



Microcredencial de IA ética

GUÍA DE APOYO

UC4 | Transparencia

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

¿Cómo utilizar esta Guía de Apoyo?

Este documento es interactivo. A lo largo del documento encontrará enlaces a información adicional.



Botón que le lleva al principio del documento. Este icono aparece en la esquina superior derecha de las páginas.



Siempre que veas esta flecha, significa que tienes un **texto interactivo en color** sobre el que hacer clic, que tiene asociado un enlace externo.

DESCARGA DE RESPONSABILIDAD: Tenga en cuenta que no podemos garantizar la disponibilidad permanente de contenidos externos, como vídeos, ya que pueden estar sujetos a cambios o ser retirados por sus autores o las plataformas que los albergan.

Índice

Haga clic en el menú

01. Introducción

02. Importancia de la transparencia en los sistemas de IA

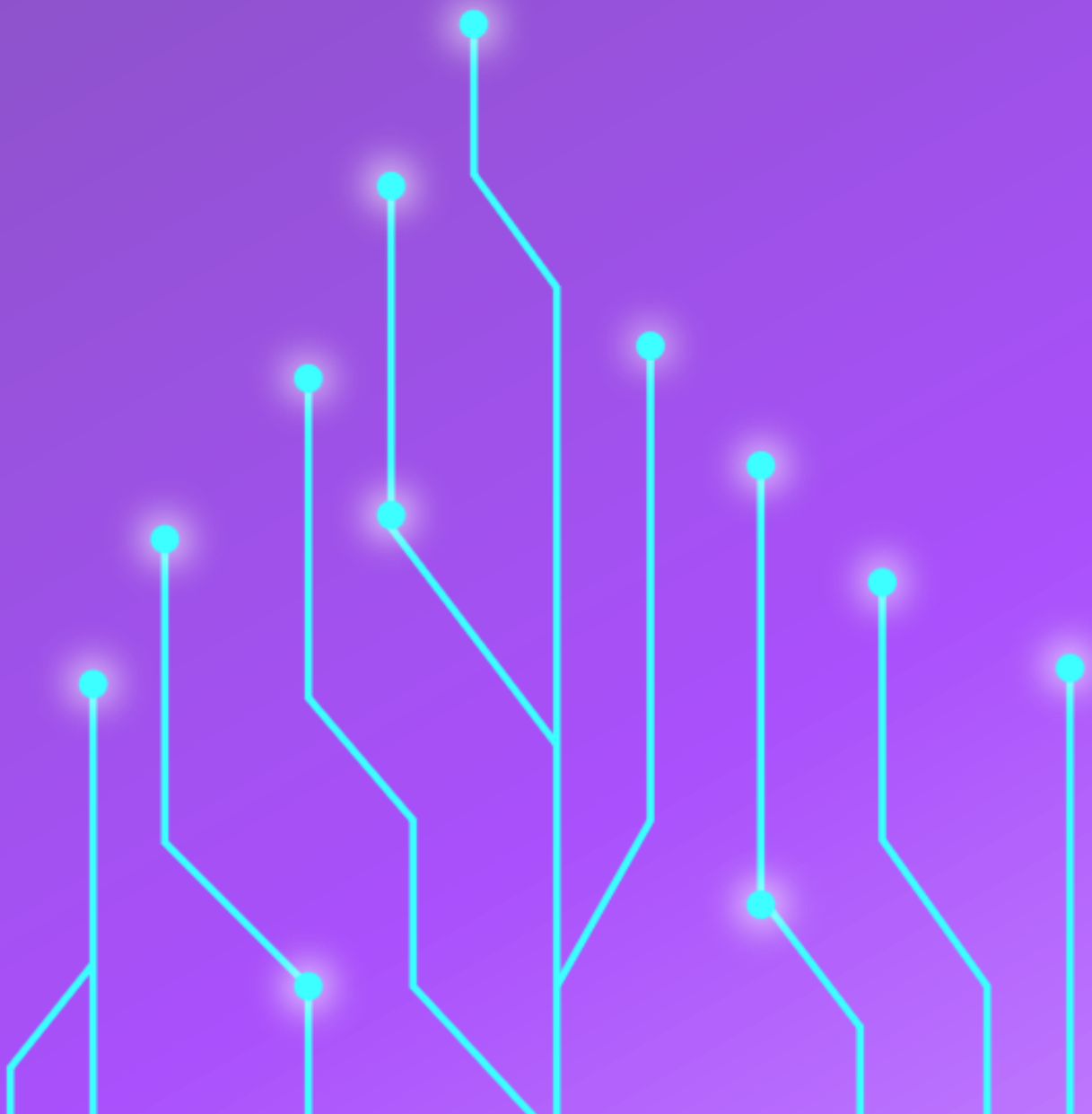
03. Relación entre transparencia y sesgo algorítmico

04. Estrategias para fomentar la transparencia en sistemas de IA

05. Conclusión

01. Introducción

UC4 | Transparencia





01. Introducción

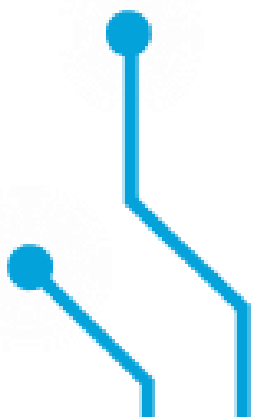
En esta unidad de competencia, los alumnos adquirirán conocimientos sobre la importancia de la transparencia en los sistemas de IA, centrándose en la comprensión de los conceptos básicos, la relación entre la transparencia y el sesgo algorítmico y la relevancia de las estrategias para garantizar que los sistemas de IA sean comprensibles, explicables y accesibles para las partes interesadas, reconociendo las implicaciones en el mundo real apreciando lo importante que pueden ser los modelos interpretables, la documentación clara y la comunicación eficaz para fomentar una cultura de transparencia, mitigando el sesgo algorítmico.

Los resultados de aprendizaje para este curso incluyen:

- **Importancia de la transparencia en los sistemas de IA** y su relevancia para garantizar que los sistemas de IA sean comprensibles, explicables y accesibles para las partes interesadas. Identificaremos los beneficios y valoraremos la importancia de la transparencia de los sistemas de IA para generar confianza y permitir la comprensión de las partes interesadas. A modo de ejemplo: Un modelo de IA diseñado para detectar el cáncer, aunque solo se equivoque en un 1%, podría poner en peligro una vida. En casos así, la IA y los humanos tienen que trabajar juntos, y la tarea resulta mucho más fácil cuando el modelo de IA puede explicar cómo ha llegado a una determinada decisión. La transparencia convierte a la IA en un jugador de equipo.

