DESAFIANDO EL SESGO EN BIG DATA PARA LA IA Y EL APRENDIZAJE AUTOMÁTICO

Matrices de competencias del WP2 para evitar el "sesgo algorítmico"

Cofinanciado por la Unión Europea

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

Índice de Contenidos

1. INTRODUCCIÓN: EL PROYECTO CHARLIE

1.1. El proyecto Charlie: visión general

CHARLIE es un proyecto ERASMUS+ KA2 con un periodo de ejecución de 30 meses, entre el 30/12/2022 y el 29/06/2025 implementado por un consorcio de seis (6) socios de cinco (5) países europeos: España, Portugal, Rumanía, Finlandia y Dinamarca.

La inteligencia artificial (IA de ahora en adelante) forma parte de nuestra vida cotidiana. Desde el algoritmo que reconoce nuestro rostro Cuando caminamos por la calle e informa a la policía hasta el algoritmo que elige la publicidad que veremos en nuestras redes sociales, la IA está en todas partes. Sin embargo, aunque el aprendizaje automático (ML, de las siglas en inglés) y la IA son matemáticas, solo a veces son adecuados, y esto sucede porque los datos procesados para llegar a cualquier conclusión pueden ser, y a menudo son, incompletos o sesgados. Las ciencias sociales han estado estudiando el sesgo humano durante muchos años. Surge de la asociación implícita que refleja un sesgo del que no somos conscientes, lo que puede dar lugar a múltiples resultados adversos. La IA y el ML no están diseñados para tomar decisiones éticas; no usan algoritmos basados en la ética. Siempre harán predicciones basadas en cómo funciona el mundo hoy, contribuyendo así a fomentar los prejuicios y las prácticas discriminatorias arraigadas sistemáticamente en nuestras sociedades actuales. Con el uso generalizado de las tecnologías de IA y ML, a menudo propiedad de grandes empresas tecnológicas con el único objetivo de obtener beneficios existe una necesidad urgente de aportar un enfoque centrado en el ser humano y utilizarlo para resolver problemas sociales en lugar de contribuir a ellos. En su Comunicación de 25/04/18 y 7/12/18, la CE expuso su visión de la inteligencia artificial, que apoya "una IA ética, segura y de vanguardia fabricada en Europa". Los sistemas de IA deben estar centrados en el ser humano y basarse en el compromiso de su uso al servicio de la humanidad y el bien común para mejorar el bienestar y la libertad humanos. Si bien ofrecen oportunidades significativas, los sistemas de IA también dan lugar a ciertos riesgos que deben manejarse de manera adecuada y proporcionada. Ahora tenemos una ventana de oportunidad esencial para dar forma a su desarrollo.

CHARLIE tiene la intención de garantizar que podamos confiar en los entornos sociotécnicos en los que están insertos. También queremos que los productores de sistemas de IA obtengan una ventaja competitiva mediante la incorporación de IA fiable en sus productos y servicios. Esto implica tratar de maximizar los beneficios de los sistemas de IA y, al mismo tiempo, prevenir y minimizar sus riesgos. La Educación Superior (ES), la Educación de Adultos (EA) y la Juventud requieren planes de estudio nuevos e innovadores que puedan cubrir esta brecha de habilidades y equipar a los estudiantes con el conocimiento y las habilidades para contribuir a un enfoque más ético del desarrollo tecnológico. La necesidad de hacer que la educación tecnológica sea más humana está alineada con el Plan de Acción de Educación Digital 2021-27, que incluye acciones específicas para abordar las implicaciones éticas y los desafíos del uso de la IA y los datos en la educación y la formación.

1.2 Objetivos del proyecto

CHARLIE tiene como objetivo desafiar el sesgo en el *big data* utilizado para la IA y el aprendizaje automático al brindar un mayor nivel de conciencia sobre los impactos negativos de la falta de un enfoque crítico y ético de la educación tecnológica. Los objetivos principales de CHARLIE son:

1. Aumentar la capacidad de las instituciones de educación superior para ofrecer a sus estudiantes oportunidades de aprendizaje en línea que satisfagan las necesidades de la sociedad, pero que también se adapten a las necesidades de aprendizaje de los estudiantes;
2. Aumentar las competencias sociales y éticas de los estudiantes de tecnología, permitiéndoles comprometerse de manera positiva, crítica y ética con la tecnología de IA/ML;
3. Equipar a los formadores y profesores con enfoques digitales atractivos para la enseñanza efectiva del tema (especialmente en la enseñanza *online*);
4. Crear sinergias entre las organizaciones de educación superior, de adultos y jóvenes en el ámbito de la educación ética en materia de IA;
5. Potenciar la transferibilidad de los cursos académicos sobre los sesgos de la IA a la educación y formación profesional (FPE);
6. Sensibilizar sobre el tema en la sociedad.

1.3. Resultados globales esperados del proyecto Charlie por WP

1- Matrices de competencias para el curso de "Sesgo algorítmico" (EQF6), la "Microcredencial de IA ética" (EQF4) y el *Serious Game* (EQF2) (**Work Package o WP2)**

2- Curso "Sesgo algorítmico" (Ed. Sup.) - enfoque bLearning: a) sesiones sincrónicas (preferiblemente presenciales), b) sesiones asincrónicas de eLearning para el componente teórico, complementadas con un *serious game* digital como evaluación formativa y evaluación sumativa en línea, entregadas como .scorm listas para ser cargadas en Cualquier LMS para estudiantes inscritos en cursos relacionados con *Big Data, IA, machine learning, deep learning*. **(WP3)**

3- " Kit de herramientas *Algorithmic Bias* para sesiones síncronas" - para Profesores y formadores que implementan las sesiones en vivo (presenciales u online) que complementan las sesiones grabadas de eLearning. **(WP3)**

4- Una **directriz** para impulsar la capacidad de los **administradores universitarios y en puestos directivos de los departamentos de Educación y Formación** a cargo de la oferta de educación digital para fomentar la impartición de educación digital sobre el sesgo algorítmico. **(WP3)**

5- Seminarios web para fomentar el aprendizaje entre pares y discutir el papel de las instituciones de educación superior en la educación tecnológica, para fomentar la interdisciplinariedad de las ciencias sociales en la educación tecnológica y para discutir enfoques para la interoperabilidad con los jóvenes y la educación sostenible. **(WP3)**

