y el diseño de directrices de equidad específicas para el contexto y el caso de uso de cada tecnología.

- **Consideración ética y gobernanza**: La consideración ética y la gobernanza en IA implican el establecimiento de marcos y directrices que garanticen que los sistemas de IA se desarrollen e implementen de manera que se mantengan las normas éticas y los valores sociales. El objetivo es proporcionar un mecanismo de supervisión regulatoria y ética para guiar el desarrollo y el uso de la IA. Desarrollar directrices éticas integrales sobre la IA, formar juntas éticas y colaborar con los responsables políticos para promulgar normas que impongan la equidad, la transparencia y la rendición de cuentas en la IA.

- **Diseño centrado en el ser humano**: El diseño centrado en el ser humano crea sistemas de IA que priorizan el bienestar humano, los valores y las consideraciones éticas. El objetivo es garantizar que las tecnologías de IA mejoren las capacidades humanas y la calidad de vida sin reemplazar ni disminuir la participación y la supervisión humanas. Esto implica interactuar con diversos grupos de usuarios durante el proceso de diseño, incorporar bucles de retroalimentación que permitan a los usuarios informar sobre el desarrollo continuo de la IA y garantizar que la IA apoye la toma de decisiones humanas en lugar de suplantarla.

- **Mitigación de sesgos algorítmicos:** La mitigación de sesgos algorítmicos implica identificar y corregir sesgos en los algoritmos de IA para evitar resultados injustos. El objetivo es perfeccionar continuamente los modelos de IA para garantizar que sean lo más objetivos e imparciales posibles. Esto incluye el uso de técnicas de detección y corrección de sesgos durante el proceso de entrenamiento del modelo, como la reponderación de los datos

de entrenamiento, el ajuste de los parámetros del modelo para equilibrar la precisión y la equidad, y el empleo de algoritmos externos de mitigación de sesgos.

Estas tendencias y cuestiones ponen de manifiesto la naturaleza dinámica del desarrollo de la IA y la necesidad constante de innovación y vigilancia para garantizar la equidad a medida que los sistemas de IA se integran más en todos los aspectos de la vida humana.

10. Conclusión

Explorar la equidad y el sesgo de la IA es crucial a medida que integramos la inteligencia artificial en más facetas de la vida cotidiana y la infraestructura crítica. Desde las definiciones y los impactos de los diversos sesgos en los sistemas de IA hasta las estrategias para su identificación, medición y mitigación, es evidente que la equidad en la IA es un desafío multifacético que requiere una atención integral y continua.

FUENTE DE LA IMAGEN | Generado por DALL-E

- **Comprender los tipos de sesgos y los impactos:** El primer paso hacia la mitigación es reconocer los diferentes tipos de sesgos, como los sesgos de selección, algorítmicos y de confirmación, y comprender su potencial para perpetuar la desigualdad. Ejemplos del mundo real, como el reconocimiento facial, las imprecisiones y los algoritmos de contratación sesgados, han ilustrado los efectos perjudiciales de los sesgos de la IA sin control.

- **Estrategias para la equidad:** Los métodos eficaces para garantizar la equidad incluyen la recopilación de datos diversos, el muestreo inclusivo y el seguimiento periódico de los sistemas de IA. Estas estrategias ayudan a

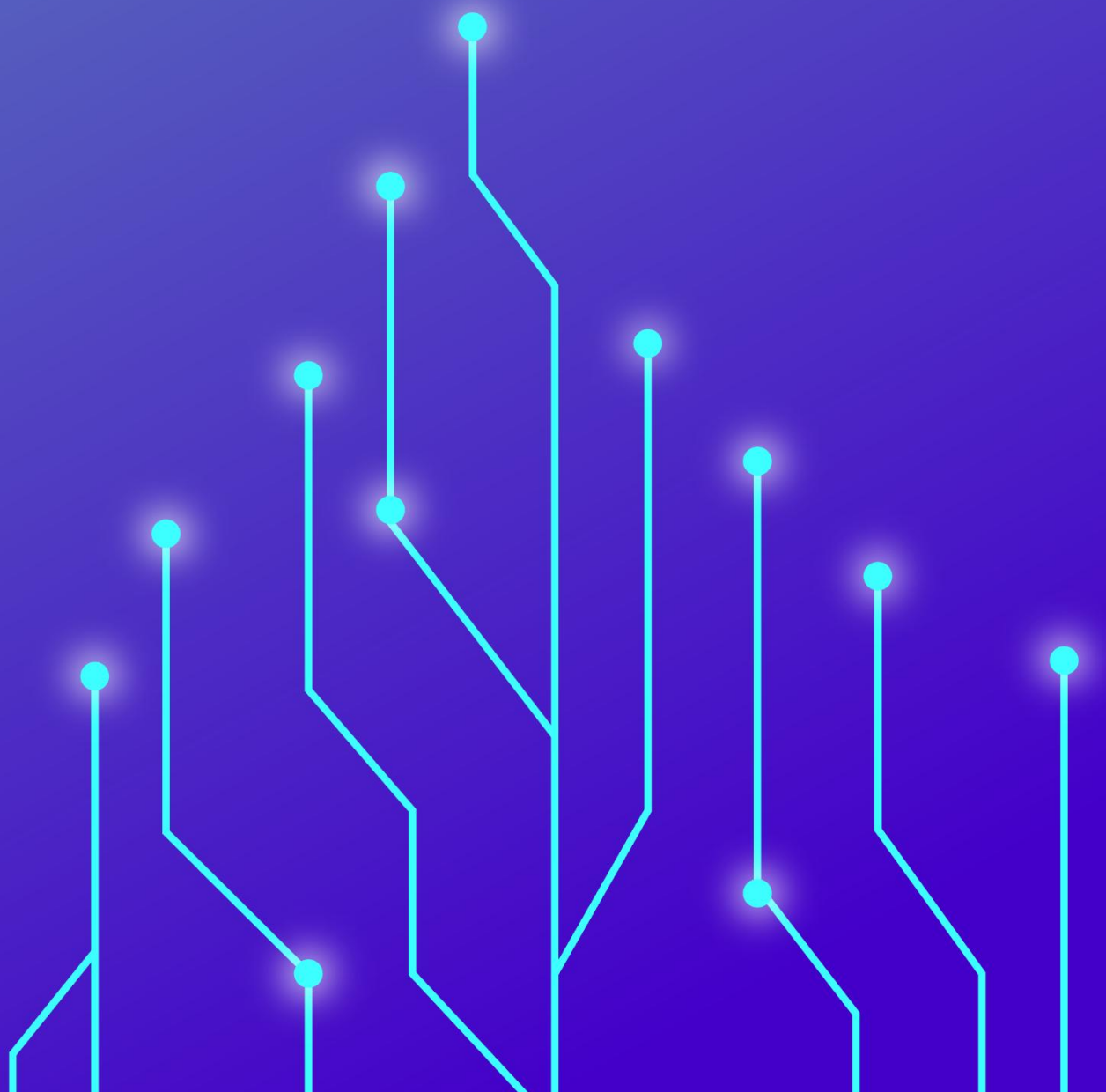
crear modelos de IA que sean representativos y justos y a reducir los errores sistemáticos que perjudican a ciertos grupos.

- **Tendencias y desafíos emergentes: La** equidad de la IA está evolucionando, con nuevos desafíos, como garantizar la equidad en las tecnologías emergentes y desarrollar defensas contra los ataques adversarios. También hay un énfasis creciente en la explicabilidad y el diseño centrado en el ser humano, que son clave para construir sistemas de IA confiables.

- **Consideraciones éticas y gobernanza:** No se puede exagerar la importancia de la gobernanza ética en el desarrollo de la IA. El establecimiento de directrices éticas sólidas y marcos de gobernanza será fundamental para guiar el despliegue responsable de las tecnologías de IA.

- **Aprendizaje y adaptación continuos: A** medida que las tecnologías de IA y las normas sociales evolucionan, también deben evolucionar nuestros enfoques de equidad. Los desarrolladores, los responsables políticos y las partes interesadas necesitan educación, vigilancia y adaptación continuas para garantizar que los sistemas de IA sigan siendo justos y beneficiosos para todos.

CU5 | Casos de estudio y proyectos



ÍNDICE

1. Introducción 134
2. Diferentes tipos de sesgo 135
3. Herramientas para evaluar la ética de una solución de IA 138
4. Proyecto: Solución de IA en reclutamiento 140
5. Identificación de posibles sesgos 144
6. Selección y preselección de los candidatos: definición del conjunto de datos 157
7. Desarrollo del modelo algorítmico 159
8. Cómo medir la equidad en las soluciones de aprendizaje automático 164
9. Implementación del modelo 170
10. Operación y monitoreo 171
11. Conclusión 172
12. Referencias 173

1. Introducción

En esta unidad didáctica, la atención se centra en examinar los sesgos en un contexto práctico. El proyecto consiste en un escenario hipotético en el que una empresa internacional desarrolla una solución de IA para el reclutamiento. Este proyecto se abordará durante la sesión final de estudio interactivo con los estudiantes. Por favor, facilite una discusión animada para apoyar su aprendizaje, alentando a los estudiantes a expresar y aplicar lo que han aprendido en sesiones anteriores.

El sesgo algorítmico es la producción de resultados injustos o discriminatorios por parte de los sistemas algorítmicos. Esto ocurre cuando un algoritmo favorece o perjudica sistemáticamente a ciertos individuos o grupos en función de características específicas, como la raza, el género o el estatus socioeconómico. Los algoritmos de aprendizaje automático operan en varios conjuntos de datos, como los datos de entrenamiento, de los que aprenden modelos aplicables a otros individuos o grupos para hacer predicciones. Sin embargo, estos algoritmos pueden tratar a individuos o elementos similares de manera diferente, lo que lleva a la replicación o magnificación de los sesgos humanos, particularmente aquellos que afectan a los grupos protegidos (Friedman y Nissenbaum, 1996; Lange y Duarte, 2018; Lee et al., 2019). Comprender las causas e implicaciones del sesgo algorítmico es crucial para desarrollar soluciones de IA éticas y confiables.

Implicaciones del sesgo algorítmico

El sesgo algorítmico puede aparecer de varias formas, afectando a diferentes grupos de distintas maneras. Los ejemplos recopilados por Lee, Resnick y Barton (2019) Ilustrar cómo el sesgo puede originarse en múltiples fuentes e impactar a los grupos de manera no intencionada o deliberada.

- Las asociaciones de palabras pueden reflejar sesgos. Investigadores de la Universidad de Princeton descubrieron que los nombres europeos se percibían de manera más positiva que los afroamericanos, y palabras como "mujer" y "niña" estaban más relacionadas con las artes que con las ciencias y las matemáticas (Lee et al., 2019). Como se demostró en el ejemplo inicial de este curso, las asociaciones como "enfermera" con las mujeres e "ingeniero" con los hombres a menudo están moldeadas por datos sesgados incrustados en los algoritmos de búsqueda, lo que refleja estereotipos sociales. Esto perpetúa el sesgo en la IA y el aprendizaje automático debido a la dependencia del contenido en línea existente, lo que plantea un importante desafío continuo en el campo de la IA y el aprendizaje automático.
- En cuanto a las herramientas de contratación en línea, Amazon suspendió el uso de un algoritmo de contratación debido al sesgo de género. El

software de IA penalizaba los currículos que contenían la palabra "de mujeres" y degradaba los de las universidades de mujeres.

- Los anuncios en línea también muestran sesgos. La investigadora de Harvard, Latanya Sweeney, descubrió que las búsquedas de nombres afroamericanos tenían más probabilidades de devolver anuncios de servicios de registros de arrestos en comparación con los nombres blancos.
- La tecnología de reconocimiento facial también puede estar sesgada. La investigadora del MIT Joy Buolamwini descubrió que los sistemas comerciales a menudo no reconocen los tonos de piel más oscuros, ya que se entrenan predominantemente en rostros masculinos y de piel más clara.
- Incluso los algoritmos de la justicia penal pueden estar sesgados. Se descubrió que el algoritmo COMPAS utilizado por los jueces para predecir las decisiones de fianza estaba sesgado contra los afroamericanos, según informó ProPublica.

2. Diferentes tipos de sesgo

Según Friedman y Nissenbaum (1996), los sesgos algorítmicos se pueden clasificar en tres tipos principales.

- En primer lugar, **el sesgo preexistente** se refiere a los sesgos que ya existen dentro de un sistema, ya sea intencionalmente o no, antes de su creación. Estos sesgos pueden entrar en un sistema a través de esfuerzos deliberados o acciones inconscientes, a veces incluso con buenas intenciones.
- En segundo lugar, **el sesgo técnico** surge de las limitaciones o consideraciones técnicas durante el proceso de diseño. Varios aspectos del diseño pueden contribuir al sesgo técnico.
- Tercero **Sesgo emergente** ocurre en el contexto del uso en el mundo real con usuarios reales. Este tipo de sesgo suele surgir una vez finalizado el proceso de diseño, a menudo debido a cambios en el conocimiento de la sociedad, la población o los valores culturales. (Friedman y Nissenbaum, 1996)

Ejemplos de sesgo en conjuntos de datos de aprendizaje automático

Los modelos de aprendizaje automático dependen en gran medida de los datos de entrenamiento, pero la participación humana en la selección y preparación de estos datos puede introducir varios tipos de sesgo. Un ejemplo prevalente es **Sesgo en la presentación de informes**, donde la frecuencia de los eventos en el conjunto de datos no refleja las ocurrencias del mundo real. Este sesgo puede llevar a que los modelos tengan dificultades para manejar situaciones ordinarias

debido a un énfasis excesivo en documentar circunstancias inusuales. (Google para desarrolladores, 2022)

Otro tipo es **Sesgo de atribución grupal**, que se manifiesta en favorecer al propio grupo o estereotipar a los demás en función de las características del grupo. El sesgo implícito, derivado de experiencias personales y modelos mentales, complica aún más las cosas, lo que a veces conduce a **Sesgo de confirmación** o **Sesgo del experimentador** durante el entrenamiento del modelo. (Google para desarrolladores, 2022)

Sesgo de automatización, en el que se favorece a los sistemas automatizados sobre el juicio humano, independientemente de su precisión, y **Sesgo de selección**, derivados de métodos de recopilación de datos no representativos, también plantean importantes retos a la hora de mitigar el sesgo en los modelos de aprendizaje automático. (Google para desarrolladores, 2022)

Ejemplos de casos del mundo real

Después de comprender el marco teórico del sesgo algorítmico y sus implicaciones, ahora es apropiado explorar ejemplos de casos del mundo real que ilustran los impactos del sesgo algorítmico. El primer caso profundiza en el algoritmo dañino de TikTok, mientras que el otro explora las preocupaciones éticas en torno a la tecnología de reconocimiento facial.