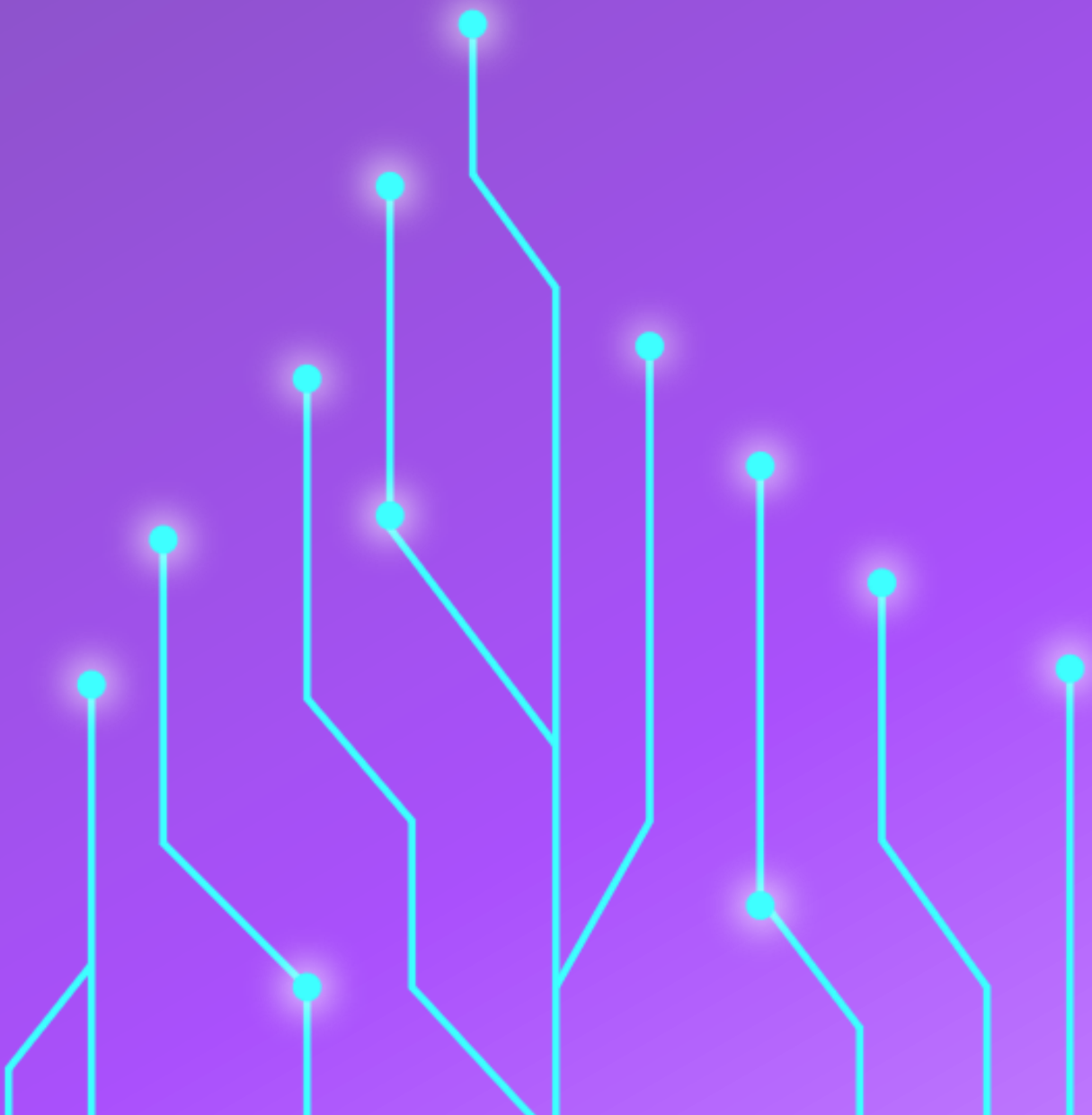


- **Relación entre transparencia y sesgo algorítmico** para encontrar la conexión entre transparencia y sesgo algorítmico, reconociendo los peligros de la opacidad y cómo una mayor transparencia puede ayudar a identificar, prevenir y mitigar los resultados sesgados en los sistemas de IA. Reconoceremos la importancia de la transparencia para abordar y mitigar el sesgo algorítmico. A modo de ejemplo: muy a menudo, los algoritmos de IA son opacos en el sentido de que sus explicaciones no están disponibles para todas las partes interesadas. Esta opacidad puede tener diferentes orígenes. A veces, las instituciones o las empresas no comunican cuándo se basan en sistemas de IA o cómo funcionan estos sistemas.
- **Estrategias para promover la transparencia en los sistemas de IA**, tales como utilizar modelos interpretables, proporcionar documentación clara y comunicar los procesos de toma de decisiones de las aplicaciones de IA. Explicaremos la importancia de estas estrategias para promover una cultura de transparencia y mitigar el sesgo algorítmico. Como ejemplo, la IA puede afectar a varias partes interesadas, como usuarios, clientes, empleados, directivos, reguladores o la sociedad. Para garantizar la transparencia y la rendición de cuentas, es necesario implicar y capacitar a las partes interesadas en la IA a lo largo de todo el ciclo de vida de la IA.



02. Importancia de la transparencia en los sistemas de IA

UC4 | Transparencia





02. Importancia de la transparencia en los sistemas de IA

La transparencia es un principio fundamental en el desarrollo y la implantación de sistemas de inteligencia artificial (IA).

La transparencia en la IA se refiere a la apertura y accesibilidad de los sistemas de IA, lo que permite a las partes interesadas comprender cómo funcionan los algoritmos, por qué se toman determinadas decisiones y qué factores influyen en sus resultados. Abarca varios aspectos, como la disponibilidad de información sobre fuentes de datos, modelos algorítmicos, procesos de toma de decisiones y posibles sesgos. Los sistemas de IA transparentes permiten a las partes interesadas, incluidos los usuarios, los desarrolladores, los responsables políticos y el público en general, examinar y cuestionar los resultados algorítmicos, fomentando la confianza y la responsabilidad.