6- Microcredencial de "IA ética" (EQF4) para estudiantes adultos que se realizará online y de forma asíncrona. **(WP4)**

7- Juego serio digital (EQF2) para jóvenes de 12 a 18 años (principalmente mujeres jóvenes desfavorecidas) para fomentar la representación de género en STEM desde una edad temprana. **(WP4)**

8- Kit de herramientas para apoyar a adultos y jóvenes en la mejora de las habilidades hacia la IA ética. **(WP4)**

9- Recomendaciones políticas para el reconocimiento de la micro credencial para estudiantes adultos en el acceso a los cursos de educación superior en los campos tecnológicos. **(WP4).**

1.4. Grupos objetivo de Charlie

Los principales grupos destinatarios del proyecto CHARLIE son:

- **Instituciones de educación superior** que proporcionan oportunidades/cursos de Big Data, IA, aprendizaje automático, aprendizaje profundo relacionados con el aprendizaje profundo y sus respectivos administradores/gestión de los departamentos de Educación y Formación
- **Estudiantes de educación superior** inscritos en cursos relacionados con Big Data, IA, aprendizaje automático y aprendizaje profundo
- **Maestros/Profesores/Conferenciantes** de Ciencias Sociales y Educación Tecnológica
- **Organizaciones de EA** y su personal
- **Adultos** de todas las edades y niveles socioeconómicos, con el objetivo de progresar hacia niveles de cualificación más altos y relevantes para el mercado laboral y para la participación en la sociedad.
- **Organizaciones juveniles** y su personal
- -**Jóvenes (**de 12 a 18 años) - especialmente mujeres jóvenes y jóvenes de contextos socialmente desfavorecidos (que se enfrentan a barreras relacionadas con el sistema educativo; diferencias culturales; barreras sociales; barreras económicas; barreras vinculadas a la discriminación; y barreras geográficas)

Principales destinatarios de las actividades y productos de comunicación y promoción de proyectos:

- Reguladores y responsables políticos nacionales;
- El público en general
- Instituciones de Educación Superior (profesores universitarios, estudiantes (pregrado y posgrado) y administradores Universitarios.
- Proveedores de educación para adultos
- Organizaciones juveniles
- Proveedores de FP
- Foros y plataformas de IA
- ONGs que abordan diferentes sesgos y discriminación
- Proveedores de tecnología educativa
- Operadores de medios de comunicación (TV, periódicos electrónicos, radio, etc.)
- Pymes y empresas de IA/ML.
- Asociaciones de proveedores de educación superior, incluidas las organizaciones de colaboración relacionadas para la promoción de las mejores

prácticas a nivel de la UE y la difusión a sus miembros universitarios (por ejemplo: Asociación Europea de Universidades, Red.es, Asociación de la Triple Hélice, AMBA, Red de Innovación de la industria universitaria, Red Europea de *Business Angels* (para invertir en innovaciones éticas de IA)
- Alumnado de toda la UE
- Asociaciones de monitoreo/desarrollo de habilidades (por ejemplo, comunidades *DigCompEdu).*

1.5. Objetivos del WP2

Las Matrices de Competencias para el "Sesgo Algorítmico" de los objetivos específicos del WP son:

- Establecer un entendimiento común sobre el propósito y los objetivos de una ruta de aprendizaje que se desarrollará para los estudiantes de educación superior en el curso "Sesgo algorítmico";
- Establecer un entendimiento común sobre el propósito y los objetivos de aprendizaje de un itinerario de aprendizaje que se desarrollará para los alumnos de EA en la "Micro credencial de IA ética";
- Establecer un entendimiento común sobre el propósito y los objetivos de aprendizaje de un itinerario de aprendizaje que se desarrollará para los jóvenes en el juego serio de la IA ética.

Estos objetivos contribuyen a los objetivos generales al sentar las bases de las siguientes actividades organizadas en el marco de los distintos paquetes de trabajo, permitiendo a los docentes, mentores y alumnos tener una visión clara de los objetivos de los diferentes itinerarios de aprendizaje, trabajando para aumentar la capacidad de las organizaciones de educación superior, de adultos y de jóvenes para ofrecer oportunidades de aprendizaje que satisfagan las necesidades de la sociedad, pero que también se adapten a las necesidades de aprendizaje de los alumnos.

1.6. Resultados esperados del WP2 y de los participantes

1.6.1. Matriz de competencias EQF6 para el curso "Sesgo algorítmico" (estudiantes de educación superior) - Enfoque de resultados de aprendizaje, 2 puntos ECTS, recomendaciones de acreditación, requisitos previos para la inscripción, horas de contacto, carga de trabajo total, integración de competencias EntreComp (por ejemplo: Pensamiento ético y sostenible), competencias DigComp 2.0 (por ejemplo: protección de la salud y el bienestar) y competencias GreenComp (por ejemplo: apoyo a la equidad). Unidades de Competencia:
CU1 - Algoritmos, Modelos y Limitaciones UIB
CU2 - Equidad y sesgo de los datos en AI UA
CU3 - Privacidad y comodidad
CU4 - Ética de la IA, un enfoque práctico VAMK
CU5 - Casos de estudio y proyecto VAMK

1.6.2. Matriz de competencias EQF 4 para la " Micro credencial de IA ética" Enfoque de resultados de aprendizaje, potencial de puntos ECVS, recomendaciones de acreditación, requisitos previos para la inscripción, horas de contacto, carga de trabajo total, integración de las competencias EntreComp (p. ej.: pensamiento ético y sostenible), competencias DigComp 2.0 (p. ej., protección de la salud y el bienestar) y competencias GreenComp (p. ej., apoyo a la equidad). Unidades de Competencia:
CU1 - ¿Qué es el sesgo algorítmico?
CU2 - No maleficiencia
CU3 – Rendición de cuentas
CU4 – Transparencia
CU5 - Derechos humanos y equidad
CU6 - Ética de la IA, un enfoque práctico

1.6.3. Resultados de aprendizaje para jóvenes EQF 2 (juego serio) - Con enfoque de resultados de aprendizaje, requisitos previos para la inscripción, horas de contacto, integración de las competencias EntreComp (por ejemplo: Pensamiento ético y sostenible), competencias DigComp 2.0 (por ejemplo: protección de la salud y el bienestar) y competencias GreenComp (por ejemplo: apoyo a la equidad). Contenidos generales: desigualdades sistémicas en la sociedad, mecánica básica de los sistemas de inteligencia artificial, agendas políticas, impactos de la IA en el mundo.