El algoritmo dañino de TikTok

En una reveladora investigación de la empresa nacional de medios de comunicación finlandesa Yle, salieron a la luz los efectos nocivos del algoritmo de TikTok. Yle realizó un estudio centrado en el contenido entregado a una niña ficticia de 13 años llamada Ella, que sufría de depresión y problemas de imagen corporal. (YLE, 2023)

En su estudio, los científicos de datos de Yle navegaron por TikTok utilizando el perfil de Ella durante cinco horas. Al principio, Ella se encontró con videos alegres que iban desde comida y mascotas hasta chistes. Sin embargo, en cuestión de minutos, el algoritmo comenzó a publicar contenido relacionado con la salud mental, incluidos los antidepresivos. A medida que continuaba la navegación, el tono de los videos se volvió más oscuro, y TikTok mostraba contenido sobre suicidio, autolesiones y trastornos alimentarios. Al final del estudio, un asombroso 95 % del contenido presentado a Ella se consideró dañino, incluida la glorificación de los trastornos alimentarios y consejos sobre cómo restringir la ingesta de alimentos. (YLE, 2023)

La psicóloga social Suvi Uski destacó el aspecto preocupante del comportamiento del algoritmo, enfatizando su tendencia a priorizar la participación del usuario sobre el daño potencial causado por el contenido

mostrado. Esta revelación subraya las importantes implicaciones éticas asociadas con las recomendaciones algorítmicas de TikTok. (YLE, 2023)

Lo que comenzó como una navegación inocente se convirtió rápidamente en una espiral preocupante de contenido dañino, lo que refleja el estado vulnerable de Ella. La rápida identificación por parte del algoritmo del interés de Ella por los comportamientos depresivos y poco saludables es alarmante en sí misma, pero lo es aún más su papel en la amplificación de estos pensamientos dañinos en lugar de ofrecer un contrapeso. En lugar de redirigir a Ella hacia recursos positivos y de apoyo, el algoritmo reforzó sus tendencias negativas, exacerbando sus problemas de salud mental. Esto plantea preguntas críticas sobre la responsabilidad ética de los diseñadores de algoritmos y los operadores de plataformas a la hora de salvaguardar el bienestar de los usuarios. El caso de Ella subraya la necesidad urgente de una mayor transparencia, rendición de cuentas y consideraciones éticas en el desarrollo y la implementación de sistemas algorítmicos, en particular aquellos que ejercen una influencia significativa sobre el comportamiento de los usuarios y la salud mental.

Recopilación de datos de reconocimiento facial poco ética

La tecnología de reconocimiento facial se ha vuelto cada vez más frecuente en varios sectores, desde la atención médica hasta la aplicación de la ley (Javaid, 2024). Sin embargo, su adopción generalizada ha planteado importantes preocupaciones éticas, en particular en lo que respecta a la recopilación de datos y su impacto en la privacidad y las libertades civiles. Un ejemplo notable de recopilación de datos de reconocimiento facial poco ética surgió en un informe del Washington Post del 7 de julio de 2019. El informe reveló que agencias federales como el FBI y el ICE habían accedido a las bases de datos de licencias de conducir con fines de reconocimiento facial, escaneando las fotos de millones de estadounidenses sin su consentimiento o conocimiento (Harwell, 2019a).

Esta alarmante revelación subraya el problema más amplio del potencial de sesgo y discriminación de la tecnología de reconocimiento facial. Un estudio federal posterior, también reportado por The Washington Post el 19 de diciembre, confirmó la prevalencia de prejuicios raciales dentro de los sistemas de reconocimiento facial. El estudio encontró que las personas asiáticas y afroamericanas tenían hasta 100 veces más probabilidades de ser identificadas erróneamente que los hombres blancos, según el algoritmo y el tipo de búsqueda utilizada (Harwell, 2019b).

3. Herramientas para evaluar la ética de una solución de IA

En el desarrollo de sistemas de IA, particularmente en áreas sensibles a sesgos como la contratación, es esencial emplear varias herramientas diseñadas para identificar y evaluar posibles sesgos. Estas **herramientas de evaluación** son fundamentales para garantizar que las aplicaciones de IA funcionen de manera justa y transparente. Al utilizar estos recursos, los desarrolladores pueden comprender mejor cómo los sesgos pueden influir en los resultados del sistema y tomar medidas proactivas para mitigar estos efectos. Este enfoque no solo mejora la fiabilidad y la equidad de los sistemas de IA, sino que también ayuda a generar confianza con los usuarios al demostrar un compromiso con las prácticas éticas de IA.

Las herramientas de diseño son cruciales para las consideraciones éticas en el desarrollo de la IA, ya que facilitan una comprensión holística de las perspectivas de las diversas partes interesadas. Estas herramientas permiten a los desarrolladores considerar con empatía los diversos impactos de las tecnologías de IA en diferentes grupos de usuarios. Al mapear las interacciones, los comportamientos y las necesidades de estos grupos, estas herramientas ayudan a identificar y mitigar los posibles sesgos inherentes a los sistemas de IA. El amplio compromiso no solo garantiza que las soluciones de IA sean más inclusivas, sino que también alinea el desarrollo de productos con los estándares éticos al abordar de manera proactiva los problemas que podrían conducir a resultados injustos o discriminatorios. En este caso del proyecto, se presentan diferentes herramientas a lo largo del proceso de diseño bajo una etapa relevante.

Si bien hay una gran cantidad de herramientas de evaluación disponibles para este propósito, este curso se enfoca en presentar solo tres de ellas. Las herramientas de diseño utilizadas en este proyecto se detallan en la descripción del proyecto que se encuentra más adelante en este manual, Capítulo E. Identificación de posibles sesgos.

La herramienta de evaluación del impacto ético de la UNESCO, desarrollado en 2023, ofrece un enfoque estructurado para evaluar las implicaciones éticas de los sistemas de IA. Comprende una serie de preguntas diversas, que abarcan formatos abiertos y de opción múltiple, diseñadas para guiar a los usuarios a través de un proceso de evaluación integral. Haciendo hincapié en las consideraciones éticas desde el principio, la herramienta comienza con investigaciones exploratorias antes de profundizar en los principios de la UNESCO. Al abordar aspectos clave como los impactos de las partes interesadas, las responsabilidades del equipo, la seguridad y las preocupaciones con respecto a la sostenibilidad, la privacidad, la transparencia, la explicabilidad y la rendición de cuentas, facilita un examen exhaustivo de las dimensiones éticas inherentes a la implementación de la IA. (UNESCO, 2023)

La sección Equidad, no discriminación y diversidad de la herramienta de evaluación del impacto ético de la UNESCO se centra en la identificación de sesgos algorítmicos, especialmente en los datos utilizados por los sistemas de IA. Examina a fondo las funciones y responsabilidades involucradas para garantizar que se aborden todos los sesgos inherentes. (UNESCO, 2023)

La Lista de Evaluación para la IA Confiable (ALTAI) por el Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial (AI HLEG), creado por la Comisión Europea, proporciona un marco estructurado para evaluar la fiabilidad de los sistemas de IA. Esta herramienta describe siete requisitos específicos para una IA confiable, con explicaciones detalladas disponibles en el sitio web que la acompaña. Además, ofrece orientación sobre cómo completar la evaluación e identifica a las partes interesadas clave que deben participar en el proceso, lo que garantiza una evaluación exhaustiva de la fiabilidad de los sistemas de IA. (ALTAI, 2020)

La sección Diversidad, No discriminación y Equidad de la herramienta ALTAI se centra en evitar los sesgos injustos. Evalúa el sesgo algorítmico a través de preguntas específicas, evaluando los datos de entrada y el diseño del algoritmo para detectar sesgos, al tiempo que promueve la inclusión, la accesibilidad y la participación de las partes interesadas para garantizar la equidad y la transparencia en los sistemas de IA. (ALTAI, 2020)

El diseño de Mario Sosa Hidalgo de un kit de herramientas éticas para el desarrollo de aplicaciones de IA ofrece un enfoque innovador, concreto y visual para integrar la ética en el proceso de desarrollo de aplicaciones de IA. La herramienta aborda los sesgos y aconseja a los desarrolladores que implementen estrategias que minimicen la injusticia, asegurando que el sistema de IA sea lo más imparcial posible. (Hidalgo, 2019)

4. Proyecto: Solución de IA en reclutamiento

Esta sección se centra en abordar el sesgo algorítmico dentro de los sistemas de IA. Presenta un proyecto que involucra un escenario hipotético en el que una empresa internacional desarrolla una solución de IA para el reclutamiento.

Para guiar esta exploración, la narrativa está contenida en cuadros de texto delineados en azul, que sirven como insumos esenciales para el proyecto. Además, esta sección proporciona información completa para ayudar a comprender el contexto y el desarrollo de la aplicación de IA con fines de reclutamiento.

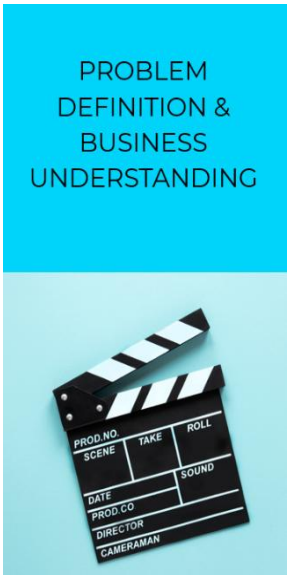
El objetivo es involucrar a los estudiantes en el reconocimiento y mitigación de los sesgos que pueden surgir durante la creación e implementación de tecnologías de IA.

Preparando la escena

Una empresa industrial global se enfrenta a ineficiencias e incertidumbres en su proceso de contratación, lo que supone una presión significativa sobre el departamento de RRHH al consumir tiempo y recursos.
Anticipando el crecimiento futuro, la compañía también reconoce la necesidad urgente de contratar una variedad de especialistas, incluidos ingenieros, comercializadores, profesionales de finanzas, codificadores y gerentes de proyectos.
La feroz competencia por empleados competentes y los altos costos asociados con los errores de reclutamiento enfatizan la importancia de un reclutamiento exitoso. La situación aumenta la presión sobre el equipo de RRHH, cuya tarea es garantizar el éxito y el buen funcionamiento del proceso de contratación.

Como resultado, la compañía está explorando el desarrollo de una solución de inteligencia artificial adaptada a sus necesidades de contratación. Al integrar la IA en el proceso de contratación, el objetivo principal es mejorar la eficiencia, la objetividad y la tasa de éxito general de la contratación.

La inteligencia artificial (IA) se está integrando cada vez más en los procesos de contratación, ofreciendo un abanico de posibilidades para mejorar las prácticas de contratación. Si bien las aplicaciones de IA en la contratación aportan ventajas significativas, también presentan ciertos desafíos. Es crucial que las organizaciones reconozcan y consideren cuidadosamente ambas caras de la moneda. Al comprender los beneficios potenciales, así como los inconvenientes, las empresas pueden tomar decisiones informadas sobre la mejor manera de implementar las tecnologías de IA. Este enfoque equilibrado garantiza que, al tiempo que se aprovecha el poder de la IA para agilizar y mejorar la contratación, también se aborden adecuadamente las implicaciones y los riesgos.