Una de las principales ventajas de los sistemas de IA transparentes es su comprensibilidad. Cuando los algoritmos de IA son transparentes, las partes interesadas pueden comprender cómo funcionan y por qué producen determinados resultados. Esta comprensión permite a los usuarios confiar en las tecnologías de IA y tomar decisiones informadas sobre su uso. Por ejemplo, en el contexto de un modelo de IA de diagnóstico médico, la transparencia permite a los profesionales sanitarios comprender cómo ha llegado el modelo a su diagnóstico, lo que les permite validar su precisión y fiabilidad antes de tomar decisiones sobre el tratamiento.



Además, la transparencia facilita la explicabilidad, que es esencial para garantizar que los sistemas de IA puedan ofrecer explicaciones interpretables de sus decisiones y acciones. Una IA explicable permite a las partes interesadas comprender los fundamentos de los resultados algorítmicos e identificar y corregir sesgos o errores. Por ejemplo, en el caso de un sistema de IA para la aprobación de préstamos, la transparencia y la explicabilidad permiten a los solicitantes de préstamos entender por qué su solicitud fue aprobada o denegada, proporcionando información sobre el proceso de toma de decisiones y vías de recurso si creen que la decisión fue sesgada o injusta.

Además, la transparencia mejora la accesibilidad de los sistemas de IA, haciéndolos más inclusivos y equitativos. Cuando los algoritmos de IA son transparentes, las partes interesadas de diversos orígenes y niveles de experiencia pueden acceder a la información sobre su funcionamiento y resultados e interpretarla. Esta accesibilidad garantiza que las tecnologías de IA no sólo sean comprensibles, sino también utilizables por una amplia gama de usuarios, incluidos aquellos con discapacidades o conocimientos técnicos limitados. Por ejemplo, en el desarrollo de herramientas de accesibilidad basadas en IA para personas con discapacidad, la transparencia permite a los usuarios comprender cómo funcionan las herramientas y cómo pueden beneficiarse de ellas.

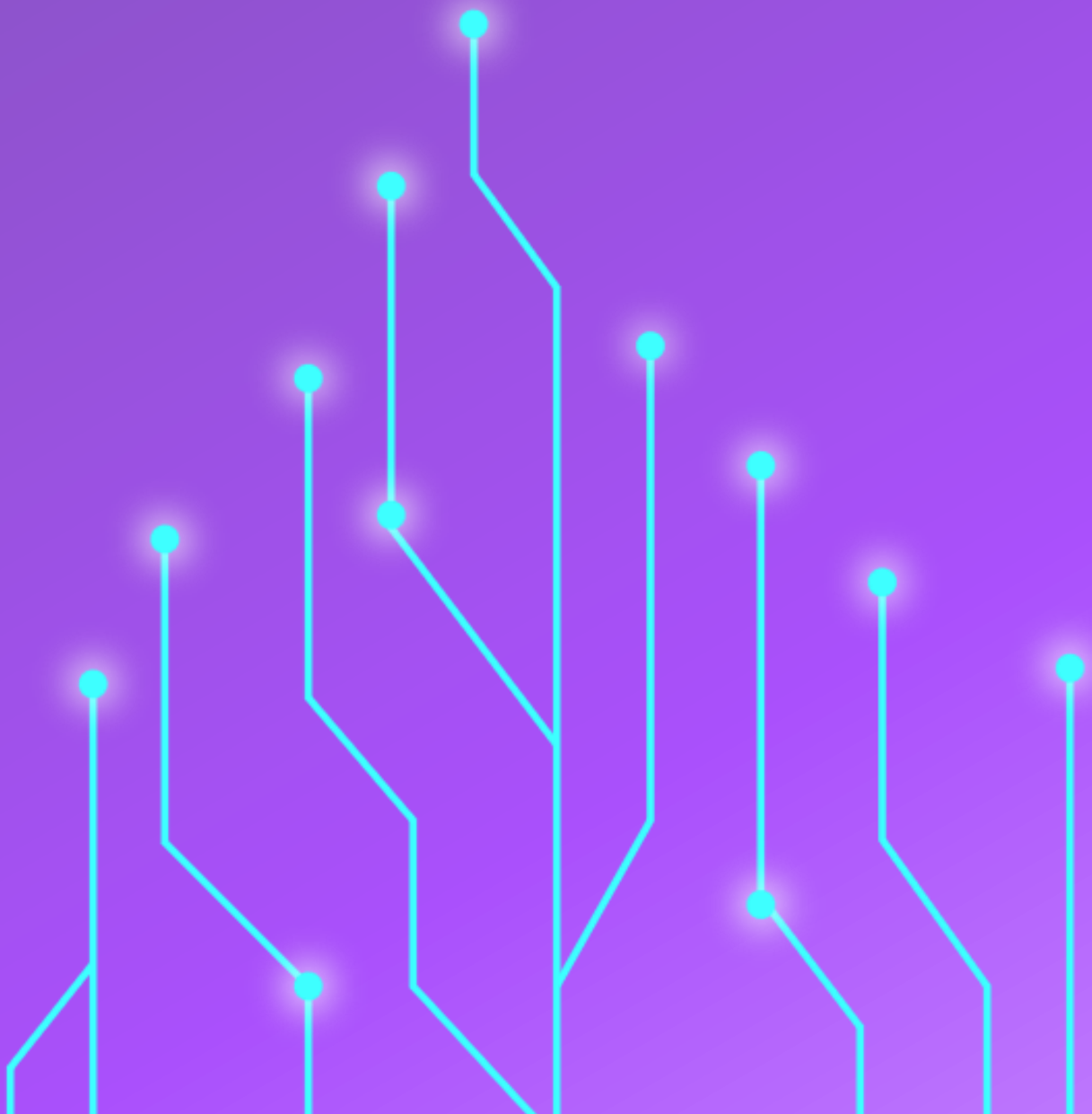
Un ejemplo ilustrativo de la importancia de la transparencia en los sistemas de IA es el desarrollo de modelos de IA para el diagnóstico médico, como la detección del cáncer. Incluso si un modelo de IA es muy preciso, con una tasa de acierto del 99%, el margen de error del 1% restante podría tener consecuencias potencialmente mortales para los pacientes. En situaciones tan críticas, la transparencia es esencial para garantizar que los profesionales sanitarios puedan entender cómo ha llegado el modelo de IA a su diagnóstico y puedan verificar su precisión antes de tomar decisiones sobre el tratamiento. Al ofrecer explicaciones transparentes sobre su proceso de toma de decisiones, el modelo de IA se convierte en una valiosa herramienta para los profesionales sanitarios, mejorando su capacidad para diagnosticar y tratar a los pacientes con eficacia.