2. INTRODUCCIÓN AL TEMA: ALGORITMOS, IA, SESGOS Y ÉTICA

Los algoritmos juegan un papel crucial en la sociedad moderna, ya que ayudan con diversos servicios como recomendar canciones, películas o amigos (Paraschakis, 2018; Milano et al., 2020). También se utilizan en escuelas, hospitales, instituciones financieras, tribunales y gobiernos para tomar decisiones cruciales (Obermeyer et al., 2019). Sin embargo, el uso de algoritmos puede conllevar riesgos éticos (Floridi et al., 2018).

Los algoritmos, como las traducciones (Prates et al., 2020) o los anuncios de empleo (Lambrecht y Tucker, 2019), pueden estar sesgados. Por ejemplo, los trabajos mejor pagados pueden mostrarse más a los hombres que a las mujeres. Esto también puede conducir a un trato injusto en la atención médica, ya que los pacientes blancos reciben una mejor atención que los pacientes negros (Obermeyer et al. 2019). Mientras la gente intenta solucionar estos problemas, siguen surgiendo más problemas.

Desde 2012, se ha prestado mucha atención a la inteligencia artificial (IA) y sus implicaciones éticas (Perrault et al. 2019). En 2016, un estudio intentó mapear estas preocupaciones (Mittelstadt et al. 2016), pero el campo ha crecido y cambiado mucho desde entonces. Los gobiernos, las organizaciones y las empresas se han involucrado más en el debate sobre la IA y los algoritmos "justos" y "éticos" (Wong, 2019).

Los sesgos en la tecnología pueden tener efectos nocivos, incluso si no son intencionados, y crear desafíos para generar confianza pública en la IA. Debemos tener en cuenta los factores específicos de la IA y sus posibles impactos en la sociedad para garantizar que sea segura y protegida. Los esfuerzos actuales para abordar el sesgo de la IA se centran en aspectos computacionales como la representatividad de los datos y la equidad en los algoritmos de aprendizaje automático. Sin embargo, los factores humanos, institucionales y sociales también contribuyen al sesgo de la IA y a menudo se pasan por alto. Para abordar con éxito este desafío, **debemos ampliar nuestra perspectiva y examinar cómo se crea la IA y cómo afecta a la sociedad**.

Al analizar si una IA es fiable y responsable no se trata solo de si verificar un sistema de IA es sesgado, justo o ético, sino también de si hace lo que dice. Prácticas como la transparencia, los conjuntos de datos y las pruebas, evaluaciones, validaciones y verificaciones son esenciales. Los factores humanos, las técnicas de diseño participativo, los enfoques de múltiples partes interesadas y el *human-in-the-loop* (o HITL, modelo que requiere interacción humana) también ayudan a mitigar los riesgos de sesgo de la IA. Sin embargo, estas prácticas por sí solas no proporcionan una solución completa. Necesitamos orientación desde una perspectiva sociotécnica más amplia que conecte estas prácticas con los valores sociales.

Los expertos en IA fiable y responsable recomiendan poner en práctica estos valores y crear nuevas normas de desarrollo e implementación de la IA. Este documento, junto con el trabajo realizado por los socios del proyecto CHARLIE sobre los sesgos de los algoritmos y la IA, adopta una perspectiva sociotécnica.

El sesgo no es exclusivo de la IA, y lograr un riesgo cero de sesgo en los sistemas de IA es imposible. El proyecto **CHARLIE tiene la intención de desarrollar recursos y métodos para aumentar las mejoras en la práctica para identificar, comprender, medir, gestionar y reducir el sesgo**. Para alcanzar este objetivo, necesitamos técnicas flexibles que puedan aplicarse en todos los contextos y comunicarse a los diferentes grupos de interés.

3. MATRIZ DE COMPETENCIAS: UN ENFOQUE TEÓRICO

Una matriz de competencias, también conocida como matriz de habilidades o marco de competencias, es una herramienta utilizada para identificar, mapear y evaluar las habilidades, conocimientos y habilidades de los individuos dentro de una organización o institución educativa (Stewart y Bartrum, 1999). Representa visualmente las competencias requeridas para roles, proyectos, tareas o programas académicos específicos. Ayuda a identificar los niveles actuales de habilidades de los empleados, los miembros del equipo, los estudiantes u otras partes interesadas (Cottrell, 2018).

Características de una Matriz de Competencias

Una matriz de competencias se caracteriza por varias características clave:

- Formato de tabla: Una matriz de competencias suele estar estructurada como una tabla, con filas que representan a los individuos y columnas que representan las competencias o habilidades requeridas (Mansfield, 2009).
- Calificaciones o niveles de competencia: Cada celda de la matriz contiene una calificación o nivel de competencia para la habilidad y el individuo correspondientes. El sistema de calificación puede ser numérico, codificado por colores o basado en términos descriptivos como "principiante", "intermedio" o "experto" (Cottrell, 2018).
- Personalizables y adaptables: Las matrices de competencias se pueden personalizar para adaptarse a las necesidades específicas de una organización o institución educativa, y se pueden actualizar fácilmente para reflejar los cambios en los roles, las responsabilidades o los requisitos de habilidades (Stewart y Bartrum, 1999).
- Integral: Una matriz de competencias debe cubrir todas las competencias relevantes requeridas para un rol, proyecto o programa académico en particular, y debe incluir tanto habilidades técnicas como blandas (Mansfield, 2009).