Entendiendo el problema

La empresa comienza una investigación de fondo para desarrollar una solución de IA, formando un equipo interno dedicado a identificar los desafíos clave en el proceso de contratación existente. El equipo realiza entrevistas con una variedad de partes interesadas, incluido el departamento de recursos humanos, los empleados recién contratados y los gerentes que han incorporado recientemente nuevos miembros al equipo. El equipo también recopila datos sobre los costos actuales de contratación y analiza la tasa de éxito de estas iniciativas.
Comprender las percepciones de los empleados potenciales sobre el uso de la IA en la contratación es crucial para la empresa. Su objetivo es evitar cualquier daño a su reputación o marca. Para lograr esto, la compañía se compromete a garantizar que la adopción de la inteligencia artificial no solo obtenga una aceptación generalizada, sino que también se adhiera a los estándares éticos y legales.
Después de discusiones e investigaciones, los pros y los contras de una solución de reclutamiento de IA se enumeran a continuación:

Beneficios de la IA en la contratación

La inteligencia artificial está revolucionando el proceso de contratación al automatizar las tareas rutinarias, lo que permite a los reclutadores concentrarse en aspectos más estratégicos de su trabajo. Esta automatización no solo mejora la eficiencia general en el manejo de las solicitudes, sino que también reduce significativamente el tiempo dedicado a los procesos de contratación. (Albaroudi et al., 2024; www.recruiter.com, 2023)

Además, la capacidad de la IA para analizar amplios conjuntos de datos permite identificar a los candidatos más adecuados, mejorando así la calidad de las contrataciones. Esta capacidad analítica reduce potencialmente la rotación de personal al garantizar una buena correspondencia entre el puesto y el candidato. La retroalimentación rápida y las actualizaciones facilitadas por la IA no solo mejoran la comunicación, sino que también mantienen a los solicitantes bien informados durante todo el proceso de selección. Esto permite a los reclutadores tener más tiempo para interactuar con los candidatos seleccionados, lo que mejora aún más la experiencia del solicitante. (Arivu Reclutamiento y Consultoría, 2023; Sheard, 2022; www.recruiter.com, 2023)

La implementación de la IA en la contratación también conduce a la reducción de costos al minimizar la dependencia del trabajo manual. (www.recruiter.com de 2023) Además, la IA se esfuerza por disminuir los sesgos humanos, aumentando así la objetividad, la coherencia y la equidad en el proceso de contratación. Este avance tecnológico es particularmente beneficioso para los grupos que enfrentan barreras a la participación económica, ya que potencialmente mejora sus perspectivas de contratación. (Bursell y Roumbanis, 2024; Köchling y Wehner, 2020; Sheard, 2022)

La capacidad de la IA para un análisis detallado se extiende a la selección de redes sociales, donde evalúa los perfiles de los candidatos para determinar su idoneidad cultural y su puesto de trabajo. También filtra el contenido inapropiado, lo que garantiza que la contratación se alinee con los valores de la organización. Además, las herramientas de IA ayudan a superar las barreras lingüísticas en la contratación, lo que permite el acceso a un grupo de talentos más amplio y diverso a nivel mundial. (Albaroudi et al., 2024)

Por último, se espera que la IA mejore la diversidad en el lugar de trabajo al eliminar los sesgos del proceso de contratación. (Sheard, 2022) Esta serie de beneficios subraya el impacto transformador de la IA en el panorama de la contratación, prometiendo un proceso de contratación más eficiente, justo e inclusivo.

Preocupaciones de la IA en la contratación

La introducción de la IA en el proceso de contratación, si bien ofrece numerosas eficiencias, trae consigo el desafío del toque humano limitado y la interacción impersonal. Esta reducción en el compromiso personal puede hacer que la experiencia de reclutamiento se sienta mecánica, lo que puede alienar a los candidatos y disminuir la calidad de la experiencia de reclutamiento. (www.recruiter.com de 2023)

Además, el riesgo de sesgo y discriminación plantea una preocupación importante. La IA, si no se diseña o entrena adecuadamente, o si se basa en datos de entrenamiento sesgados, puede introducir inadvertidamente sesgos en el proceso de contratación. Esto puede tener un impacto negativo en la diversidad, la inclusión y la equidad en las prácticas de contratación. (Albassam, 2023; www.recruiter.com, 2023)

La protección de datos y la seguridad también son primordiales, dada la gran cantidad de datos personales recopilados y procesados por los sistemas de IA. Estas preocupaciones plantean problemas en torno a la privacidad, la protección de datos y el riesgo de ciberataques, lo que subraya la necesidad de medidas de seguridad sólidas y transparencia en el uso de los datos. (Albassam, 2023; www.recruiter.com, 2023)

La implementación de la IA en la contratación no está exenta de costes y recursos. Para las pequeñas empresas, en particular, la inversión financiera y de recursos requerida puede ser sustancial, lo que representa una barrera de entrada y puede limitar la adopción de la IA en sus procesos de contratación. (www.recruiter.com de 2023)

También hay preocupaciones con respecto a la falta de juicio humano y los posibles problemas de precisión. La dependencia exclusiva de la IA, sin supervisión humana, puede dar lugar a que se pasen por alto candidatos cualificados debido a datos incompletos o a la falta de reconocimiento de la transferibilidad de las competencias. Además, la precisión de la IA puede verse comprometida por los candidatos que manipulan sus datos para parecer más cualificados. (Albassam, 2023; Arivu Reclutamiento y Consultoría, 2023)

La evolución del marco legal en torno al uso de la inteligencia artificial en la contratación requiere que las empresas se adapten y cumplan para garantizar la equidad. Este panorama en evolución pone de manifiesto la necesidad de una vigilancia y adaptabilidad constantes en el uso de herramientas de IA en el proceso de contratación. (Fernandes, 2021; www.recruiter.com, 2023)

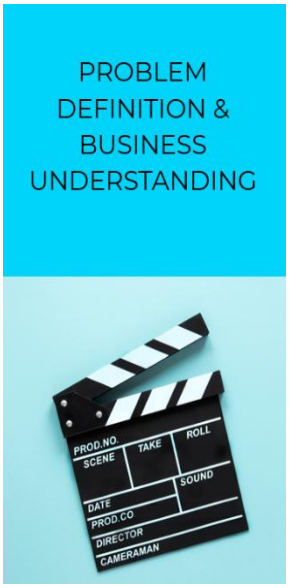
Las limitaciones en la capacidad de la IA para evaluar la idoneidad cultural o las habilidades de trabajo en equipo pueden dar lugar a la pérdida de oportunidades para contratar candidatos que podrían sobresalir si se les diera la oportunidad. Albassam (2023) hace hincapié en la importancia de tener en cuenta estos factores, que a menudo son críticos para el rendimiento laboral, pero difíciles de cuantificar con precisión para los sistemas de IA.

El dominio del mercado y la falta de transparencia en las herramientas de contratación de IA, en particular las desarrolladas por empresas privadas en los EE. UU., complican los esfuerzos hacia la mitigación de sesgos y el escrutinio externo. La naturaleza propietaria de estos sistemas, como propiedad intelectual, dificulta la evaluación y mejora de su equidad y eficacia. (Sheard, 2022)

Por último, la exacerbación de la desigualdad a través de plataformas de contratación automatizadas es una preocupación creciente. Estos sistemas pueden reforzar las desigualdades sociales existentes a través de algoritmos sesgados y el importante papel de las redes y las estructuras informales en la adquisición de empleos. Esta preocupación apunta a la necesidad de una cuidadosa consideración y acción para eliminar los sesgos de las plataformas de contratación impulsadas por IA. (Escuela de Ingeniería y Ciencias Aplicadas John A. Paulson de Harvard, 2023)

Cada uno de estos puntos subraya la complejidad y la naturaleza multifacética de la incorporación de la IA en el proceso de contratación, destacando la necesidad de un enfoque equilibrado que tenga en cuenta la eficiencia, la equidad y la inclusión.

5. Identificación de posibles sesgos



Entendiendo a los usuarios del sistema

Una de las personas entrevistadas es la Lead Recruiter María, quien cuenta con varios años de experiencia en la empresa. Ha visto de primera mano la acumulación de tareas de selección y la creciente presión que provoca en el departamento de RRHH, que se intensifica a la hora de contratar talento global.

María es de mente abierta e interesada en las posibilidades de la IA para ayudar en las tareas de contratación, aunque no está muy familiarizada con la tecnología de inteligencia artificial.

Herramienta: Persona

Las personas se utilizan en el diseño para crear perfiles detallados que representan tipos específicos de usuarios o grupos. Estos perfiles capturan características esenciales sin recurrir a estereotipos. Al incluir detalles realistas como comportamientos, necesidades e información demográfica, las personas se vuelven identificables y útiles para comprender a los diferentes grupos de usuarios. Ayudan a los equipos de diseño y desarrollo a generar empatía, alinear sus esfuerzos y desarrollar soluciones adaptadas a necesidades específicas. Las personas son particularmente valiosas porque ayudan a romper los segmentos de marketing tradicionales, lo que lleva a un diseño más inclusivo y efectivo. (Kumar, 2013; Stickdorn et al., 2018a, 2018b)

Sesgo de la perspectiva de Maria

A la hora de evaluar al personaje, es importante tener en cuenta que sus puntos de vista personales podrían influir en las recomendaciones y opiniones que proporcione con respecto a la adopción de la herramienta. Existe un debate en curso sobre si la IA representa un desafío o un riesgo para los roles tradicionales de reclutamiento. Los profesionales de RRHH podrían percibir los sistemas de contratación de IA como una amenaza para sus puestos de trabajo, lo que podría dar lugar a una reticencia a adoptar herramientas de IA para la contratación (Chen, 2023). Otros sesgos individuales basados en su personalidad podrían ser, por ejemplo, los siguientes:

- Apertura o escepticismo relacionado con la edad hacia las nuevas tecnologías y su impacto en los roles tradicionales.
- La influencia de estar en una ciudad tecnológica como Helsinki, posiblemente sesgada hacia soluciones innovadoras.
- Su origen guatemalteco podría afectar su percepción del impacto de la IA en la diversidad y la inclusión en el reclutamiento.
- Su formación en gestión de RRHH puede fomentar la creencia en la importancia del juicio humano, lo que posiblemente provoque reticencias a confiar en la IA para tareas críticas de contratación.
- Al vivir sola, sus experiencias personales podrían hacer que valore más la interacción humana, lo que la lleva al escepticismo sobre la naturaleza impersonal de la IA.
- Sus objetivos de trabajo hacen hincapié en la eficiencia y el compromiso personal de los candidatos, lo que podría entrar en conflicto con su visión de la contratación de IA.
- Las frustraciones con las altas cargas de trabajo y la lentitud de los procesos podrían hacer que se abriera a la eficiencia de la IA, pero el miedo a la despersonalización y a perder el contacto con los candidatos podría causar dudas.

- Pasatiempos como la natación, la escalada y la lectura indican un enfoque equilibrado y multifacético de la vida, lo que podría traducirse en un deseo de un enfoque matizado para el reclutamiento.
- La preocupación de que la IA altere las funciones de RRHH puede llevarla a cuestionarse cómo se alinea con su conjunto de habilidades y la necesidad futura de su función.

Proceso de contratación

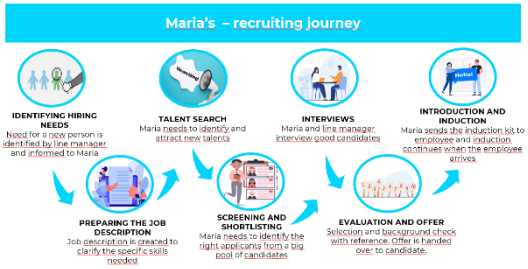
Comprender el proceso de contratación

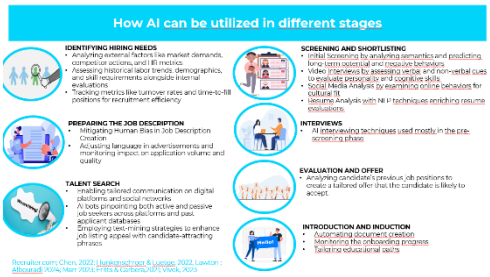
Para obtener una comprensión integral del proceso de reclutamiento, la empresa desarrolla un viaje de reclutamiento que involucra a María y al equipo de reclutamiento. En colaboración con los gerentes de línea, navegan a través de cada etapa del viaje para identificar frustraciones significativas y pasos que requieren mucho tiempo. Además, exploran la posible integración de la IA en cada etapa para mejorar la eficiencia y la eficacia del proceso de contratación.

Herramienta: Mapa de viaje

Un mapa de recorrido es una herramienta visual utilizada en el diseño para representar la experiencia paso a paso que un usuario tiene con un servicio. Representa la perspectiva del usuario, detallando lo que sucede en cada etapa de interacción, los puntos de contacto involucrados y cualquier desafío u obstáculo que pueda enfrentar. Además, los mapas de viaje suelen incluir capas que indican las respuestas emocionales del usuario, destacando las experiencias positivas o negativas a lo largo de su viaje. Estos mapas se pueden utilizar para ilustrar tanto las experiencias actuales (mapas de viaje de estado actual) como las interacciones futuras previstas (mapas de viaje de estado futuro). Varían en alcance, desde descripciones generales de alto nivel de una experiencia completa de extremo a extremo hasta mapas muy detallados que se centran en momentos específicos. (Kumar, 2013; Stickdorn et al., 2018a).

En este proyecto, el mapa de viaje se presenta inicialmente como una representación del estado actual, seguida de una proyección del estado futuro, que describe las capacidades que las herramientas de IA pueden proporcionar para el reclutamiento. El uso de la IA en las diferentes etapas de la contratación se aborda con más detalle en el capítulo "Presentación de soluciones de IA: cómo abordar las etapas del proceso de contratación y sus sesgos asociados".





Las partes interesadas y sus posibles sesgos

Reconocer los sesgos personales

Después de una declaración clara del problema, la empresa ha optado por proceder con una solución impulsada por IA. Ha sido nombrado desarrollador principal de IA responsable de supervisar la implementación de esta tecnología de IA elegida.

La compañía reconoce la importancia de reunir un grupo diverso y de múltiples partes interesadas para asegurar una variedad de perspectivas para el éxito del proyecto.

En un esfuerzo por abordar los posibles sesgos inconscientes dentro del equipo, la compañía ha ordenado que todos los miembros participen en un ejercicio inspirado en Alan Turing llamado "Matriz de

Herramienta: Matriz de posicionalidad de Alan Turing

Reflexionar sobre la "posicionalidad" (comprender el origen social y cultural de cada uno) es crucial en el desarrollo de la IA para la toma de decisiones éticas. El Instituto Alan Turing (2022) subraya la importancia de la "Matriz de Posicionalidad", una herramienta para evaluar cómo los rasgos individuales como la edad, la raza y la educación influyen en la investigación y la innovación. Esta autoconciencia es clave para evitar sesgos y desequilibrios de poder en los sistemas de IA. Al tener en cuenta estos factores personales, los desarrolladores pueden asegurarse de que las soluciones de IA sean equitativas y respeten a las

diversas comunidades a las que sirven, contribuyendo así positivamente sin perpetuar las injusticias.