Como ya hemos leído en este folleto, el sesgo algorítmico se refiere a errores sistemáticos o injusticias en los algoritmos de IA que dan lugar a resultados discriminatorios para determinadas personas o grupos. Estos sesgos pueden proceder de diversas fuentes, como datos de entrenamiento sesgados, un diseño algorítmico defectuoso o sesgos humanos codificados en el sistema. Las consecuencias de los sesgos algorítmicos pueden ser de gran alcance, ya que perpetúan las desigualdades, refuerzan los estereotipos y socavan la confianza en los sistemas de IA.



03. Relación entre transparencia y sesgo algorítmico

UC4 | Transparencia





03. Relación entre transparencia y sesgo algorítmico

La opacidad, o la falta de transparencia, agrava los riesgos asociados al sesgo algorítmico.

Muy a menudo, los algoritmos de IA son opacos, lo que significa que las explicaciones sobre sus decisiones y acciones no están fácilmente disponibles para todas las partes interesadas. Esta opacidad puede deberse a diversas causas, como el secreto institucional, la confidencialidad empresarial o la complejidad técnica. Cuando las partes interesadas carecen de acceso a la información sobre los sistemas de IA, no pueden evaluar la imparcialidad, fiabilidad o implicaciones éticas de los resultados algorítmicos, lo que conduce a una falta de responsabilidad y a un daño potencial.

La transparencia es un antídoto crucial contra la opacidad de los sistemas de IA, ya que permite a las partes interesadas examinar y cuestionar las decisiones algorítmicas, mitigando así los riesgos de sesgo algorítmico. Al aumentar la transparencia, los desarrolladores y profesionales de la IA pueden ofrecer a las partes interesadas información sobre cómo funcionan los sistemas de IA, por qué se toman determinadas decisiones y qué factores influyen en sus resultados. Los sistemas de IA transparentes permiten a las partes interesadas identificar y abordar los sesgos, validar la precisión algorítmica y responsabilizar a los desarrolladores del uso ético y equitativo de las tecnologías de IA.

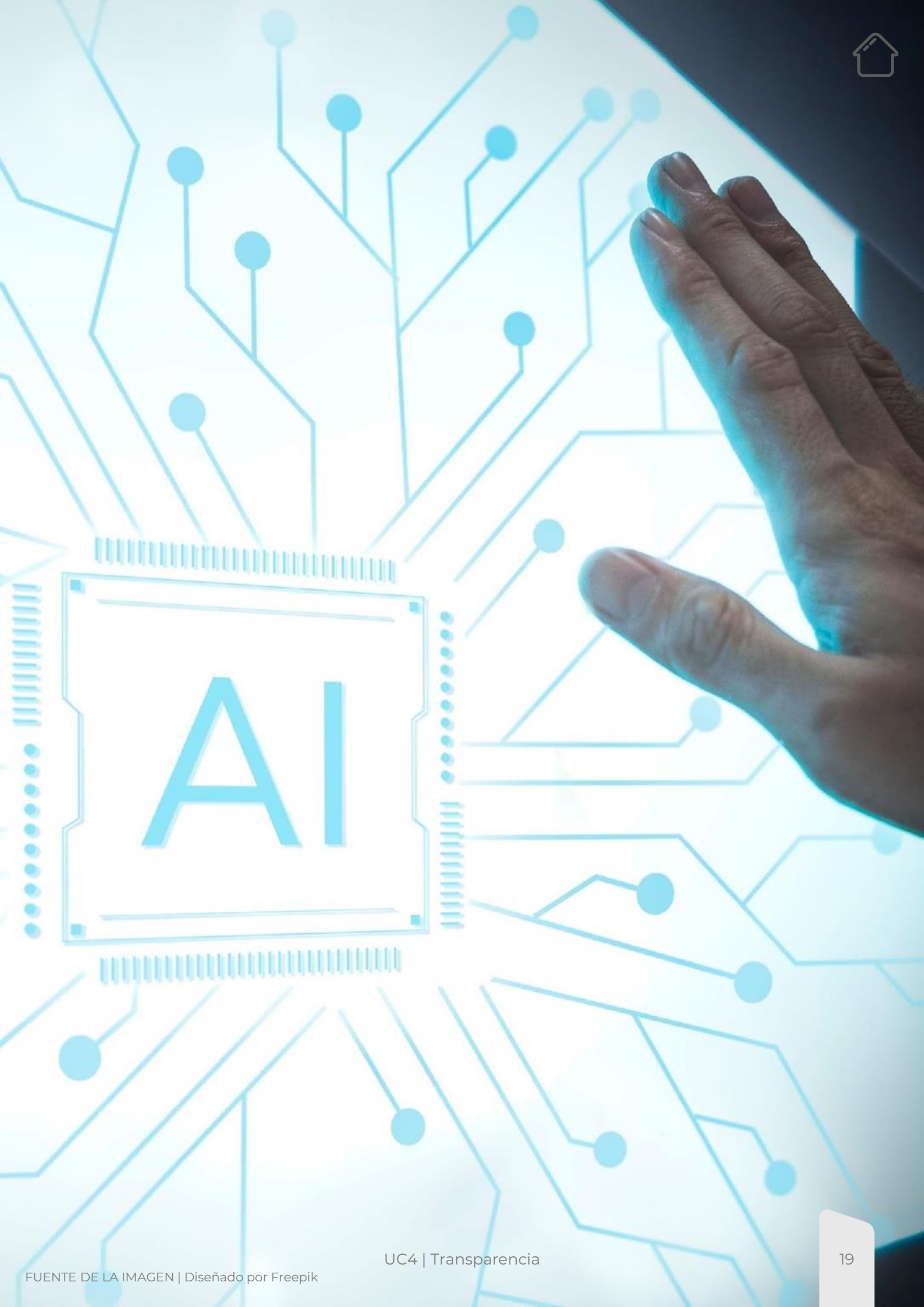


Uno de los principales beneficios de la transparencia a la hora de abordar el sesgo algorítmico es la capacidad de detectar y mitigar los resultados sesgados. Cuando los algoritmos de IA son transparentes, las partes interesadas pueden examinar el proceso de toma de decisiones e identificar los casos en los que puede haber sesgo. Por ejemplo, en el contexto de un sistema de IA de contratación, la transparencia permite a las partes interesadas evaluar si el sistema discrimina injustamente a determinados grupos demográficos en el proceso de selección. Al identificar los resultados sesgados, las partes interesadas pueden tomar medidas correctivas para mitigar el daño causado por el sesgo algorítmico y promover la justicia y la equidad.