Objetivos principales de una matriz de competencias

Una matriz de competencias sirve a varios objetivos importantes dentro de una organización o institución educativa:

- Identificación de brechas de habilidades: Al proporcionar una visión clara de las habilidades y competencias requeridas para un rol o proyecto en particular, una matriz de competencias puede ayudar a identificar áreas en las que las personas pueden necesitar un mayor desarrollo o capacitación (Stewart y Bartrum, 1999).
- Apoyar el desarrollo de empleados o estudiantes: Se puede utilizar una matriz de competencias para guiar los planes de desarrollo personal e informar las iniciativas de capacitación y desarrollo, ayudando a las personas a identificar sus fortalezas y debilidades, y establecer metas de mejora (Cottrell, 2018).
- Optimización de la asignación de recursos: Una matriz de competencias puede ayudar a las organizaciones e instituciones educativas a asignar recursos de manera más efectiva, mediante la identificación de las personas más apropiadas para roles o proyectos específicos en función de sus habilidades y competencias (Mansfield, 2009).
- Mejorar el desempeño: Al ayudar a garantizar que las personas tengan las habilidades y competencias adecuadas para sus funciones, una matriz de competencias puede contribuir a mejorar el desempeño general dentro de una organización o institución educativa (Stewart y Bartrum, 1999).
- Facilitar la comunicación y la colaboración: Una matriz de competencias puede servir como un lenguaje común para discutir habilidades y competencias dentro de una organización o institución educativa, promoviendo una mayor comprensión y colaboración entre los miembros del equipo, los departamentos o las unidades académicas (Cottrell, 2018).

Aplicaciones de la Matriz de Competencias en Diferentes Contextos

Las matrices de competencias se han aplicado en diversos contextos, incluidos los negocios, el gobierno y las instituciones de educación superior, para lograr una serie de objetivos:

- Planificación de la fuerza laboral: Las matrices de competencias se pueden utilizar para apoyar la planificación estratégica de la fuerza laboral, identificando las habilidades y competencias necesarias para alcanzar las metas y objetivos organizacionales, y asegurando que estos se reflejen en las iniciativas de reclutamiento, capacitación y desarrollo (Stewart y Bartrum, 1999).
- Gestión del desempeño: Las matrices de competencias pueden integrarse en los sistemas de gestión del desempeño, proporcionando un marco para establecer expectativas de desempeño, realizar evaluaciones de desempeño e identificar áreas de mejora (Mansfield, 2009).
- Planificación de la sucesión: Las matrices de competencias pueden apoyar los esfuerzos de planificación de la sucesión, identificando las habilidades y competencias necesarias para los roles de liderazgo y ayudando a identificar y desarrollar posibles sucesores dentro de una organización o institución educativa (Cottrell, 2018).
- Planificación y desarrollo curricular: En las instituciones de educación superior, las matrices de competencias se pueden utilizar para guiar la planificación y el desarrollo del currículo, asegurando que los programas académicos estén diseñados para desarrollar las habilidades y competencias requeridas por los estudiantes para tener éxito en los campos elegidos (Billett, 2009). Ésta, y las tres siguientes, son la orientación principal del presente documento.
- Desarrollo del profesorado y del personal: Las matrices de competencias pueden aplicarse para apoyar el desarrollo profesional del profesorado y el personal de las instituciones de educación superior, identificando las habilidades y competencias necesarias para desempeñar funciones eficaces en la enseñanza, la investigación y la administración, e informando el diseño de programas de formación y desarrollo (Eraut, 2004).
- Formación de equipos de investigación: Las matrices de competencias pueden utilizarse para facilitar la formación de equipos de investigación en instituciones de educación superior, identificando individuos con habilidades y competencias complementarias y promoviendo la colaboración interdisciplinaria (Börner et al., 2010).
- Asesoramiento y apoyo a los estudiantes: Las matrices de competencias pueden utilizarse para apoyar los servicios de asesoramiento y apoyo a los estudiantes en las instituciones de educación superior, ayudando a los asesores

a identificar las fortalezas y debilidades de los estudiantes y guiarlos hacia los recursos, las intervenciones o las trayectorias académicas adecuadas (Cuaeo, 2010).

Desafíos y limitaciones de la matriz de competencias

Si bien las matrices de competencias pueden proporcionar información valiosa y respaldar una serie de objetivos estratégicos, también presentan algunos desafíos y limitaciones:

- Subjetividad: Las evaluaciones de competencias pueden ser subjetivas y pueden estar influenciadas por factores como sesgos personales, diferencias culturales o variaciones en la interpretación de las definiciones de competencias (Stewart y Bartrum, 1999).
- Uso intensivo de tiempo y recursos: El desarrollo y mantenimiento de una matriz de competencias integral puede requerir mucho tiempo y recursos, especialmente en grandes organizaciones o instituciones educativas con diversas funciones y responsabilidades (Mansfield, 2009).
- Énfasis excesivo en las habilidades técnicas: Las matrices de competencias pueden poner mayor énfasis en las habilidades y competencias técnicas, a expensas de las habilidades blandas, como la comunicación, el trabajo en equipo o la resolución de problemas, que a menudo son fundamentales para el éxito en muchos roles y contextos (Cottrell, 2018).
- Rigidez: Las matrices de competencias pueden ser percibidas como rígidas e inflexibles, lo que puede limitar la creatividad, la innovación o la adaptabilidad dentro de una organización o institución educativa (Eraut, 2004).

A pesar de estos desafíos y limitaciones, las matrices de competencias pueden servir como una herramienta valiosa para las organizaciones e instituciones educativas que buscan optimizar la asignación de recursos, apoyar el desarrollo individual y mejorar el rendimiento general. Al adoptar un enfoque sistemático e integral para identificar, mapear y evaluar las habilidades y competencias, e incorporar una variedad de perspectivas y partes interesadas en el desarrollo y mantenimiento de la matriz, las organizaciones e instituciones educativas pueden maximizar los beneficios de este enfoque al tiempo que minimizan los posibles inconvenientes (Billett, 2009).