La matriz de posicionalidad de Alan Turing es un marco reflexivo que revela el impacto de los antecedentes socioeconómicos individuales, los contextos culturales y las experiencias de vida en el juicio y la toma de decisiones dentro de un equipo. Sirve como herramienta para discernir y abordar la influencia de diversos atributos personales, como la edad y el origen étnico, en las perspectivas de investigación y desarrollo. Esta matriz es fundamental para descubrir sesgos inconscientes y navegar por las dinámicas de poder, fomentando así un compromiso con la diversidad y la equidad en el desarrollo de la IA. (Instituto Alan Turing, 2022)

Tener en cuenta a las partes interesadas

Además del grupo de partes interesadas internas que participan en el proyecto, la empresa también debe tener en cuenta a otras partes interesadas relacionadas.

Por lo tanto, la empresa creó un mapa de partes interesadas siguiendo la definición de G.J. Millers del mapeo de partes interesadas del proceso de desarrollo de IA que define a las partes interesadas en el desarrollo y el uso, así como a las partes interesadas externas.

Con base en el mapa, el gerente del proyecto necesita organizar a las partes interesadas en grupos para saber cómo mantener informadas a las partes interesadas y ver cuál es su influencia en el proyecto.

Herramienta: Mapa de partes interesadas

Un mapa de partes interesadas es una herramienta visual que se utiliza para identificar y mostrar las funciones y relaciones de todas las partes interesadas involucradas en un proyecto. Puede abarcar desde un simple cuadrante que muestra los niveles de influencia y compromiso hasta una matriz detallada que detalla las interacciones entre las partes interesadas. El mapa de partes interesadas ayuda a clarificar la dinámica de las partes interesadas y es crucial para comprender y gestionar las relaciones de la red. (Giordano et al., 2018; SDT, s.f.)

Herramienta: Matriz de análisis de grupos de interés

La matriz de análisis de los grupos de interés es una herramienta visual que se utiliza para categorizar y gestionar a los interesados en un proyecto en función de sus niveles de influencia e interés. Al organizar a las partes interesadas en una cuadrícula de cuadrantes, esta matriz ayuda a priorizar las estrategias de participación. El cuadrante superior izquierdo representa a las partes interesadas con alta influencia, pero bajo interés, lo que requiere esfuerzos para mantenerlas satisfechas sin abrumarlas. El cuadrante superior derecho incluye a las partes interesadas y altamente influyentes que necesitan una gestión cercana. El cuadrante inferior derecho, para las partes interesadas con alto interés, pero menor influencia, sugiere actualizaciones periódicas para mantenerlas informadas. Por último, el cuadrante inferior izquierdo para aquellos con un interés e influencia mínimos requiere un compromiso mínimo, centrándose en el monitoreo. (Hoory y Botorff, 2022; Wallbridge, 2023)

Sesgos en las etapas de reclutamiento; Taller

Cuando los grupos de interés están claros, la empresa decide realizar un taller ético en el que se repasen todos los posibles sesgos en las diferentes etapas de la contratación.

Se invita a las partes interesadas del campo "Trabajar juntos" (diapositiva anterior) a unirse.

¿Qué tipo de post habrías rellenado?

Herramienta: Taller

Un taller es una herramienta de diseño colaborativo en la que un grupo de participantes realiza actividades interactivas para abordar un problema específico o trabajar en un proyecto. Por lo general, facilitados por un líder, los talleres pueden variar en duración desde unas pocas horas hasta varios días, dependiendo de los objetivos y la complejidad de las tareas en cuestión. Este método es eficaz para generar ideas, resolver problemas y fomentar el trabajo en equipo y la innovación dentro de un grupo. (Wirtz, 2022)

Presentación de soluciones de IA: abordar las etapas del proceso de contratación y sus sesgos asociados

Los algoritmos y las tecnologías de aprendizaje automático se utilizan cada vez más desde las etapas iniciales del proceso de contratación. Estos enfoques predictivos ayudan a anticipar la evolución del negocio mediante el análisis de factores externos como las demandas del mercado, los movimientos de los competidores y los cambios en las necesidades de los consumidores. Las métricas de RRHH juegan un papel crucial en la creación de estas previsiones. Incluyen un análisis de las tendencias laborales históricas, los cambios demográficos y la evolución de las necesidades de competencias. Además, las evaluaciones internas, como las evaluaciones de la carga de trabajo, las observaciones de los directivos, los comentarios de los empleados y la demografía de la fuerza laboral, son fundamentales para comprender el estado actual y la evolución continua de la base de empleados. Métricas como las tasas de rotación, el tiempo de cobertura de los puestos y el coste por contratación son cruciales para medir la eficacia de la contratación de una organización e identificar las áreas de mejora de los procesos. (www.recruiter.com de 2023)

Los sesgos en la contratación predictiva pueden surgir de varias fuentes, como el nivel de rendimiento de los empleados actuales en funciones específicas, lo que puede conducir a un sesgo de anclaje. Por ejemplo, si un empleado que sobresale en su función necesita ser reemplazado, los análisis pueden sugerir un reemplazo. Sin embargo, el alto nivel de rendimiento del empleado saliente podría requerir dos reemplazos. Por el contrario, el rendimiento de un empleado de bajo rendimiento podría sugerir incorrectamente la necesidad de contrataciones adicionales para compensar su falta de productividad. (Chen, 2023, p. 145) Además, la resistencia al cambio entre los empleados puede conducir a un sesgo del statu quo, en el que hay reticencia a adaptar las estructuras organizativas en respuesta a necesidades empresariales claras. (Sukernek, 2024)

Además, el análisis predictivo a veces puede pasar por alto factores críticos debido a un enfoque limitado. Por ejemplo, no tener en cuenta las tendencias más amplias de la industria o los cambios demográficos, como el envejecimiento de la población, puede hacer que una empresa no esté preparada para los futuros requisitos de la fuerza laboral .

Al igual que otras etapas del proceso de contratación, esta fase también es susceptible al sesgo del bucle de retroalimentación. Este tipo de sesgo puede surgir cuando los sistemas de IA utilizados en la contratación se entrenan con datos históricos y, sin darse cuenta, refuerzan sus propias decisiones prejuiciosas. Esto ocurre en entornos dinámicos donde las decisiones de la IA influyen en los datos de los que aprenderá en el futuro. Si una IA, ya sesgada por datos anteriores, continúa tomando decisiones de contratación sesgadas, estas decisiones retroalimentan su conjunto de datos de aprendizaje, perpetuando e incluso amplificando los sesgos originales (Vivek, 2023, p. 0107).

Sesgos en la descripción del puesto de trabajo

La inteligencia artificial puede facilitar la creación de descripciones de trabajo precisas. Al ajustar el lenguaje de los anuncios y monitorear el efecto de estas modificaciones en el volumen y la calidad de las aplicaciones, las organizaciones pueden mejorar la eficiencia de sus esfuerzos promocionales. Además, la IA puede determinar qué facetas de una empresa, incluida su cultura y sus logros, deben destacarse a los posibles candidatos para obtener las respuestas más favorables. (Chen, 2023, p. 140)

Las descripciones de los puestos de trabajo desempeñan un papel fundamental a la hora de comunicar las necesidades de una empresa a los posibles empleados. Sin embargo, los sesgos presentes en estas descripciones pueden impedir que las personas calificadas presenten su solicitud. Estos sesgos suelen girar en torno a la edad, el género o las cualificaciones educativas (Ongig, 2024).

Curiosamente, los artículos sugieren que la IA tiene el potencial de reducir algunos de estos problemas al revisar las descripciones de los puestos de trabajo e identificar el lenguaje que puede contribuir al sesgo. Aunque las descripciones de trabajo generadas por la IA pueden mostrar menos sesgos que las creadas por los humanos, persiste el riesgo de perpetuar los prejuicios existentes si las herramientas de IA no se diseñan meticulosamente (CareerExperts, 2023). Por ejemplo, puede surgir un sesgo de anclaje de las decisiones de contratación anteriores de un gerente, donde las características y competencias de los mejores empleados actuales influyen en la creación de nuevas descripciones de trabajo (Chen, 2023, p. 145).

Además, la sugerencia de la IA de utilizar ciertos términos en los anuncios de empleo, como "ambicioso" o "líder seguro", puede favorecer o disuadir involuntariamente a grupos demográficos específicos, introduciendo así predisposición (Lawton, 2022). Si Los sistemas de IA se entrenan utilizando

descripciones de puestos anteriores, corren el riesgo de replicar el lenguaje y los requisitos que históricamente han atraído a un grupo menos diverso de solicitantes, lo que podría continuar esta tendencia.

Sesgos en la búsqueda de talento

Se emplean varias metodologías de IA durante la fase inicial de interacción con los posibles candidatos. Los algoritmos facilitan la comunicación personalizada a través de plataformas digitales y redes sociales, mientras que los bots de IA se utilizan para identificar a los solicitantes de empleo activos y pasivos en varias plataformas, o incluso entre la base de datos de solicitantes anteriores de una empresa. Además, la IA puede desplegar estrategias de minería de textos para mejorar el atractivo de las ofertas de empleo incorporando frases que atraigan a la mayor cantidad de candidatos. Además, la implementación de una métrica de tono por parte de la IA puede proporcionar recomendaciones para refinar las descripciones de los puestos, haciéndolos más inclusivos (Hunkenschroer y Luetge, 2022, p. 992).

Los anuncios de empleo impulsados por IA conllevan inherentemente el riesgo de discriminación indirecta. Esto se debe a que los empleadores pueden especificar la audiencia de sus anuncios utilizando criterios como las habilidades lingüísticas, la formación académica, la experiencia profesional o incluso la edad y el género, lo que puede impedir que ciertos grupos vean estas oportunidades laborales, lo que contribuye a la homogeneidad en el lugar de trabajo (Chen, 2023, p. 144; Sheard, 2022, p. 624). Por ejemplo, la decisión de Verizon de dirigir un anuncio de Facebook a personas de entre 25 y 36 años, que viven o han visitado la capital de EE.UU., con interés en las finanzas, podría excluir inadvertidamente a otros posibles solicitantes de empleo del alcance del anuncio (Sheard, 2022, p. 624). Los algoritmos diseñados para mejorar la eficiencia de la adición de empleo tienden a centrarse en los grupos demográficos que históricamente interactúan más con estos anuncios, lo que reduce la diversidad del grupo de solicitantes con el tiempo (Köchling y Wehner, 2020).

Los sesgos de contratación se ven reforzados por las preferencias de los reclutadores por plataformas publicitarias específicas y su tendencia a interactuar más con los candidatos de esas plataformas. Esto puede conducir a una falta de diversidad, ya que las diferentes plataformas atraen a diferentes tipos de usuarios (Lawton, 2022). El sesgo en la contratación no es únicamente el resultado de algoritmos o prácticas de los empleadores; También se deriva de los comportamientos de los propios solicitantes. Confiar en las redes sociales u otras actividades en línea para identificar candidatos potenciales puede sesgar el grupo hacia aquellos que están activos en línea o tienen los conocimientos necesarios para crear una presencia digital atractiva. Esto es particularmente evidente en industrias altamente competitivas como TI, donde los profesionales más ocupados pueden no actualizar regularmente sus perfiles de LinkedIn. Por el contrario, hay quienes podrían administrar estratégicamente sus perfiles en línea

para mejorar su atractivo, lo que plantea preguntas sobre la autenticidad de tales personas seleccionadas digitalmente. (Naka, 2018, págs. 46-47)

La cuestión de la transparencia es una preocupación ética importante en la contratación asistida por IA. La complejidad de los algoritmos puede dar lugar a un efecto de "caja negra", en el que el razonamiento de las decisiones sigue siendo oscuro para los solicitantes (DigitalOcean, s.f.). Además, existen cuestiones éticas sobre el uso de los perfiles y las actividades de los candidatos en las redes sociales y diversas plataformas con fines de contratación. Los estándares para el manejo de dichos datos varían a nivel internacional, siendo el Reglamento General de Protección de Datos (GDPR) europeo uno de los más estrictos (Hunkenschroer y Luetge, 2022, p. 995).

Sesgos en la selección y preselección de candidatos

Las tecnologías de IA desempeñan un papel crucial en la selección inicial de los candidatos, ya que dan prioridad a aquellos que parecen más adecuados para el puesto. Históricamente, las empresas dependían del escaneo de currículos en busca de palabras clave específicas. Sin embargo, los enfoques modernos son más sofisticados, incluidos los chatbots y las herramientas de análisis de currículums que buscan correlaciones semánticas y evalúan otras calificaciones. Algunas herramientas incluso predicen el rendimiento futuro potencial de un candidato buscando indicadores de compromiso o productividad a largo plazo, así como la ausencia de marcadores negativos como la impuntualidad habitual o los problemas disciplinarios (Hunkenschroer y Luetge, 2022, p. 992).

En el proceso de preselección de candidatos, la IA también se emplea en el análisis de entrevistas en vídeo. Los asistentes de IA realizan estas entrevistas, haciendo una lista establecida de preguntas y evaluando no solo las respuestas, sino también analizando el tono de voz, las expresiones microfaciales y las respuestas emocionales del candidato para inferir rasgos de personalidad. Además, los videojuegos impulsados por IA se utilizan para evaluar rasgos como el comportamiento de toma de riesgos, las habilidades de planificación, la perseverancia y la motivación (Hunkenschroer y Luetge, 2022, p. 992).

Además, las soluciones de IA analizan las huellas digitales de los candidatos, incluidas las actividades en las redes sociales y otros comportamientos en línea, a través del análisis lingüístico para medir su compatibilidad con la cultura de la empresa (Hunkenschroer y Luetge, 2022, p. 992). Más allá del escrutinio de las redes sociales, las técnicas de programación neurolingüística (PNL) se aplican en varias etapas de reclutamiento, incluido el análisis del currículum, para proporcionar una visión más profunda (Albaroudi et al., 2024, p. 391).