Además, la transparencia facilita la rendición de cuentas y la confianza en los sistemas de IA. Cuando las partes interesadas tienen acceso a la información sobre los algoritmos de IA, pueden exigir responsabilidades a los desarrolladores y profesionales por el uso ético y equitativo de las tecnologías de IA. Los sistemas de IA transparentes generan confianza entre los usuarios, los reguladores y el público en general, fomentando la confianza en la fiabilidad e imparcialidad de los resultados algorítmicos. Por ejemplo, en el despliegue de sistemas de IA para la justicia penal o la atención sanitaria, la transparencia permite a las partes interesadas comprender cómo se toman las decisiones y garantizar que estas se ajustan a los principios éticos y las normas jurídicas.

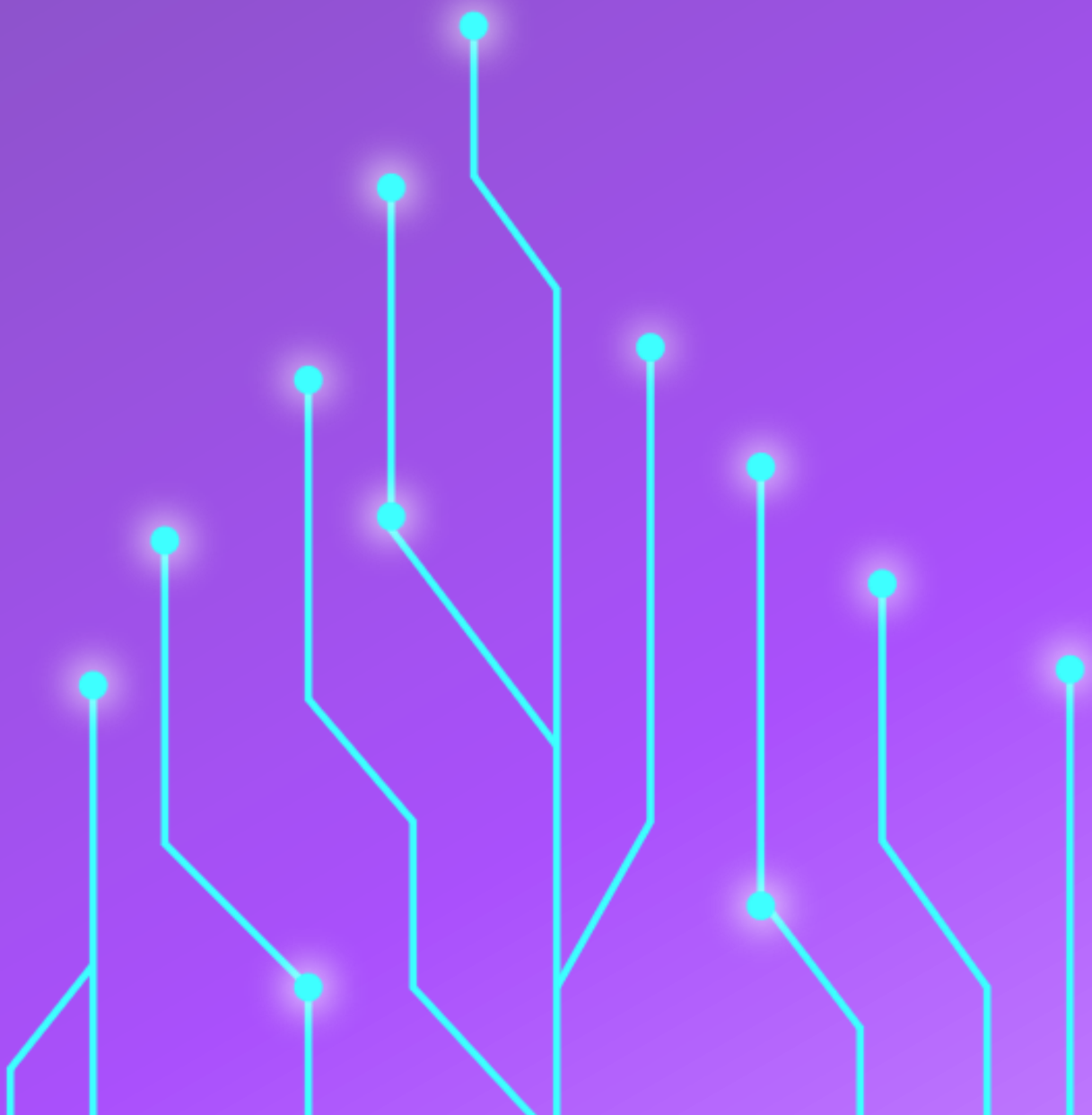
La transparencia desempeña un papel fundamental a la hora de abordar y mitigar el sesgo algorítmico en los sistemas de IA. Al aumentar la transparencia, las partes interesadas pueden detectar y mitigar los resultados sesgados, promover la rendición de cuentas y generar confianza en las tecnologías de IA. A medida que la IA sigue evolucionando e integrándose en diversos aspectos de la sociedad, la transparencia sigue siendo esencial para garantizar que los sistemas de IA se desarrollen y desplieguen de una manera que respete las normas éticas y promueva la justicia y la equidad. Mediante una comprensión global de la relación entre transparencia y sesgo algorítmico, los alumnos pueden contribuir al desarrollo responsable y ético de las tecnologías de IA, creando así un futuro más equitativo e inclusivo.