4. MATRICES CHARLIE WP2

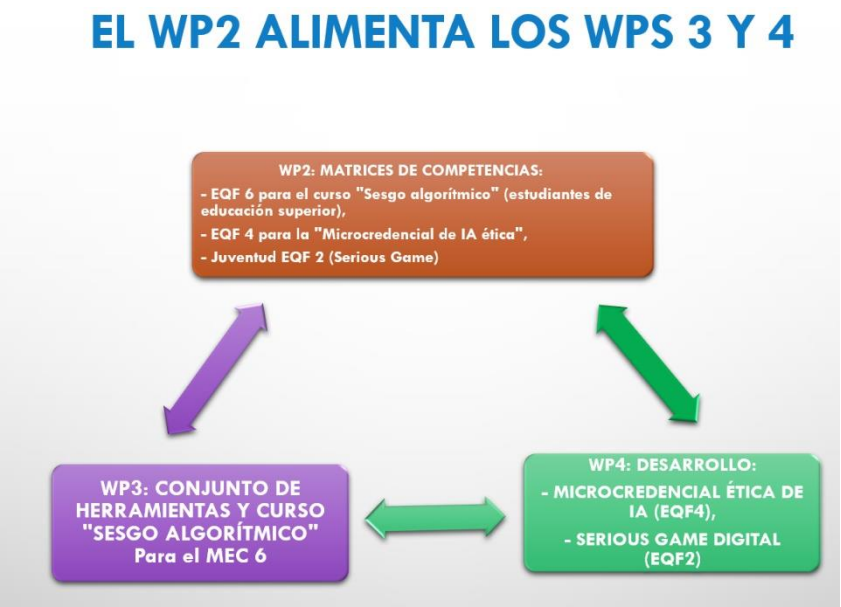
La matriz de competencias del curso "Sesgo algorítmico" del EQF6 sirve como elemento fundamental para otros dos componentes:

- la matriz de competencias del EQF 4 "Microcredencial ética de la IA" para adultos
- y los resultados de aprendizaje del EQF 2 (juego serio) para jóvenes

Estos tres componentes están interconectados y diseñados para complementarse y reforzarse entre sí, garantizando una comprensión y aplicación integrales de los principios éticos de la IA en diversos niveles y contextos educativos.

La matriz de competencias del curso "Sesgo algorítmico" sienta las bases para el desarrollo de la matriz de competencias EQF4 "Micro credencial de AI ética". Al proporcionar una base sólida para comprender los sesgos algorítmicos y sus implicaciones, este curso permite a los alumnos profundizar en las dimensiones éticas de la IA a medida que avanzan hacia el nivel EQF4. La matriz de competencias del EQF4 "Micro credencial ética de la IA" se basa en las competencias iniciales, ampliando el alcance para incluir consideraciones éticas más amplias y aplicaciones prácticas de los principios éticos de la IA en diversos sectores y escenarios.

Al mismo tiempo, la matriz de competencias del curso "Sesgo algorítmico" EQF6 también informa los resultados de aprendizaje de "Resultados de aprendizaje para adultos y jóvenes EQF2 (*Serious Game*)". Esta conexión garantiza que los conceptos y habilidades fundamentales desarrollados en el curso sean accesibles y atractivos para los alumnos del nivel EQF2. El formato de juego serio está diseñado para presentar estos temas críticos de una manera entretenida y educativa, fomentando una comprensión más profunda del sesgo algorítmico y los principios éticos de la IA entre un público más amplio.



Gráfica 1. Interconexión entre los WP del proyecto y los resultados

En conclusión, los tres elementos, la matriz de competencias del curso EQF6 "Sesgo algorítmico", la matriz de competencias del EQF4 "Micro credencial ética de la IA" y los resultados de aprendizaje "Resultados de aprendizaje para adultos y jóvenes del EQF2 (*Serious Game*)", están vinculados. Esta estructura interconectada permite un enfoque integral y multinivel para aprender sobre el sesgo algorítmico y la IA ética, lo que garantiza que los alumnos de diversos orígenes educativos y grupos de edad puedan desarrollar los conocimientos y las habilidades necesarios para navegar por los desafíos y oportunidades de la IA de manera ética y responsable. Además, los resultados del WP2 se incorporarán directamente al WP3 y al WP4, lo que ilustra aún más la estrecha relación entre estos componentes, como se muestra en el Gráfico 1.

Esta perfecta integración de WP2, WP3 y WP4 da como resultado un marco educativo sólido y cohesivo que aborda los aspectos cruciales del sesgo algorítmico y la IA ética, fomentando una comprensión y aplicación completas de estos principios en diversos contextos y perfiles de aprendizaje.

4.1. *CURSO "SESGO ALGORÍTMICO"* EQF6

Este programa educativo está diseñado para **proporcionar una comprensión integral de los sesgos algorítmicos y sus implicaciones en la sociedad actual impulsada por los datos**. Dado que los algoritmos desempeñan un papel importante en la configuración de nuestras vidas y en la toma de decisiones en diversos sectores, incluidos la atención médica, la educación, las finanzas y el gobierno, es crucial que los profesores, profesionales y estudiantes desarrollen una comprensión sólida de los posibles sesgos dentro de estos sistemas.

El objetivo de este curso es dotar a los participantes de los conocimientos y habilidades necesarios para identificar, mitigar y abordar los sesgos algorítmicos. A través de una combinación de fundamentos teóricos y aplicaciones prácticas, los alumnos obtendrán información sobre los aspectos éticos, sociales y técnicos del sesgo algorítmico.

Estos objetivos contribuyen a los objetivos generales al sentar las bases de las siguientes actividades, permitiendo a los docentes, mentores y educandos tener una visión clara de los objetivos de los diferentes itinerarios de aprendizaje, trabajando para aumentar la capacidad de las organizaciones de educación superior, adultos y jóvenes para proporcionar oportunidades de aprendizaje que satisfagan las necesidades de la sociedad, pero que también se adapten a las necesidades de aprendizaje de los educandos. Concretamente, los resultados esperados de este curso están dirigidos a:

1) aumentar la capacidad de las instituciones de educación superior para ofrecer a sus estudiantes oportunidades de aprendizaje en línea que satisfagan las necesidades de la sociedad, pero que también se adapten a las necesidades de aprendizaje de los estudiantes, y que las instituciones de educación superior se beneficien del enfoque de resultados de aprendizaje, en el que profesores y estudiantes tienen un terreno común con respecto a los resultados esperados al final de una ruta de aprendizaje;
2) aumentar las competencias sociales y éticas de los estudiantes de tecnología, lo que les permite comprometerse de manera positiva, crítica y ética con la tecnología de IA/ML, y los estudiantes de educación superior se benefician del enfoque de resultados de aprendizaje en el que saben claramente lo que se espera de ellos al final de la ruta de aprendizaje;

3) equipar a los maestros/profesores con enfoques digitales y atractivos para la enseñanza efectiva del tema (especialmente en la enseñanza en línea), ya que el enfoque del alumno es el punto de partida para el diseño de las experiencias de aprendizaje;
4) crear sinergias entre las instituciones de educación superior y las actividades agroambientales y juveniles a nivel regional y local en el ámbito de la educación ética en materia de IA, mediante el establecimiento de resultados de aprendizaje para REA complementarios dirigidos a diferentes niveles (EQF 6, 4, 2) en la misma área de conocimiento;
5) potenciar la transferibilidad de los cursos académicos sobre los sesgos de la IA para adultos y jóvenes, estableciendo un punto de contacto en los diferentes resultados como base para un terreno común para la discusión sobre el potencial de transferibilidad de las vías de aprendizaje;
6) sensibilizar sobre el tema a nivel de la sociedad, ya que establece el primer paso para llevar los recursos de aprendizaje dedicados a los diferentes niveles de la sociedad y a los grupos destinatarios.