La IA promete fomentar un entorno de contratación más justo al reducir los sesgos inconscientes. Al hacer hincapié en las habilidades y credenciales predefinidas, las soluciones impulsadas por la IA pueden pasar por alto detalles demográficos como la edad, el sexo y la etnia, lo que podría influir

involuntariamente en el juicio de un reclutador. Este enfoque tiene el potencial de mejorar la diversidad en el lugar de trabajo, ya que garantiza que los candidatos sean evaluados en función de sus calificaciones y logros. Además, ciertas aplicaciones de IA se desarrollan específicamente para ayudar a las empresas a alcanzar sus objetivos de diversidad e inclusión. Lo logran examinando las tendencias de contratación y proponiendo ajustes para abordar las disparidades. (DigitalOcean, s.f.)

A pesar del potencial de los procesos de selección para reducir ciertos sesgos, pueden introducir otros inadvertidamente si no se manejan meticulosamente. El sesgo puede provenir de los principios detrás del diseño del modelo, la selección de características y los datos utilizados para el entrenamiento (Hunkenschroer y Luetge, 2022, p. 994).

Cuando los datos de capacitación se basan en un grupo de empleados existentes o en un grupo estrechamente definido que no representa proporcionalmente a poblaciones diversas, puede dar lugar a una discriminación involuntaria contra grupos infrarrepresentados (Hunkenschroer y Luetge, 2022, p. 994). Los sistemas de IA, que aprenden de datos que ya pueden contener sesgos históricos, tienen el potencial de amplificar aún más estos sesgos. Por ejemplo, si un sistema de IA se entrena predominantemente con currículos de un grupo demográfico específico, puede favorecer injustamente a los candidatos de ese grupo (Vivek, 2023). Un caso notable ocurrió en 2018 cuando el sistema de contratación de IA de Amazon mostró una preferencia por los patrones dominantes entre los solicitantes masculinos, discriminando así a las candidatas femeninas. Esto se debió a que los datos de entrenamiento de Amazon consistían principalmente en currículos de los mejores artistas, la mayoría de los cuales eran hombres, lo que llevó al algoritmo a penalizar las características típicamente asociadas con las mujeres (Albaroudi et al., 2024, p. 386; Hunkenschroer y Luetge, 2022, p. 994).

Además, el sesgo de formación también puede surgir cuando una herramienta algorítmica no tiene en cuenta adecuadamente todas las habilidades y rasgos relevantes para el trabajo. Si la herramienta se centra demasiado en las habilidades técnicas sin tener en cuenta las habilidades interpersonales, es posible que se pasen por alto competencias cruciales como la comunicación. También puede producirse un sesgo de agregación, en el que se hacen falsas generalizaciones sobre poblaciones enteras, lo que lleva a la exclusión de individuos basándose en suposiciones sobre las características del grupo. Por ejemplo, los algoritmos podrían suponer que las personas más jóvenes tienen habilidades tecnológicas superiores, pasando por alto a los candidatos mayores que están igualmente calificados (Albaroudi et al., 2024, p. 386).

La evaluación de la huella digital de un candidato como parte de su selección también puede perpetuar los sesgos, reflejando las diferentes percepciones de lo que debe presentarse en las redes sociales y las limitaciones a las que se enfrentan las personas para actualizar sus perfiles. Un estudio de la Universidad Estatal de Carolina del Norte en el que se entrevistó a 61 profesionales de RR.HH. reveló una tendencia a valorar los intereses personales, como el senderismo o la

celebración de la Navidad, por encima de los rasgos profesionales. Este enfoque beneficia a ciertos grupos demográficos sobre otros y asume que características como ser "enérgico" o "activo" son inherentemente superiores, lo que puede discriminar a las personas mayores o discapacitadas. La introducción de la IA en este proceso corre el riesgo de incrustar estos prejuicios en los algoritmos (Shipman, 2021).

El uso de software de reconocimiento emocional en las evaluaciones de video debe considerar cuidadosamente las diversas entonaciones en todos los idiomas y el potencial de desventaja sistemática para razas o grupos étnicos específicos. Además, a algunas personas les puede resultar difícil comportarse de forma natural frente a una cámara, lo que lleva a evaluaciones inexactas de la herramienta (Hunkenschroer y Luetge, 2022, p. 995).

La privacidad y el consentimiento informado en el uso de la información de los candidatos para la contratación son consideraciones éticas críticas, con regulaciones que varían a nivel internacional. El Reglamento General de Protección de Datos (RGPD) europeo es uno de los marcos más estrictos, diseñado para salvaguardar los derechos de los ciudadanos de la UE al regular la recopilación, el almacenamiento y el procesamiento de datos personales y exigir el consentimiento informado para cualquier actividad de procesamiento de datos personales. Sin embargo, el dilema surge al considerar si los solicitantes correrían el riesgo de optar por no participar, por temor a que pudiera poner en peligro sus perspectivas laborales (Hunkenschroer y Luetge, 2022, p. 995).

Sesgos en las entrevistas a los candidatos

Una de las razones para utilizar las técnicas de entrevista basadas en IA puede ser las barreras lingüísticas en la contratación multinacional, que requiere la adquisición de talento local y profesionales de RRHH multilingües. La IA mitiga esto al permitir la entrevista de hablantes de diversos idiomas sin la necesidad de que RR.HH. conozca todos los dialectos locales, gracias a los avances en el procesamiento del lenguaje natural (NLP). Esta tecnología permite a las empresas cerrar de manera eficiente la brecha lingüística en la contratación global sin depender de traductores. (Albaroudi et al., 2024, p. 392)

Sin embargo, las herramientas de entrevista basadas en IA se utilizan con mayor frecuencia para automatizar la primera ronda de entrevistas y, por lo tanto, contribuir a la evaluación de los candidatos (Fritts y Cabrera, 2021, p. 792). Por lo tanto, las soluciones de entrevista basadas en IA y los posibles sesgos que están creando, se manejan en el capítulo anterior "Selección y preselección de los candidatos".

Vivek (2023) sugiere que, debido a la comprensión matizada y la inteligencia emocional que poseen los seres humanos, el poder final de toma de decisiones debe permanecer en los reclutadores humanos. Fritts & Cabrera (2021) mencionar también que las preocupaciones en torno a los procesos impulsados por la IA

para identificar, seleccionar, entrevistar e incorporar candidatos podrían conducir a la deshumanización al eliminar las interacciones personales. Esto podría ser éticamente problemático, ya que los valores artificiales incorporados en los algoritmos de contratación podrían socavar la relación empleado-empleador al eliminar los aspectos inherentemente humanos de las interacciones tradicionales.

Sesgos en la evaluación y la creación de ofertas

Después de elegir un candidato, la empresa debe extender una oferta de trabajo. Los algoritmos de IA pueden evaluar las posiciones pasadas del candidato para formular una oferta que el candidato probablemente acepte. Sin embargo, estas herramientas pueden exacerbar las desigualdades existentes en los salarios iniciales y continuos entre diferentes géneros, razas y otros grupos demográficos. Esto sucede porque estas herramientas a menudo se basan en datos salariales históricos, que pueden contener sesgos inherentes. (Lawton, 2022)

Sesgos en el *onboarding*

La implementación de la IA en la incorporación puede reducir las tareas rutinarias de los profesionales de RRHH mediante la automatización de la gestión de documentos, la programación de comunicaciones y el uso de *chatbots* para los comentarios y consultas iniciales. También evalúa el progreso de los nuevos empleados y garantiza el cumplimiento de los requisitos de formación, al tiempo que proporciona análisis sobre las tendencias de incorporación. (Marr, 2023)

Las interfaces de IA pueden adaptar las rutas educativas para que se ajusten a las habilidades individuales y los estilos de aprendizaje, lo que garantiza que los nuevos empleados reciban el apoyo que necesitan para tener éxito. Los departamentos de RRHH pueden utilizar los conocimientos de las interacciones automatizadas y los comentarios para comprender las capacidades y las áreas de crecimiento de los nuevos empleados. El análisis predictivo puede anticiparse a las necesidades de los nuevos empleados, ofreciendo recursos y soporte proactivos. (Marr, 2023). La IA también se puede utilizar para hacer coincidir al recién llegado con el mentor laboral en función de sus habilidades, intereses y experiencia. (Destacado, 2023)

Al igual que en las fases anteriores, los sesgos en la incorporación impulsada por la IA pueden surgir del uso de datos de entrenamiento basados en registros históricos que conllevan prejuicios anteriores. Además, los algoritmos podrían incorporar inadvertidamente los sesgos subconscientes de los desarrolladores a la hora de dar forma a las trayectorias educativas.

6. Selección y preselección de los candidatos: definición del conjunto de datos

Selección y preselección de los Candidatos; Definición de conjunto de datos

Si bien la empresa está estudiando todos los aspectos y posibilidades en la contratación de IA, actualmente la empresa está evaluando intensamente la IA en la fase de selección y preselección. Esto implica un análisis en profundidad del escaneo de currículums y la alineación de los candidatos con las vacantes de trabajo adecuadas.

Mediante la implementación de herramientas de IA para automatizar el proceso de selección de currículums, el objetivo es identificar de forma rápida y precisa a los candidatos que cumplen los requisitos del puesto. Se espera que la adopción de dicha tecnología conduzca a un uso más estratégico de los recursos de recursos humanos, mejore el calibre de los candidatos que llegan a la lista de finalistas y, al alinear las habilidades y antecedentes de los candidatos con los requisitos de la empresa, aumente la efectividad general del proceso de contratación.

En primer lugar, la empresa debe definir los conjuntos de datos, que son la base del sistema de contratación y de los que los algoritmos toman los datos.

Entrenamiento de IA

Los conjuntos de datos de entrenamiento desempeñan un papel crucial en el desarrollo de algoritmos. Estos conjuntos de datos son la base sobre la que se construyen los algoritmos. El propósito de estos conjuntos de datos es servir como la "verdad fundamental" o base fáctica que instruye al algoritmo sobre cómo interpretar y procesar grandes cantidades de datos. Enseñan al algoritmo a comprender las relaciones entre las variables de entrada (la información introducida en el sistema) y una variable de salida (la predicción o decisión que toma el sistema en función de las entradas). Un desafío significativo surge cuando los propios datos de entrenamiento contienen sesgos. Estos sesgos pueden ser el reflejo de desigualdades históricas, prejuicios o simplemente el resultado de una recopilación de datos no representativa o incompleta. Cuando un algoritmo se entrena con datos sesgados, aprende estos sesgos y los perpetúa en sus predicciones y decisiones. Este fenómeno se describe comúnmente como el problema del "sesgo hacia adentro, sesgo hacia afuera". Esto significa que si los datos de entrada (sesgo de entrada) están sesgados, la salida del modelo de aprendizaje automático (sesgo de salida) también estará sesgada. Esto puede dar lugar a resultados injustos, discriminatorios o defectuosos que afecten a los candidatos que se seleccionan o rechazan en función de criterios sesgados en lugar de sus verdaderas cualificaciones. potencial. (Sheard, 2022)

Entrenar la IA con los datos internos de la empresa

Al entrenar aplicaciones de IA para la detección de CV, las empresas tienen acceso a una variedad de fuentes de datos internas que pueden ser fundamentales en el proceso de aprendizaje. (Fernandes, 2021; Ma, 2024; Sheard, 2022)

Una de esas fuentes son los datos de rendimiento de los empleados junto con los datos históricos de empleo, que abarcan los registros internos de las evaluaciones de los empleados y los resultados de las contrataciones anteriores. Además, las empresas pueden utilizar la acumulación de CV enviados a lo largo del tiempo, que incluye una gran cantidad de CV históricos y currículos que se recibieron directamente para las solicitudes de empleo.

Otro recurso valioso son los datos de "casos falsos negativos", que se refieren a la información sobre candidatos que fueron rechazados por error en ciclos de contratación anteriores. El análisis de estos casos puede proporcionar información sobre patrones que la IA debe evitar repetir. Además, las empresas pueden extraer palabras clave y frases específicas que se encuentran con frecuencia en los CV recopilados, utilizándolas para refinar la capacidad de la IA para identificar calificaciones y experiencia relevantes.

Por último, la información de antecedentes educativos, que es un componente estándar de las solicitudes de los candidatos, también se puede aprovechar para

mejorar la comprensión de la IA sobre las calificaciones académicas y su relevancia para los puestos de trabajo.

Entrenar la IA con los datos externos de la empresa

Se han sugerido posibles fuentes de datos externos en el entrenamiento de IA e incluyen varias plataformas y métodos para recopilar información. (Banerjee, 2022) Las plataformas de redes sociales son una rica fuente en la que las personas comparten intereses personales, experiencias y logros profesionales, proporcionando información sobre la personalidad, las habilidades y la red profesional de un candidato. (Albaroudi et al., 2024; Albassam, 2023; Chaker, 2018; Filtrado, s.f.; Heymans, 2022; Shipman, 2021) Además, los sitios web profesionales diseñados para la creación de redes y el desarrollo profesional, como LinkedIn, contienen perfiles detallados que incluyen historial laboral, educación, habilidades, avales y logros profesionales. Estos perfiles ofrecen una visión integral de los talentos potenciales. Por último, las búsquedas en Internet realizadas por las empresas pueden desenterrar información disponible públicamente sobre un candidato a partir de fuentes como artículos de noticias, blogs personales, publicaciones u otras presencias en la web. Esta amplia gama de información proporciona un contexto valioso sobre las calificaciones, los logros y el carácter de una persona. (Chaker, 2018)

7. Desarrollo del modelo algorítmico

Anuncio de empleo; Selección de palabras clave para el algoritmo

La empresa ha publicado recientemente una descripción del puesto de trabajo y ha recibido cientos de solicitudes en respuesta.