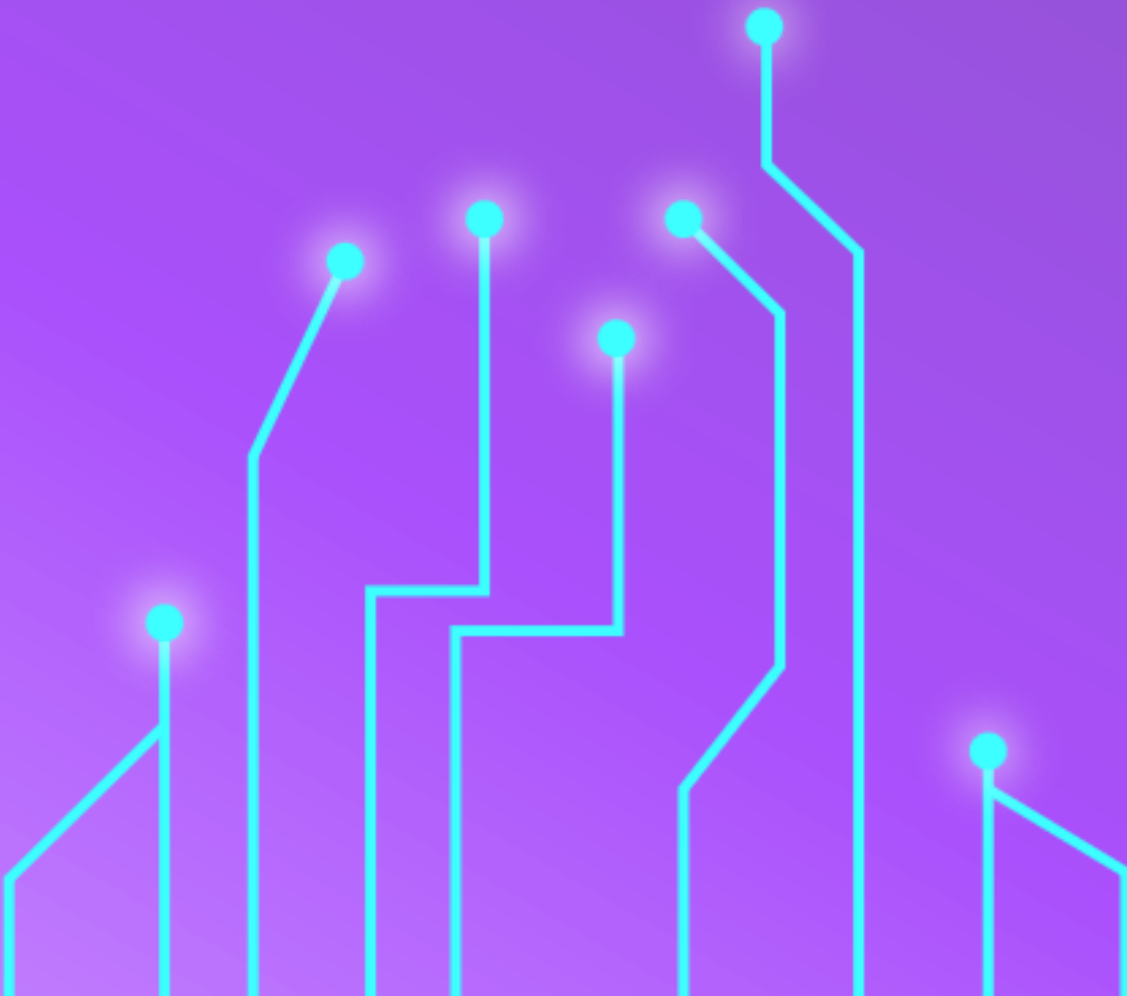




04. Estrategias para fomentar la transparencia en los sistemas de IA

UC4 | Transparencia





04. Estrategias para fomentar la transparencia en los sistemas de IA

Existen varias estrategias para fomentar la transparencia en los sistemas de **inteligencia artificial (IA)**, tales como utilizar modelos interpretables, proporcionar documentación clara y comunicar los procesos de toma de decisiones.

> Modelos interpretables

Los modelos interpretables representan una estrategia clave para fomentar la transparencia en los sistemas de IA. Son modelos de aprendizaje automático que producen resultados fáciles de entender e interpretar por los humanos. He aquí algunos ejemplos:

- **Regresión lineal:** La regresión lineal es un modelo sencillo e interpretable que suele utilizarse para predecir resultados numéricos. Funciona ajustando una línea recta a los puntos de datos, lo que facilita la interpretación de la relación entre las variables de entrada y el resultado.
- **Árboles de decisión:** Los árboles de decisión son modelos jerárquicos que toman decisiones basándose en una serie de afirmaciones del tipo si-entonces. Cada nodo del árbol representa una decisión basada en una característica de los datos, lo que facilita el seguimiento de la lógica que subyace a las predicciones del modelo.



- **Regresión logística:** La regresión logística es un modelo estadístico utilizado para tareas de clasificación binaria. Calcula la probabilidad de que una instancia pertenezca a una clase determinada en función de sus características de entrada, lo que lo hace interpretable y fácil de entender.
- **Modelos basados en reglas:** Los modelos basados en reglas, como los árboles de clasificación y regresión (CART) o las reglas de decisión, traducen directamente las características de entrada en reglas de decisión. Estas reglas son fáciles de interpretar y pueden proporcionar información sobre el modo en que el modelo realiza predicciones.
- **Modelos aditivos generalizados (GAM):** Los GAM son modelos flexibles que pueden captar relaciones complejas entre las variables de entrada y la variable objetivo sin dejar de ser interpretables. Utilizan funciones suaves para representar la relación entre cada variable de entrada y la salida, lo que permite interpretar fácilmente las predicciones del modelo.

Los modelos interpretables permiten a las partes interesadas comprender cómo toman decisiones los algoritmos de IA y qué factores influyen en sus resultados. Mediante el uso de modelos interpretables, los desarrolladores de IA pueden mejorar la transparencia y la rendición de cuentas, permitiendo validar los resultados algorítmicos e identificar posibles sesgos o errores.

Por ejemplo, en el contexto de un sistema de IA de calificación crediticia, el uso de modelos interpretables permite a las partes interesadas comprender los factores que contribuyen a las decisiones crediticias, como los ingresos, el historial crediticio y los niveles de deuda, fomentando así la transparencia y la equidad en las prácticas de préstamo.

➤ **Documentación clara**

Una documentación clara es otra estrategia crucial para promover la transparencia en los sistemas de IA. La documentación proporciona a las partes interesadas información sobre el diseño, el desarrollo y el despliegue de los algoritmos de IA, incluidas las fuentes de datos, las técnicas de preprocesamiento, las arquitecturas de los modelos y las métricas de evaluación.

Al documentar exhaustivamente los sistemas de IA, los desarrolladores pueden mejorar la transparencia y la rendición de cuentas, permitiendo a las partes interesadas comprender los procesos y supuestos subyacentes de las tecnologías de IA. Por ejemplo, en el desarrollo de un sistema de IA de mantenimiento predictivo para equipos industriales, una documentación clara permite a las partes interesadas evaluar la fiabilidad y precisión de los modelos predictivos, comprender las recomendaciones de mantenimiento y verificar el cumplimiento de las normas de seguridad.