Al finalizar este curso, los participantes habrán desarrollado una comprensión integral de los sesgos algorítmicos, su impacto potencial en varios sectores y las estrategias y herramientas disponibles para abordar estos sesgos. Este conocimiento será invaluable para profesionales y académicos/estudiantes que trabajan en campos donde se utilizan algoritmos para la toma de decisiones y ayudará a garantizar resultados más equitativos y justos en un mundo impulsado por datos.

4.1.1. Contenido del curso "Sesgo algorítmico"

El curso está estructurado en torno a cinco Unidades de Competencia (UC) principales, cada una diseñada para equipar a los participantes con el conocimiento y las habilidades necesarias para navegar por los desafíos y oportunidades en el ámbito del sesgo algorítmico.

CU1 - Modelos y limitaciones de los algoritmos: esta unidad presenta 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 derivados de su uso. Los temas que se tratarán incluirán la complejidad algorítmica, la optimización y las limitaciones de la toma de decisiones algorítmicas.

CU2 - Equidad y sesgo de los datos en la IA: esta unidad profundizará en los conceptos de equidad y sesgo en los sistemas de IA, explorando las diversas fuentes y manifestaciones del sesgo en los procesos algorítmicos. Los participantes aprenderán sobre los métodos para identificar y medir los sesgos en los datos, así como las estrategias para abordar y mitigar estos sesgos para garantizar resultados más equitativos.

CU3 - Privacidad y conveniencia de la IA: en esta unidad, los participantes examinarán las compensaciones entre la privacidad y la conveniencia en las aplicaciones de IA. Las discusiones abarcarán los desafíos de mantener la privacidad del usuario mientras se brindan servicios personalizados y eficientes, así como las tecnologías existentes y emergentes diseñadas para equilibrar estos intereses contrapuestos.

CU4 - Ética de la IA, un enfoque práctico: esta unidad se centrará en la aplicación práctica de los principios éticos en el desarrollo y la implementación de la IA. Los participantes explorarán varios marcos y pautas éticas, aprendiendo a aplicarlos a escenarios del mundo real que involucran sistemas de IA. La unidad también abordará la importancia de la transparencia, la rendición de cuentas y la participación de las partes interesadas en el desarrollo ético de la IA.

CU5 - Estudios de casos y proyectos: la unidad final brindará a los participantes la oportunidad de aplicar sus conocimientos y habilidades a estudios de casos del mundo real y a un proyecto práctico. A través de estas aplicaciones prácticas, los alumnos obtendrán una comprensión más profunda de las complejidades y matices del sesgo algorítmico, así como de las estrategias y herramientas disponibles para abordar estos problemas.

4.1.2. Descripción del grupo destinatario, características del EQF6

El Marco Europeo de Cualificaciones (EQF) es un marco de referencia común que vincula los sistemas de Cualificaciones de diferentes países europeos, lo que facilita que las personas reconozcan sus Cualificaciones en toda Europa (Comisión Europea, 2023). El EQF consta de ocho niveles de referencia, cada uno de los niveles está definido por un conjunto de descriptores que indican los resultados del aprendizaje, las capacidades, las competencias y la autonomía que se esperan en ese nivel. El nivel 6 del EQF (EQF6) corresponde a las cualificaciones obtenidas a nivel de grado o de formación profesional

superior en el Espacio Europeo de Educación Superior (EEES). Las características y características del EQF6 incluyen (Comisión Europea, 2023; CEDEFOP, 2023):

- **Conocimientos avanzados**: Las calificaciones EQF6 requieren que los alumnos posean conocimientos avanzados de un campo específico, demostrando una comprensión crítica de teorías, principios y métodos. Este nivel de conocimiento es necesario para desarrollar ideas originales y resolver problemas complejos.
- **Habilidades cognitivas**: En el EQF6, los alumnos deben desarrollar la capacidad de aplicar sus conocimientos a situaciones nuevas o desconocidas, analizar problemas complejos y sintetizar información de diversas fuentes. Esto incluye el pensamiento crítico, la resolución de problemas, la toma de decisiones y las habilidades de pensamiento creativo.
- **Habilidades prácticas**: Las Cualificaciones del EQF 6 implican el desarrollo de habilidades prácticas avanzadas, que permiten a los alumnos gestionar actividades, proyectos o procesos técnicos o profesionales complejos. Estas habilidades pueden incluir la investigación, la gestión de proyectos o la capacidad de utilizar herramientas y técnicas especializadas de manera efectiva.
- **Autonomía y responsabilidad**: Se espera que los alumnos del EQF6 demuestren un alto grado de autonomía y responsabilidad en su trabajo. Esto implica asumir la responsabilidad de su propio aprendizaje y desarrollo profesional, gestionar y supervisar el trabajo de los demás y tomar decisiones informadas basadas en consideraciones éticas y sociales.
- **Comunicación y colaboración**: Las Cualificaciones del EQF 6 requieren que los alumnos desarrollen habilidades efectivas de comunicación y colaboración, lo que les permite presentar información, argumentos o propuestas claras y detalladas tanto a audiencias especializadas como no especializadas. Esto incluye habilidades de comunicación escrita, oral e interpersonal, así como la capacidad de trabajar eficazmente en equipo y en todas las disciplinas.
- **Aprendizaje a lo largo de toda la vida:** En el EQF 6, los educandos deben desarrollar la capacidad de participar en el aprendizaje a lo largo de toda la vida, reflexionando sobre sus propias experiencias de aprendizaje e identificando áreas de mayor desarrollo. Esto incluye la capacidad de adaptarse a nuevas situaciones, tecnologías o entornos profesionales y de participar en el desarrollo profesional continuo.