El prototipo de selección y preselección de IA escaneará los CV de los solicitantes en busca de palabras clave y otra información que se considere relevante para las contrataciones exitosas, como experiencia, títulos laborales, empleadores anteriores, universidades y títulos. Sobre la base de esto, el sistema crea un perfil de solicitante estructurado y todos los candidatos son calificados y clasificados (Sheard, 2022).

La IA agiliza significativamente el proceso de selección de candidatos, que tradicionalmente es uno de los pasos que más tiempo requiere en la contratación. Durante la etapa de "selección" del embudo de reclutamiento, la evaluación de los candidatos a un puesto de trabajo en función de su experiencia, habilidades y otras características pertinentes es crucial para crear una lista de candidatos a la entrevista. La IA se puede utilizar para escanear los CV en busca de palabras clave y otros datos asociados con las contrataciones exitosas, como cargos, empleadores anteriores, instituciones educativas y calificaciones, lo que agiliza el proceso de identificación de candidatos adecuados. (Dennis, 2023; Fernandes, 2021; Sheard, 2022)

Al utilizar herramientas de IA, los reclutadores pueden identificar y priorizar rápidamente a los mejores candidatos, mejorando así la eficiencia y mejorando las métricas clave de contratación, como el tiempo de contratación. Los sistemas de IA seleccionan automáticamente a los candidatos que no cumplen con los criterios específicos preestablecidos, analizan los currículos para resaltar la experiencia relevante y clasifican a los solicitantes según su idoneidad para el puesto. Esta automatización permite a los reclutadores centrarse más en la evaluación de los candidatos en las últimas etapas del embudo de contratación. Además, la IA puede escanear los currículos durante la fase inicial de selección de currículums, identificando a los candidatos que cumplen con criterios específicos. A continuación, los algoritmos de aprendizaje automático evalúan las habilidades y cualificaciones de estos candidatos en función de las especificaciones del puesto para compilar una lista de candidatos adecuados. (Hoja de vida, 2023; Dennis, 2023)

La descripción del puesto proporciona detalles esenciales sobre el puesto, incluidas las habilidades, cualificaciones y experiencia requeridas, que la IA utiliza para filtrar los CV. Sin una descripción detallada del puesto, la IA carecería de los parámetros necesarios para saber qué buscar en los CV.

Job description for Communication Assistant
We are looking for a communications assistant to be responsible for the creation of content such as media releases, blogs, and social media posts on behalf of our company. You will also be monitoring media and campaign coverage and attending internal and external events.

Responsibilities
- Develop and manage diverse content, including media releases, blogs, and social media posts. aligned with the company's strategic goals; monitor media and campaign coverage.
- Support the implementation of internal and external communications strategies and assist in managing the company's image.
- Organize marketing events and provide comprehensive administrative support. including maintaining event calendars and updating contact lists.
- Compile and prepare presentations and reports while tracking project progress and media exposure.

Requirements
- Holds a Bachelor's degree in communications. marketing, or related field, with excellent verbal and written communication abilities.
- Proficient in social media strategies and media relations, with a creative and innovative approach.
- Strong organizational skills and attention to detail, capable of multitasking and maintaining positive interpersonal relationships.
- Skilled in using office management and design software, such as Photoshop and InDesign, and knowledgeable in various social media platforms.

Los avances en IA han ido más allá de la simple coincidencia de palabras clave. Inicialmente, las empresas utilizaban algoritmos para escanear los currículos en busca de palabras clave o frases preseleccionadas. Las tecnologías de IA actuales, incluidos los chatbots y las sofisticadas herramientas de análisis de currículums, ahora evalúan la relevancia semántica y los términos relacionados para determinar con mayor precisión las calificaciones de un candidato. Además, los algoritmos de aprendizaje automático se emplean para predecir el rendimiento laboral futuro en función de varios indicadores, como la antigüedad en el trabajo y la productividad. Estas herramientas también pueden sugerir las ofertas de trabajo más adecuadas para los candidatos en función de sus perfiles. (Hunkenschroer y Luetge, 2022)

Sin embargo, en este curso, utilizamos una simple concordancia de palabras clave como ejemplo para ilustrar cómo una aplicación de cribado de CV con IA procesa las solicitudes.

Lista de palabras clave específicas utilizadas para la selección de candidatos

- "Licenciatura en comunicación"
- "Grado en marketing" "Creación de contenidos"
- "Comunicados de prensa"
- "Blogs"
- "Publicaciones en redes sociales"
- "Monitoreo de medios" y "cobertura de campañas"
- "Apoyo a la implementación de estrategias de comunicación"
- "Gestión de la imagen de la empresa"
- "Organización de eventos de marketing"
- "Brindar soporte administrativo integral" "Actualización de listas de contactos"
- "Compilación de presentaciones" e "informes", "seguimiento del progreso del proyecto" y "exposición a los medios"
- "Competente en estrategias de redes sociales" y "relaciones con los medios"
- "Enfoque creativo e innovador"
- "Fuerte capacidad de organización" y "atención al detalle"
- "Capaz de realizar múltiples tareas"
- "Mantener relaciones interpersonales positivas", "Hábil en el uso de la gestión de oficinas" y "Diseño de software"
- "Photoshop" e "InDesign"
- "Conocedor de varias plataformas de redes sociales"

Según Sheard (2022)Por lo general, los currículos son revisados por ojos humanos solo si se alinean con los criterios de los anuncios de empleo, mientras que el resto puede ignorarse, lo que pone de relieve una posible preocupación en la que una parte significativa de los currículos podría excluirse prematuramente.

A continuación, nos centraremos en el análisis de dos aplicaciones para entender cómo el prototipo de aplicación de IA evalúa a los candidatos durante la fase de selección. Este análisis ayudará a identificar si hay algún sesgo que influya en el proceso de selección.

Comparar los CV de dos candidatos con la frase clave

La empresa organiza un taller para explorar los detalles de la detección de CV impulsada por IA, utilizando un estudio de caso de sus propias operaciones.

Durante el taller, analizan dos CV uno al lado del otro para descubrir cómo las diferencias en la redacción y la redacción pueden llevar inadvertidamente a pasar por alto a los candidatos calificados.

JOHN
DOE

Communications assistant

About Me

A dedicated communications professional with a Bachelor's degree in Marketing, seeking the role of Communications Assistant to leverage extensive experience in content creation, social media management, and event coordination to contribute to the company's strategic goals.

+123-456-7890
hello@doeland.com
123 Anywhere St., Any City

EXPERIENCE

Communications assitant
ABC Corporation
2019- Present

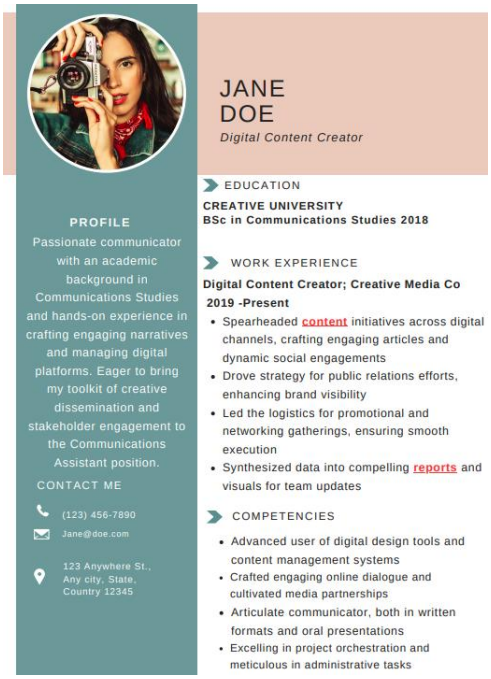
- Developed and managed diverse content, including media releases, blogs, and social media posts
- Supported the implementation of comprehensive communications strategies
- Organized multiple marketing events and maintained event calendars
- Compiled presentations and tracked project progress, ensuring alignment with strategic goals

EDUCATION

University of communications
Bachelor's degree in Marketing
2014-2018

SKILLS SUMMARY

- Proficient in Photoshop and InDesign
- Skilled in social media strategies and media relations
- Excellent verbal and written communication abilities
- Strong organizational skills and attention to detail



JANE
DOE

Digital Content Creator

PROFILE

Passionate communicator with an academic background in Communications Studies and hands-on experience in crafting engaging narratives and managing digital platforms. Eager to bring my toolkit of creative dissemination and stakeholder engagement to the Communications Assistant position.

CONTACT ME

(123) 456-7890
Jane@doe.com
123 Anywhere St., Any city, State, Country 12345

EDUCATION

CREATIVE UNIVERSITY
BSc in Communications Studies 2018

WORK EXPERIENCE

Digital Content Creator; Creative Media Co
2019 -Present

- Spearheaded content initiatives across digital channels, crafting engaging articles and dynamic social engagements
- Drove strategy for public relations efforts, enhancing brand visibility
- Led the logistics for promotional and networking gatherings, ensuring smooth execution
- Synthesized data into compelling reports and visuals for team updates

COMPETENCIES

- Advanced user of digital design tools and content management systems
- Crafted engaging online dialogue and cultivated media partnerships
- Articulate communicator, both in written formats and oral presentations
- Excelling in project orchestration and meticulous in administrative tasks

Este ejemplo ilustra el impacto de las diferencias de redacción y redacción en dos CV y cómo pueden afectar al resultado de un proceso de selección de IA. A pesar de poseer calificaciones relevantes, el segundo CV puede pasarse por alto debido a que sus descripciones no se alinean estrechamente con las palabras clave específicas establecidas por la empresa para la selección. Esto demuestra la importancia de hacer coincidir el idioma de un CV con las palabras clave esperadas por un sistema de IA.

Aspect	CV1	CV 2
Job Titles and Terminology	Uses standard terms like "Communications Assistant"	Uses titles like "Digital Content Creator"
Academic Background	Explicitly mentions "Bachelor's Degree in Marketing"	States "BSc in Communications Studies"
Describing Tasks and Skills	Direct match with keywords like "Content creation", "Media releases"	Uses varied language like "crafting engaging narratives", "spearheaded content initiatives"
Technical and Software Skills	Specifies "Photoshop" and "InDesign"	General mention of "digital design tools and content management systems"
Direct Keyword Matches	Closely matches many specified keywords	Uses different phrasing that may not match the specific Keywords set for screening

8. Cómo medir la equidad en las soluciones de aprendizaje automático

Probar la equidad de la solución

La evaluación del prototipo de algoritmo reveló inconsistencias en el análisis de CV, lo que provocó refinamientos para una comprensión más completa de los atributos de CV. El modelo ahora está preparado para una fase de prueba antes de la implementación en el mundo real.

Si bien varios aspectos requieren evaluación, la atención se centrará aquí en probar la equidad de la aplicación para determinar el alcance de los sesgos incurridos durante el proceso. Las métricas de equidad se emplean para medir el nivel de sesgo presente en las decisiones de la aplicación.

Para crear sistemas que sean equitativos para todos y libres de sesgos como la discriminación racial o de género, es importante evaluar el rendimiento de la solución en diversos segmentos de la población. Si bien existen varios métodos para evaluar la equidad de una solución, en este contexto, nos centraremos exclusivamente en las pruebas de clasificación binaria. (*Aprendizaje automático y equidad*, 2021)

La equidad se ha convertido en un tema crucial en el aprendizaje automático debido a su creciente aplicación en diversos campos. Una amplia investigación se ha centrado en esta área porque los sistemas de aprendizaje automático dependen en gran medida de los datos, que, cuando se obtienen de humanos, pueden ser inherentemente sesgados. Definir la equidad en términos matemáticos ha demostrado ser un desafío, lo que ha llevado a una falta de consenso sobre las formulaciones estándar. Sin embargo, como señaló Zhong (2018), la mayoría de las definiciones de equidad se agrupan en torno a varios conceptos clave:

- Desconocimiento
- Paridad demográfica
- Cuotas igualadas
- Paridad de tasa predictiva
- Equidad individual
- Equidad contrafáctica

La paridad demográfica, las cuotas igualadas y la paridad de tasas predictivas generalmente se agrupan bajo el paraguas de la "equidad de grupo". (Zhong, 2018). Este concepto se explorará más a fondo en esta sección, particularmente a través de la lente de la IA en el reclutamiento.

Además de la equidad de grupo, Castelnovo et al. (2022) dividir las categorías de equidad en equidad individual y equidad basada en la causalidad, incluyendo también los temas mencionados por las definiciones de Zhong (2018).

Matriz de confusión

Uno de los métodos de clasificación binaria es La matriz de la confusión (*Aprendizaje automático y equidad*, 2021). Para Para comprender las métricas de equidad, comenzamos con el concepto de una matriz de confusión. Esta herramienta resume las predicciones realizadas por un modelo en comparación con los resultados reales, o la verdad sobre el terreno, en los que se entrenó. La matriz de confusión muestra los recuentos de predicciones correctas e incorrectas, lo que proporciona una visión clara del rendimiento del modelo y su validez explicable. (Saplicki, 2022)

	Unqualified	Qualified
Reject	True Negative (TN)	False Negative (FN)
Hire	False Positive (FP)	True Positive (TP)

FUENTE DE LA IMAGEN | Imagen 1 - Matriz de confusión (Modificado de (Machine Learning and Fairness, 2021)

Una matriz de confusión nos ayuda a categorizar los resultados del proceso de contratación. Los candidatos calificados que son contratados con éxito se consideran verdaderos positivos (TP) y los candidatos no calificados rechazados adecuadamente son verdaderos negativos (TN). Por el contrario, los falsos positivos (FP) se refieren a candidatos no calificados que son contratados por error, mientras que los falsos negativos (FN) son candidatos calificados rechazados erróneamente. Una matriz de confusión simplemente categoriza a los solicitantes en distintos grupos en función de sus resultados de clasificación. Los cuatro valores producidos por una matriz de confusión se utilizan para calcular las métricas estándar de aprendizaje automático, incluida la exactitud y la precisión, en diferentes grupos demográficos. (*Aprendizaje automático y equidad*, 2021; Teodorescu, 2020)

En el escenario de contratación que se está discutiendo, nos centraremos específicamente en la demografía de género. Aunque es importante tener en cuenta otros factores demográficos, este ejemplo se centrará únicamente en el género.