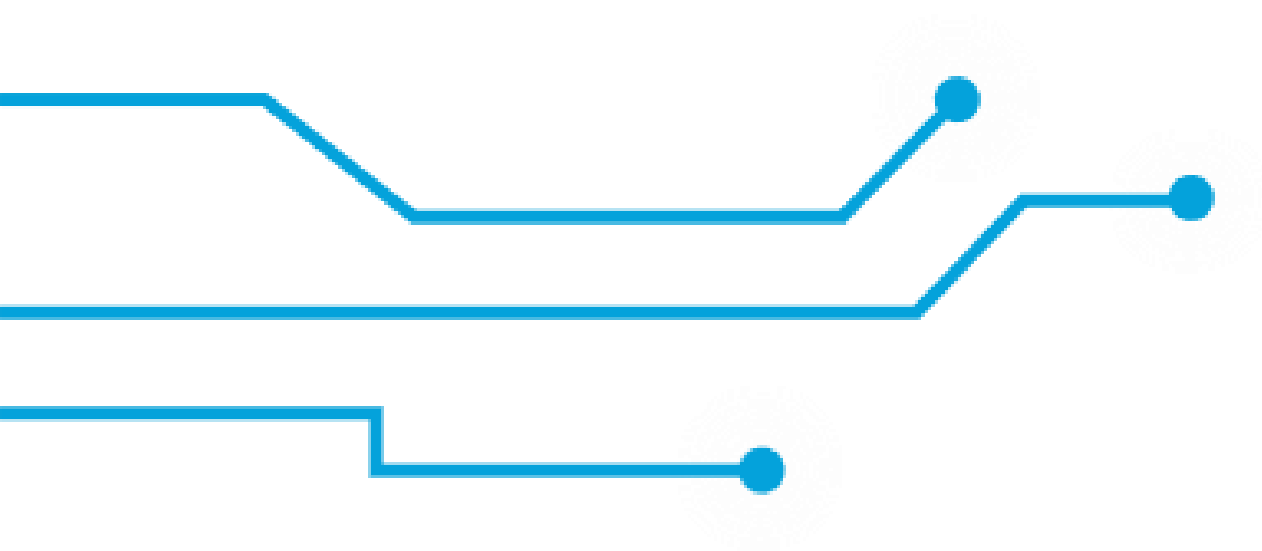




> Comunicación eficaz de la toma de decisiones

La comunicación efectiva de los procesos de toma de decisiones es esencial para promover la transparencia en los sistemas de IA. La comunicación garantiza que las partes interesadas estén informadas de los fundamentos, la lógica y las implicaciones de las decisiones algorítmicas de la IA.

Al comunicar los procesos de toma de decisiones de forma clara y transparente, los desarrolladores de IA pueden generar confianza entre los usuarios, los reguladores y el público en general. Por ejemplo, en el despliegue de sistemas de IA para el diagnóstico sanitario, una comunicación eficaz garantiza que los profesionales sanitarios y los pacientes entiendan cómo se toman las decisiones de diagnóstico, lo que les permite confiar y verificar la exactitud de los diagnósticos generados por la IA.



> Participación y capacitación de las partes interesadas en la IA

La participación y la capacitación de las partes interesadas en la IA a lo largo de todo su ciclo de vida son cruciales para garantizar la transparencia y la rendición de cuentas. La participación de las partes interesadas implica a usuarios, clientes, empleados, directivos, reguladores y a la sociedad en el diseño, desarrollo, despliegue y evaluación de los sistemas de IA.

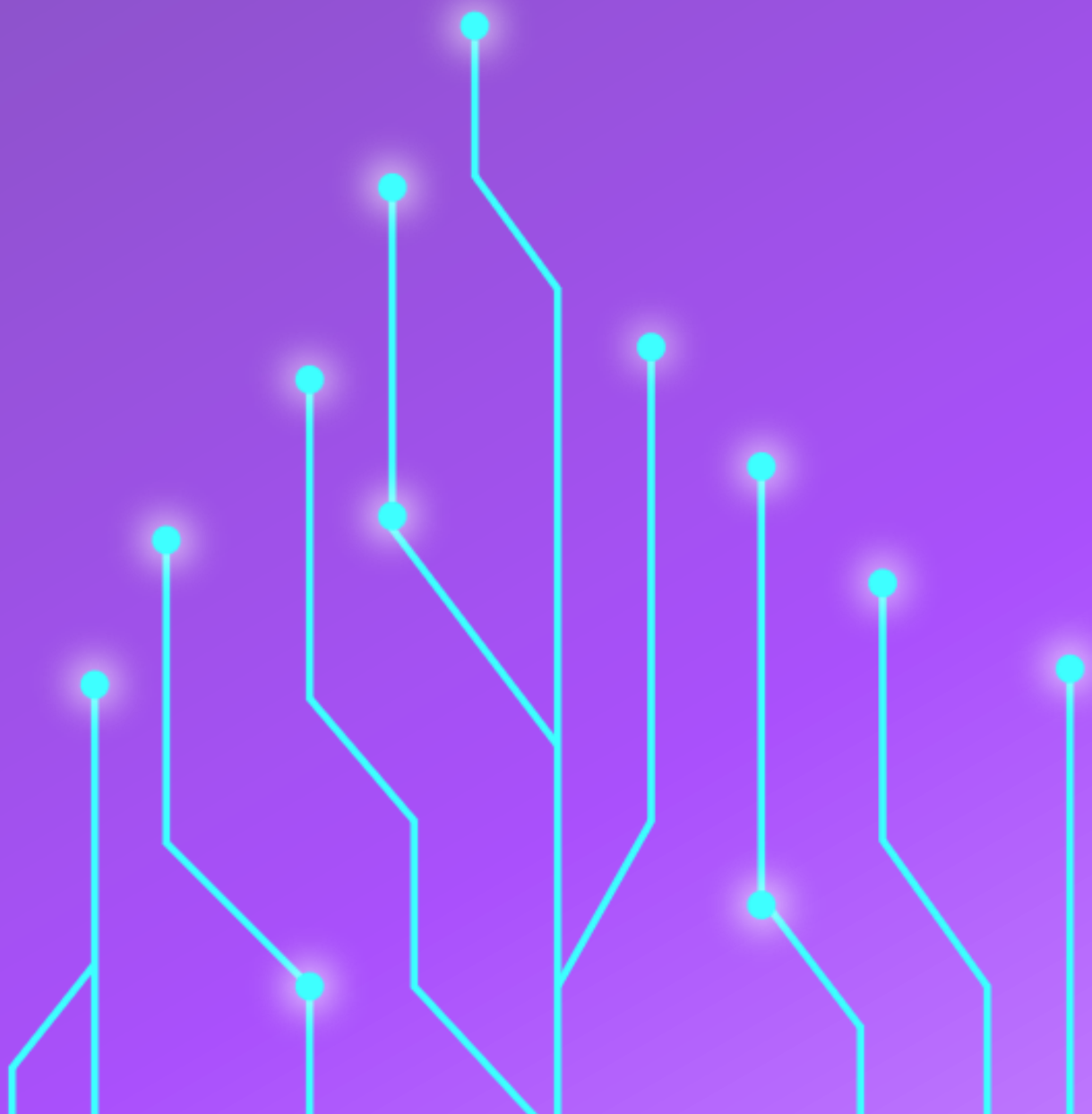
Al involucrar a las partes interesadas, los desarrolladores de IA pueden obtener información valiosa sobre sus necesidades, preferencias y preocupaciones, promoviendo así la transparencia, la responsabilidad y la toma de decisiones éticas. Por ejemplo, en el desarrollo de vehículos autónomos impulsados por IA, la colaboración con los reguladores y la sociedad garantiza que se tengan en cuenta la seguridad, la privacidad y las consideraciones éticas, aumentando la transparencia y la confianza en la tecnología.