Tratando de traducir las características del EQF6 al tema del proyecto CHARLIE, estos son algunos ejemplos de competencias sobre sesgo algorítmico adaptadas al nivel del EQF6:

- **Conocimientos avanzados:** Los alumnos deben tener un conocimiento profundo del sesgo algorítmico, sus fuentes y su impacto en los sistemas de IA/ML. Deben estar familiarizados con los diferentes tipos de sesgos, las métricas de equidad y las implicaciones éticas de los algoritmos sesgados.
- **Habilidades cognitivas:** Los alumnos deben ser capaces de identificar y analizar casos de sesgo algorítmico en situaciones del mundo real, evaluar críticamente la equidad de los sistemas de IA/ML y proponer métodos para mitigar o eliminar el sesgo.
- **Habilidades prácticas:** Los alumnos deben desarrollar la capacidad de implementar técnicas para detectar y mitigar el sesgo algorítmico, como el remuestreo, la reponderación o el entrenamiento de adversarios. También deben ser capaces de evaluar la eficacia de estas técnicas para reducir el sesgo y mejorar la equidad.
- **Autonomía y responsabilidad:** Los alumnos deben ser capaces de tomar decisiones informadas sobre el uso ético de los sistemas de IA/ML, teniendo en cuenta los posibles riesgos y beneficios asociados al sesgo algorítmico. También deben ser capaces de asumir la responsabilidad del impacto de sus decisiones en los individuos, las comunidades y la sociedad en su conjunto.
- **Comunicación y colaboración:** Los alumnos deben ser capaces de comunicar su comprensión del sesgo algorítmico y sus implicaciones tanto a audiencias técnicas como no técnicas. También deben ser capaces de colaborar eficazmente con equipos interdisciplinarios para abordar cuestiones de equidad, transparencia y rendición de cuentas en los sistemas de IA/ML.
- **Aprendizaje permanente:** Los alumnos deben comprometerse a mantenerse actualizados con las últimas investigaciones, métodos y herramientas relacionadas con el sesgo algorítmico, la equidad y la ética en IA/ML. Deben ser capaces de adaptar sus conocimientos y habilidades para hacer frente a los nuevos retos y oportunidades sobre el terreno.

En resumen, las Cualificaciones del EQF6 se caracterizan por conocimientos avanzados, habilidades cognitivas y prácticas, autonomía y responsabilidad, comunicación y colaboración efectivas, y un compromiso con el aprendizaje permanente. Estas características y características garantizan que los graduados de este nivel estén bien equipados para tener éxito en la profesión elegida o continuar su educación en niveles superiores, como cursar un máster (EQF 7) o un doctorado (EQF 8).

4.1.3. Matriz de competencias generales para el curso "Sesgo algorítmico" EQF6

La tabla a continuación presenta la matriz de competencias general para el curso "Sesgo algorítmico" con un equivalente EQF6:

RESULTADOS DE APRENDIZAJE		
Al finalizar la experiencia de formación, el alumno será capaz de:		
CONOCIMIENTO	**HABILIDADES**	**COMPETENCIAS**
1. Asociar la naturaleza interdisciplinaria de la IA y el papel de los algoritmos en la configuración de diversos aspectos de la sociedad, la economía y la tecnología. 2. Obtenga una comprensión integral de las implicaciones éticas, legales y sociales de los sistemas de IA, incluidas las cuestiones relacionadas con la equidad, la privacidad y la conveniencia. 3. Desarrollar habilidades de pensamiento crítico y resolución de problemas para identificar, analizar y abordar el sesgo algorítmico y otros desafíos éticos en la IA. 4. Evaluar la eficacia de diversos enfoques para mitigar el sesgo algorítmico y promover la equidad en los sistemas de IA, teniendo en cuenta las limitaciones y compensaciones. 5. Comprender la importancia de un enfoque centrado en el usuario en el diseño y la implementación de sistemas de IA que respeten la privacidad, la autonomía y las preferencias individuales. 6. Reconocer la importancia de la colaboración interdisciplinaria y la participación de las partes interesadas en el desarrollo y la implementación de soluciones éticas de IA. 7. Desarrollar una base sólida en teorías y marcos éticos aplicables a la IA, y aprender a aplicar estos principios en escenarios del mundo real. 8. Apreciar la importancia de la transparencia, la	1. Analizar y evaluar algoritmos y modelos de IA para identificar posibles sesgos y limitaciones y proponer mejoras o alternativas cuando sea necesario. 2. Aplicar diversos métodos y técnicas para identificar, medir y mitigar los sesgos en los datos y los sistemas de IA, garantizando la equidad y la equidad de los resultados. 3. Equilibre las compensaciones entre la privacidad y la comodidad en las aplicaciones de IA, tomando decisiones informadas para proteger la privacidad del usuario y ofreciendo servicios personalizados y eficientes. 4. Implementar principios, marcos y directrices éticas en el diseño, desarrollo e implementación de sistemas de IA, garantizando el cumplimiento de las normas de transparencia, rendición de cuentas y participación de las partes interesadas. 5. Evaluar críticamente los estudios de casos y escenarios del mundo real que involucran sistemas de IA, identificando desafíos éticos y proponiendo soluciones basadas en los conocimientos y habilidades adquiridos en el curso. 6. Colaborar eficazmente con equipos interdisciplinarios, demostrando una comprensión de las diversas perspectivas y conocimientos necesarios	1. Reconocer y abordar de forma independiente los problemas éticos y los sesgos en los sistemas de IA, asumiendo la responsabilidad de garantizar la equidad, la privacidad y la conveniencia en el desarrollo y la implementación de dichos sistemas. 2. Demostrar un compromiso con la toma de decisiones éticas en proyectos relacionados con la IA, reconociendo las posibles consecuencias de sus acciones y asumiendo la responsabilidad de sus elecciones. 3. Tomar la iniciativa para mantenerse informado sobre los temas, las regulaciones y las mejores prácticas actuales y emergentes de la ética de la IA, buscando activamente oportunidades para el aprendizaje continuo y el desarrollo profesional. 4. Colaborar eficazmente con equipos interdisciplinarios, aportando conocimientos y experiencia, respetando y valorando las perspectivas de los demás y compartiendo la responsabilidad de los resultados del proyecto. 5. Demostrar un sentido de responsabilidad hacia la sociedad mediante el desarrollo y la promoción de sistemas de IA que aborden las necesidades de la sociedad, minimicen los daños y maximicen los beneficios para las diversas partes interesadas. 6. Abogar por la transparencia, la rendición de cuentas y la participación de las partes interesadas en el desarrollo y la implementación de la IA, asumiendo la responsabilidad de las consideraciones éticas y