Supongamos que hay un número igual de 100 solicitantes masculinos y femeninos. El sistema ha identificado con precisión a 75 personas de cada grupo como calificadas o no calificadas, lo que demuestra que la tasa de precisión es consistente en ambos grupos demográficos. (*Aprendizaje automático y equidad*, 2021)

MEN	Unqualified	Qualified	WOMEN	Unqualified	Qualified
Reject	**15 (TN)**	5 (FN)	Reject	**60 (TN)**	20 (FN)
Hire	20 (FP)	**60 (TP)**	Hire	5 (FP)	**15 (TP)**

FUENTE DE LA IMAGEN | Imagen 2 - Ejemplo de matriz de confusión de reclutamiento (Modificado de Machine Learning and Fairness, 2021)

Paridad demográfica

Sobre la base de los datos proporcionados, examinaremos cómo se aplican varias métricas de equidad en este contexto, comenzando con la paridad demográfica. Esta métrica solo tiene en cuenta la salida del clasificador, sin tener en cuenta las etiquetas reales. La paridad demográfica es una de las métricas sugeridas a la hora de evaluar un sistema de contratación automatizado. (*Aprendizaje automático y equidad*, 2021)

La paridad demográfica, comúnmente conocida como independencia o paridad estadística, es un criterio de equidad en el que la probabilidad de un resultado favorable, como ser contratado, no se ve influenciada por una característica protegida como el género (Zhong, 2018).

MEN	Unqualified	Qualified
Reject	15 (TN)	5 (FN)
Hire	**20 (FP)**	**60 (TP)**

WOMEN	Unqualified	Qualified
Reject	60 (TN)	20 (FN)
Hire	**5 (FP)**	**15 (TP)**

FUENTE DE LA IMAGEN | Imagen 3 - Ejemplo de paridad demográfica de reclutamiento (Modificado de Machine Learning and Fairness, 2021)

En el escenario de ejemplo de reclutamiento (imagen 3), si se contrata a 80 de cada 100 solicitantes masculinos y solo a 20 de cada 100 solicitantes femeninas (imagen 3), no se cumple la paridad demográfica. Sin embargo, esto podría ser aceptable si los solicitantes masculinos están realmente más calificados que las solicitantes femeninas, como parece ser en este caso. (*Aprendizaje automático y equidad*, 2021).

Sin embargo, la paridad demográfica es difícil de lograr cuando las personas pertenecen a varios grupos protegidos, ya que es posible que las probabilidades iguales no sean factibles en todos los grupos demográficos. La paridad demográfica a veces puede favorecer inadvertidamente a las personas menos cualificadas debido a su enfoque a nivel de grupo, lo que puede conducir a la injusticia a nivel individual. Por ejemplo, aumentar la proporción de candidatas menos calificadas podría disminuir las posibilidades de contratar a hombres calificados. (Teodorescu, 2020) Por lo tanto, esta medida solo considera el resultado, no las etiquetas verdaderas, lo que significa que no tiene en cuenta las calificaciones reales de los solicitantes. Sostiene que la proporción de solicitantes contratados con respecto al total debe ser similar en todos. (*Aprendizaje automático y equidad*, 2021).

Paridad predictiva

La suficiencia, vista desde la perspectiva de aquellos que reciben la misma decisión de un modelo, exige tasas de error iguales independientemente de los atributos sensibles. Este concepto se engloba en la paridad predictiva, también conocida como prueba de resultados. Garantiza que las tasas de error estén equilibradas para los individuos agrupados por las decisiones que reciben, en lugar de por los resultados reales. A diferencia de la paridad demográfica, la paridad predictiva tiene en cuenta tanto las decisiones del clasificador como los resultados reales. Evalúa si la probabilidad de que un candidato sea adecuado para un puesto es coherente en los diferentes grupos, siempre que la IA lo haya seleccionado para su contratación. Se espera que esta probabilidad sea casi idéntica para todos los grupos considerados. (Castelnovo et al., 2022; *Aprendizaje automático y equidad*, 2021).

MEN	Unqualified	Qualified	WOMEN	Unqualified	Qualified
Reject	15	5	Reject	60	20
Hire	20	**60**	Hire	5	**15**

FUENTE DE LA IMAGEN | Imagen 4 - Ejemplo de paridad predictiva de reclutamiento (Modificado de *Aprendizaje automático y equidad*, 2021)

En nuestro ejemplo de reclutamiento, el sistema cumple con los criterios de paridad predictiva, ya que tres cuartas partes de los solicitantes contratados de ambos grupos están calificados (*Aprendizaje automático y equidad*, 2021).

Saldo de tasas de falsos positivos, saldo de tasas de falsos negativos y cuotas igualadas

La métrica conocida como **Tasa de falsos positivos** afirma que la probabilidad de que un candidato no calificado sea contratado erróneamente debe ser consistente en todos los grupos demográficos. En términos más simples, el sistema debería juzgar erróneamente a los candidatos no calificados, independientemente de su género. (*Aprendizaje automático y equidad*, 2021)

False positive rate

MEN	Unqualified	Qualified
Reject	15	5
Hire	**20**	60

WOMEN	Unqualified	Qualified
Reject	60	20
Hire	**5**	15

FUENTE DE LA IMAGEN | Imagen 5 - Ejemplo de tasa de falsos positivos de reclutamiento (Modificado de *Aprendizaje automático y equidad*, 2021)

En este ejemplo (imagen 5), las tasas de falsos positivos son diferentes, con una fracción sustancialmente mayor de hombres no calificados contratados que de mujeres no calificadas.

El saldo de la tasa de falsos negativos es esencialmente lo mismo que la tasa de falsos positivos, pero para las tasas de falsos negativos. El concepto de paridad de tasas de falsos negativos sugiere que la probabilidad de rechazar erróneamente a candidatos calificados debería ser uniforme en todos los grupos.

False negative rate

MEN	Unqualified	Qualified
Reject	15	**5**
Hire	20	60

WOMEN	Unqualified	Qualified
Reject	60	**20**
Hire	5	15

FUENTE DE LA IMAGEN | Imagen 6. Ejemplo de tasa de falsos negativos de reclutamiento (Modificado de *Aprendizaje automático y equidad*, 2021)

En este caso (imagen 6) hay una discrepancia notable en las tasas de falsos negativos, ya que las candidatas calificadas se enfrentan al rechazo a una tasa mayor que sus homólogos masculinos.

Cuotas igualadas combina estos dos atributos satisfaciendo tanto el equilibrio de tasas de falsos positivos como el equilibrio de tasas de falsos negativos, lo que significa que el sistema debe manejar ambos tipos de errores (candidatos no calificados aceptados incorrectamente y candidatos calificados rechazados incorrectamente) de manera equitativa en diferentes grupos, asegurando que todos los candidatos calificados de manera similar reciban un tratamiento consistente. (*Aprendizaje automático y equidad*, 2021). Las cuotas igualadas son difíciles de lograr, pero cuando se logran pueden considerarse como uno de los niveles más altos de equidad algorítmica. (Teodorescu, 2020)

Compensaciones con las métricas

Un sistema no puede cumplir simultáneamente la paridad predictiva, el balance de tasas de falsos positivos y el balance de tasas de falsos negativos debido a restricciones matemáticas. Si un sistema cumple con dos de estos criterios, inevitablemente se quedará corto en el tercero. (*Aprendizaje automático y equidad*, 2021)

Este teorema subraya las compensaciones inherentes entre los diferentes objetivos de equidad. A medida que nos esforzamos por ser justos, a menudo nos encontramos con objetivos contradictorios, y lograr el equilibrio perfecto en todas las dimensiones sigue siendo difícil de alcanzar.

En la práctica, los responsables de la toma de decisiones deben sopesar cuidadosamente estas compensaciones, considerar el contexto y tomar decisiones informadas basadas en las necesidades específicas de la aplicación. Si bien la perfección puede ser inalcanzable, los esfuerzos continuos hacia la equidad y la transparencia son esenciales en los sistemas algorítmicos de toma de decisiones.

Por último, es importante reconocer que numerosas facetas de la equidad eluden la cuantificación debido a su complejidad sociotécnica. En consecuencia, no todos los aspectos de la equidad pueden ser encapsulados por métricas. (*Aprendizaje automático y equidad*, 2021)

Equidad individual

La equidad individual difiere de los criterios de equidad basados en el grupo (las métricas discutidas anteriormente) al centrarse en el trato a los individuos. Este concepto afirma que los individuos con características similares deben recibir resultados comparables. Determinar una métrica para medir la similitud individual es complejo. Considere tres candidatos para el puesto de trabajo: A con

una licenciatura y un año de experiencia relacionada; B con una maestría y la misma cantidad de experiencia; y C con una maestría pero sin experiencia. Decidir si A se parece más a B o a C, y cuantificar esa similitud, es un desafío sin poder comparar su desempeño laboral, lo que no es posible si no se contrata a los tres. (Zhong, 2018)

9. Implementación del modelo

Informar a los solicitantes sobre el uso de la IA en selección de personal

Tras cuidadosas evaluaciones internas, la empresa está lista para lanzar su herramienta de selección selectiva en los puestos vacantes para recopilar los comentarios iniciales de los usuarios y solicitantes.

Es crucial que los solicitantes sean conscientes de que están interactuando con un sistema de contratación impulsado por la IA, lo que fomenta la confianza. La concienciación sobre las ventajas de las herramientas de IA debe establecerse desde el principio, lo que mejora la disposición de los solicitantes a utilizar tecnologías interactivas como los *chatbots*. Además, perfeccionar los procesos de contratación de IA y garantizar la claridad sobre el funcionamiento del sistema son clave para su adopción y éxito entre los candidatos.

Además, debe asegurarse de que haya un humano que supervise e intervenga cuando sea necesario. (Chen, 2023)

10. Operación y monitoreo

Después de la implementación

Si bien se realizaron evaluaciones internas de equidad antes del lanzamiento, la empresa también debe estar preparada para auditorías externas independientes.

También es ventajoso buscar continuamente comentarios después de la implementación y establecer canales de participación continuos para las partes interesadas. Esto implica mantener a los trabajadores informados e involucrados en consultas y procesos participativos durante toda la implementación del sistema de IA dentro de la organización. (HLEG IA fiable)

Auditoría de IA

Si bien las mediciones internas de equidad se realizan antes de la implementación, una auditoría externa también es algo para lo que la empresa debe estar preparada.

Según Ahmed (2020) Los auditores que evalúan las aplicaciones de IA deben priorizar dos aspectos: garantizar el cumplimiento de los derechos de los interesados y evaluar los riesgos tecnológicos en el aprendizaje automático y la ciberseguridad. La auditoría comienza con la definición de su alcance, objetivos y los riesgos relacionados con la IA para la organización, generalmente documentados en una matriz de riesgos y control dentro de los marcos establecidos.

Los riesgos clave incluyen la desalineación de las estrategias de TI con los objetivos comerciales, la gobernanza ineficaz y las estructuras de responsabilidad. En cuanto al cumplimiento, los auditores deben comprender los principios de privacidad de datos y sus implicaciones en los derechos de las personas en virtud de regulaciones como el GDPR (Ahmed, 2020).

Existen varias directrices para auditar la IA. A medida que la IA remodela los negocios y la sociedad, los auditores deben garantizar la preparación para abordar los desafíos emergentes y verificar la eficacia de los controles y la gobernanza en torno a la IA (Ahmed, 2020).

11. Conclusión

Esta unidad del curso se centró en examinar los sesgos algorítmicos en un contexto práctico. El tema se exploró a través de un escenario hipotético en el que una empresa internacional desarrolla una solución de IA para la contratación. A medida que las tecnologías de IA se integran en la contratación, ofrecen beneficios sustanciales al mejorar la eficiencia y potencialmente aumentar la diversidad en el lugar de trabajo al enfatizar las calificaciones sobre las características demográficas. Sin embargo, la implementación de estas tecnologías también plantea importantes desafíos éticos, particularmente en la identificación y mitigación de sesgos algorítmicos.

Los sistemas de IA pueden perpetuar inadvertidamente los sesgos existentes en sus datos de entrenamiento. Estos sesgos pueden reflejar disparidades históricas o desigualdades sociales, lo que conduce a prácticas discriminatorias en los procesos de contratación. Abordar estos sesgos no es solo un imperativo ético, sino también una necesidad legal, ya que el panorama regulatorio que rodea a la IA en la contratación está en continua evolución. Las organizaciones deben permanecer vigilantes y proactivas en el cumplimiento de estas normas para evitar la discriminación.

Para mitigar eficazmente estos desafíos, es crucial utilizar herramientas sofisticadas diseñadas específicamente para identificar y corregir sesgos en los sistemas de IA. El seguimiento periódico y las auditorías independientes son esenciales, ya que ayudan a las organizaciones a identificar los sesgos emergentes y a ajustar sus sistemas de IA en consecuencia.

Además, es fundamental mantener un equilibrio entre los procesos impulsados por la IA y la supervisión humana. Si bien la IA puede agilizar significativamente los procesos de contratación, el elemento humano sigue siendo indispensable para interpretar contextos que la IA podría juzgar erróneamente. La supervisión humana garantiza que la contratación siga siendo un proceso sensible e inclusivo, que respeta los matices de las experiencias y los valores humanos.