05. Conclusión

UC4 | Transparencia

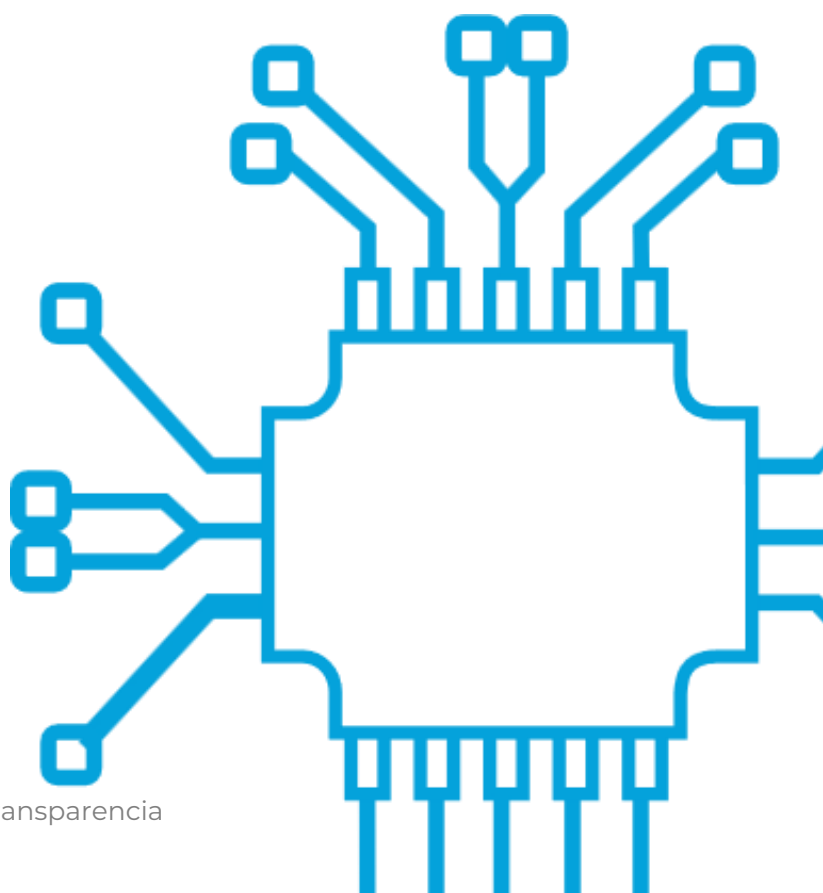


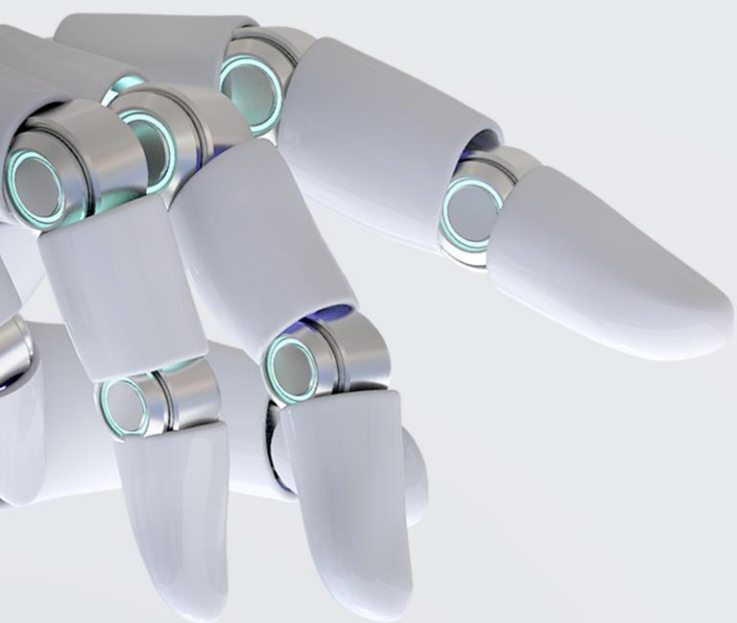


05. Conclusión

En conclusión, nunca se insistirá lo suficiente en la importancia de la transparencia en los sistemas de IA, ya que constituye la base para generar confianza, rendir cuentas y mitigar el sesgo algorítmico. Además, la comprensión de la relación entre la transparencia y el sesgo algorítmico pone de relieve la necesidad de abordar la opacidad como medio para identificar, prevenir y mitigar los resultados sesgados en los sistemas de IA.

Por último, la exploración de estrategias para promover la transparencia dota a los estudiantes de herramientas prácticas para mejorar la rendición de cuentas y fomentar la confianza en las tecnologías de IA. Gracias a una comprensión global de estos conceptos, los estudiantes están mejor preparados para afrontar los retos éticos del desarrollo y el despliegue de la IA, contribuyendo al avance de sistemas de IA responsables y equitativos.







Charlie



Cofinanciado por
la Unión Europea

Financiado por la Unión Europea. Las opiniones y puntos de vista expresados solo comprometen a su(s) autor(es) y no reflejan necesariamente los de la Unión Europea o los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA pueden ser considerados responsables de ellos.



Universitat
de les Illes Balears



helixconnect



2022-1-ES01-KA220-HED-000085257