En resumen, si bien la IA en el reclutamiento ofrece ventajas considerables, es crucial aprovechar esta tecnología de manera responsable. Las organizaciones deben centrarse en identificar continuamente los sesgos y emplear herramientas avanzadas para mitigarlos. Este enfoque equilibrado garantiza que los procesos de contratación mantengan los más altos estándares de equidad y prácticas éticas, lo que permite a las empresas adoptar los avances tecnológicos al tiempo que salvaguardan la diversidad y la equidad en el lugar de trabajo.

12. Referencias

Ahmed, H. S. A. (21 de diciembre de 2020). *Directrices de auditoría para la inteligencia artificial*. ISACA. https://www.isaca.org/resources/news-and-trends/newsletters/atisaca/2020/volume-26/auditing-guidelines-for-artificial-intelligence

Instituto Alan Turing. (2022). *Reflexionar sobre el propósito, la posición y el poder: Turing Commons*. https://alan-turing-institute.github.io/turing-commons/skills-tracks/aeg/chapter5/reflect/

Albaroudi, E., Mansouri, T., & Alameer, A. (2024). Una revisión exhaustiva de las técnicas de IA para abordar el sesgo algorítmico en la contratación de empleos. *AI, 5*(1), 383–404. https://doi.org/10.3390/ai5010019

Albassam, W. A. (2023). El poder de la inteligencia artificial en el reclutamiento: una revisión analítica de las estrategias actuales de reclutamiento basadas en IA. *Revista Internacional de Revista de Negocios Profesionales*, *8*(6), e02089. https://doi.org/10.26668/businessreview/2023.v8i6.2089

ALTAI. (2020). *La lista de evaluación de la inteligencia artificial confiable*. https://altai.insight-centre.org/

Arivu Reclutamiento y Consultoría. (13 de julio de 2023). *Pros y contras de utilizar la inteligencia artificial (IA) en el reclutamiento | LinkedIn*. https://www.linkedin.com/pulse/pros-cons-using-artificial-intelligence/

Banerjee, S. (20 de febrero de 2022). *Big data en el sector de la contratación | LinkedIn*. https://www.linkedin.com/pulse/big-data-recruitment-sector-shantanu-banerjee-phd/

Bursell, M., & Roumbanis, L. (2024). Después de los algoritmos: Un estudio de los juicios metaalgorítmicos y la diversidad en el proceso de contratación en una gran empresa multisitio. *Big Data y Sociedad*, *11*(1), 20539517231221758. https://doi.org/10.1177/20539517231221758

CarrerasExpertos. (24 de abril de 2023). Reducción del sesgo en la contratación con la generación de descripciones de puestos de trabajo impulsada por IA. *Expertos en carreras*. https://www.careerexperts.co.uk/management-leadership/ai-powered-job-description

Castelnovo, A., Crupi, R., Greco, G., Regoli, D., Penco, I. G., & Cosentini, A. C. (2022). Una aclaración de los matices en el panorama de las métricas de equidad. *Scientific Reports*, *12*(1), 4209. https://doi.org/10.1038/s41598-022-07939-1

Chaker, N. (24 de octubre de 2018). *Fuentes de datos de candidatos: La lista de verificación del reclutador*. https://beamery.com/resources/blogs/candidate-data-sources-the-recruiter-s-checklist

Chen, Z. (2023). Colaboración entre reclutadores e inteligencia artificial: Eliminando los prejuicios humanos en el empleo. *Cognición, Tecnología y Trabajo*, *25*(1), 135–149. https://doi.org/10.1007/s10111-022-00716-0

Dennis, J. (22 de mayo de 2023). *Reclutamiento de IA: usos, ventajas y desventajas 2024*. Asesoramiento tecnológico. https://technologyadvice.com/blog/human-resources/ai-recruiting/

Océano Digital. (s.f.). *Cómo utilizar la IA en la contratación: técnicas y herramientas*. Recuperado el 26 de abril de 2024, de https://www.digitalocean.com/resources/article/ai-in-hiring

Recomendados. (2023, 22 de diciembre). *Cómo estos 6 líderes están utilizando la incorporación impulsada por IA*. Empresa rápida. https://www.fastcompany.com/90996367/how-leaders-using-ai-powered-onboarding

Fernandes, R. F. (2021). El Big Data como herramienta para potenciar los procesos de selección. *Revista Electrónica de Estudios Laborales Internacionales y Comparados, 10*(03).

Filtrada. (s.f.). *¿Qué es la selección de currículums con IA y cómo puede beneficiar a su empresa? | Blog filtrado*. Recuperado el 14 de marzo de 2024, de https://www.filtered.ai/blog/ai-resume-screening-fd

Friedman, B., & Nissenbaum, H. (1996). Sesgo en los sistemas informáticos. *ACM Transactions on Information Systems*, *14*(3), 330–347.

Fritts, M., & Cabrera, F. (2021). Los algoritmos de reclutamiento de IA y el problema de la deshumanización. *Ética y Tecnología de la Información*, *23*(4), 791–801. https://doi.org/10.1007/s10676-021-09615-w

Giordano, F., Morelli, N., Götzen, A. D., & Hunziker, J. (2018, junio). *El mapa de grupos de interés: una herramienta de conversación para diseñar servicios públicos liderados por las personas*. ServDes2018 - Prueba de concepto de diseño de servicios, Politecnico di Milano.

Google para desarrolladores. (2022). *Equidad: Tipos de sesgo | Aprendizaje automático*. Google para desarrolladores. https://developers.google.com/machine-learning/crash-course/fairness/types-of-bias

Escuela de Ingeniería y Ciencias Aplicadas John A. Paulson de Harvard. (6 de diciembre de 2023). *¿Cómo se puede eliminar el sesgo de las plataformas de contratación impulsadas por inteligencia artificial?* https://seas.harvard.edu/news/2023/06/how-can-bias-be-removed-artificial-intelligence-powered-hiring-platforms

Harwell, D. (2019a, 7 de julio). El FBI y el ICE descubren que las fotos de las licencias de conducir del estado son una mina de oro para las búsquedas de reconocimiento facial. *Washington Post*. https://www.washingtonpost.com/technology/2019/07/07/fbi-ice-find-state-drivers-license-photos-are-gold-mine-facial-recognition-searches/

Harwell, D. (2019b, 19 de diciembre). Un estudio federal confirma el sesgo racial de muchos sistemas de reconocimiento facial y pone en duda su uso en expansión. *Washington Post*. https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/

Heymans, Y. (7 de septiembre de 2022). *Machine learning en el reclutamiento: una inmersión profunda*. https://www.herohunt.ai/blog/machine-learning-in-recruitment-a-deep-dive

Hidalgo, M. A. S. (2019). *Diseño de un conjunto de herramientas éticas para el desarrollo de aplicaciones de IA* [Universidad Tecnológica de Delft]. http://resolver.tudelft.nl/uuid:b5679758-343d-4437-b202-86b3c5cef6aa

Hoory, L., & Botorff, C. (7 de agosto de 2022). *¿Qué es un análisis de las partes interesadas? Todo Lo Que Necesitas Saber – Forbes Advisor*. https://www.forbes.com/advisor/business/what-is-stakeholder-analysis/

Hunkenschroer, A. L., & Luetge, C. (2022). Ética del reclutamiento y la selección habilitados por la IA: una agenda de revisión e investigación. *Revista de Ética Empresarial*, *178*(4), 977–1007. https://doi.org/10.1007/s10551-022-05049-6

Javaid, S. (2024, 12 de enero). *Reconocimiento facial: mejores prácticas y casos de uso en 2024*. AIMultiple: Casos de uso de alta tecnología & Herramientas para hacer crecer tu negocio. https://research.aimultiple.com/facial-recognition/

Köchling, A., & Wehner, M. C. (2020). Discriminado por un algoritmo: Una revisión sistemática de la discriminación y la equidad mediante la toma de decisiones algorítmica en el contexto de la contratación y el desarrollo de RRHH. *Investigación Empresarial*, *13*(3), 795–848. https://doi.org/10.1007/s40685-020-00134-w

Kumar, V. (2013). *101 Métodos de diseño: un enfoque estructurado para impulsar la innovación en su organización*. John Wiley & Sons, Inc. https://learning.oreilly.com/library/view/101-design-methods/9781118083468/cvi.xhtml

Lange, A. R., & Duarte, N. (2018, 4 de abril). Comprensión del sesgo en el diseño algorítmico. *ASME ISHOW / IDEA LAB*. https://medium.com/impact-engineered/understanding-bias-in-algorithmic-design-db9847103b6e

Lawton, G. (29 de agosto de 2022). *Sesgo de contratación de IA: todo lo que necesitas saber | TechTarget*. Software de RRHH. https://www.techtarget.com/searchhrsoftware/tip/AI-hiring-bias-Everything-you-need-to-know

Lee, N. T., Resnick, P., & Barton, G. (22 de mayo de 2019). *Detección y mitigación de sesgos algorítmicos: mejores prácticas y políticas para reducir los daños a los consumidores | Brookings*. https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

Ma, Y. (2024). Método inteligente de recomendación de recursos de talento basado en Big Data. *Tecnología de computación inteligente y automatización*. https://doi.org/10.3233/ATDE231176

Aprendizaje automático y equidad. (24 de mayo de 2021). Seminarios web y tutoriales de Microsoft Research. https://youtu.be/ZtN6Qx4KddY?si=kWo9MxkNQ7ko1HUB

Marr, B. (2023, 12 de diciembre). *Incorporación de empleados mejorada con IA: una nueva era en las prácticas de RRHH*. Forbes. https://www.forbes.com/sites/bernardmarr/2023/12/12/ai-enhanced-employee-onboarding-a-new-era-in-hr-practices/

Naakka, S.-S. (2018). *BIG DATA COMO VENTAJA COMPETITIVA EN LA BÚSQUEDA DE CANDIDATOS A LA CONSULTORÍA EXECUTIVE SEARCH ¿Más, más*

rápidos y mejores expertos en TI con big data? [Universidad de Turku]. https://www.utupub.fi/bitstream/handle/10024/145309/Naakka%20Sara-Selia.pdf?sequence=1&isAllowed=y

Engig. (2024). *Software de descripción de puestos | Ongig*. https://www.ongig.com

Hoja de vida. (16 de octubre de 2023). *De los currículos a los algoritmos: el papel de la IA en la selección de candidatos | LinkedIn*. https://www.linkedin.com/pulse/from-resumes-algorithms-role-ai-screening-candidates-resumement/

Saplicki, C. (2022, 11 de noviembre). Explicación de la equidad: definiciones y métricas. *Medio*. https://medium.com/ibm-data-ai/fairness-explained-definitions-and-metrics-9690f8e0a4ea

SDT. (s.f.). *Herramientas | Herramientas de diseño de servicios*. Recuperado el 18 de abril de 2024, de https://servicedesigntools.org/tools.html

Sheard, N. (2022). Discriminación laboral por algoritmo: ¿se puede responsabilizar a cualquiera? *Revista de Derecho de la Universidad de Nueva Gales del Sur*, *45*(2). https://doi.org/10.53637/XTQY4027

Shipman, M. (4 de marzo de 2021). Las comprobaciones de las redes sociales pueden generar sesgos en la contratación. *Futuridad*. https://www.futurity.org/cybervetting-human-resources-bias-2527382-2/

Stickdorn, M., Hormess, M. E., Lawrence, A., & Schneider, J. (2018a). *Esto es Service Design Doing*. O'Reilly Media, Inc. https://learning.oreilly.com/library/view/this-is-service/9781491927175/ch03.html

Stickdorn, M., Hormess, M., Lawrence, A., & Schneider, J. (2018b, julio). *#TISDD Biblioteca de métodos*. Esto es hacer diseño de servicios. https://www.thisisservicedesigndoing.com/methods

Sukernek, W. (2024). *Cómo detectar sesgos en la contratación*. FloCareer. https://blog.flocareer.com/how-to-spot-bias-in-hiring

Teodorescu, M. (2020). *Criterios de equidad | Explorando la equidad en el aprendizaje automático para el desarrollo internacional | Recursos complementarios*. MIT OpenCourseWare. https://ocw.mit.edu/courses/res-ec-001-exploring-fairness-in-machine-learning-for-international-development-spring-2020/pages/module-three-framework/fairness-criteria/

UNESCO. (2023). *Evaluación de impacto ético. Una herramienta de la Recomendación sobre la Ética de la Inteligencia Artificial*. UNESCO. https://doi.org/10.54678/YTSA7796

Vivek, R. (2023). Mejorar la diversidad y reducir el sesgo en la contratación a través de la IA: una revisión de las estrategias y los desafíos. *Информатика. Экономика. Управление - Informatics. Economía. Gestión*, *2*(4), 0101–0118. https://doi.org/10.47813/2782-5280-2023-2-4-0101-0118

Wallbridge, A. (2023, 13 de abril). *Cómo llevar a cabo una matriz de análisis de partes interesadas a prueba de balas en 4 pasos útiles*. Formación TSW. https://www.tsw.co.uk/blog/leadership-and-management/stakeholder-analysis-matrix/

Wirtz, D. (3 de agosto de 2022). *¿Qué es un taller? (+2 Ejemplos) | Escuela de Facilitadores*. https://www.facilitator.school/blog/what-is-a-workshop

www.recruiter.com. (26 de abril de 2023). *Pros y contras de utilizar la IA en el reclutamiento*. Recruiter.Com. https://www.recruiter.com/recruiting/pros-and-cons-of-using-ai-in-recruiting/

YLE. (11 de noviembre de 2023). *Creamos a Ella, de 13 años: Tiktok le ofreció contenido sobre el suicidio y el conteo de calorías*. Noticias de Yle. https://yle.fi/a/74-20059318

Zhong, Z. (2018, 22 de octubre). *Un tutorial sobre la equidad en el aprendizaje automático*. Medio. https://towardsdatascience.com/a-tutorial-on-fairness-in-machine-learning-3ff8ba1040cb