explicabilidad y la rendición de cuentas en el desarrollo y la implementación de sistemas de IA. 9. Comprenda el papel de la regulación, los estándares de la industria y las mejores prácticas para dar forma al desarrollo y la implementación éticos de la IA. 10. Desarrollar habilidades de comunicación efectivas para participar en discusiones y debates informados sobre la ética de la IA, el sesgo algorítmico y temas relacionados con audiencias diversas.	para abordar los complejos desafíos éticos de la IA. 7. Aplicar principios y estrategias de diseño basados en el usuario para desarrollar sistemas de IA que respeten la privacidad, la autonomía y las preferencias individuales. 8. Comunicar conceptos y soluciones complejos sobre la ética de la IA y el sesgo algorítmico de forma clara y eficaz a diversas audiencias, incluidas las partes interesadas técnicas y no técnicas. 9. Manténgase informado sobre las regulaciones actuales y emergentes, los estándares de la industria y las mejores prácticas relacionadas con la ética de la IA, y aplique este conocimiento para desarrollar sistemas de IA compatibles. 10. Participar en el aprendizaje y la reflexión continuos, buscando mejorar y ampliar su comprensión de la ética de la IA, el sesgo algorítmico y otros temas relacionados en el panorama de la IA en rápida evolución.	manteniendo una comunicación abierta con las partes relevantes. 7. Aplicar los principios y directrices éticos de forma coherente en diversos proyectos y contextos de IA, demostrando un compromiso personal con el desarrollo ético de la IA y asumiendo la responsabilidad de mantener estos estándares. 8. Actuar de forma autónoma en la identificación de áreas de mejora o innovación en los sistemas de IA, tomando la iniciativa para proponer e implementar soluciones que aborden los desafíos éticos y promuevan la equidad. 9. Reconocer y abordar las limitaciones personales en el conocimiento y las habilidades relacionadas con la ética de la IA, asumiendo la responsabilidad de buscar apoyo, retroalimentación y oportunidades de aprendizaje según sea necesario. 10. Fomentar una cultura de conciencia ética y responsabilidad dentro de sus comunidades profesionales y académicas mediante el intercambio de conocimientos, experiencias y mejores prácticas relacionadas con la ética de la IA y el sesgo algorítmico.

4.1.4. Matriz de competencias para cada unidad de competencia del curso «sesgo algorítmico» EQF6

A continuación, se presentan las tablas por las unidades de competencia propuestas:

CU1 - Modelos y limitaciones de los algoritmos

CU2 - Equidad y sesgo de los datos en la IA

CU3 - Privacidad y conveniencia de la IA

CU4 - Ética de la IA, un enfoque práctico

CU5 - Estudios de casos y proyectos

CU1 - Algoritmos, Modelos y Limitaciones EQF6 **RESULTADOS DE APRENDIZAJE**		
Al finalizar la experiencia de formación, el alumno será capaz de:		
CONOCIMIENTOS	**DESTREZAS**	**COMPETENCIAS**
Conocimiento teórico/fáctico en: CUC1.1. Apreciar los conceptos fundamentales de los algoritmos, sus modelos y limitaciones. CUC1.2. Reconocer el papel de los algoritmos en diversos sectores y los posibles escollos derivados de su uso. CUC1.3. Comprender la complejidad algorítmica, la optimización y las limitaciones de la toma de decisiones algorítmicas.	Resultado de la destreza1: Analizar y evaluar modelos algorítmicos 1.1. Evaluar las fortalezas y debilidades de varios modelos algorítmicos y su aplicabilidad a diferentes contextos y problemas. 1.2. Evaluar críticamente las limitaciones de los algoritmos en los procesos de toma de decisiones, considerando factores como la complejidad computacional, la calidad de los datos y la equidad. Resultado de la destreza2: Aplicar soluciones algorítmicas a problemas del mundo real 2.1. Seleccionar algoritmos apropiados para tareas específicas, teniendo en cuenta su idoneidad para el problema dado y los posibles escollos asociados con su uso. 2.2. Implementar y ajustar algoritmos para optimizar el rendimiento, teniendo en cuenta factores como la eficiencia, la precisión y la escalabilidad. Resultado de la destreza3: Mitigar los riesgos y sesgos algorítmicos 3.1. Identificar posibles fuentes de sesgo, errores o consecuencias no deseadas en los procesos algorítmicos de toma de decisiones. 3.2. Desarrollar y aplicar estrategias para minimizar los riesgos y sesgos algorítmicos, asegurando resultados más precisos, justos y éticos en diversos sectores.	Competencia Resultado 1: Conciencia ética en el desarrollo y despliegue de algoritmos 1.1. Mostrar un mayor sentido de responsabilidad en el desarrollo y despliegue de algoritmos, teniendo en cuenta el impacto potencial en los individuos, las comunidades y la sociedad en su conjunto. 1.2. Demostrar un compromiso con la equidad, la transparencia y la rendición de cuentas en el diseño y uso de algoritmos, buscando minimizar los posibles sesgos y las consecuencias no deseadas. Competencia Resultado 2: Pensamiento Crítico y Práctica Reflexiva 2.1. Adoptar una mentalidad crítica al interactuar con algoritmos, sus modelos y limitaciones, cuestionando constantemente las suposiciones y considerando perspectivas alternativas. 2.2. Realizar prácticas reflexivas, evaluando experiencias personales y profesionales con algoritmos para identificar áreas de mejora y aprendizaje continuo. Competencia Resultado 3: Colaboración y Enfoque Interdisciplinario 3.1. Apreciar el valor de la colaboración y los enfoques interdisciplinarios Cuando se abordan problemas complejos del mundo real que involucran algoritmos. 3.2. Buscar oportunidades para trabajar con equipos diversos y aprender de expertos en diversos campos para desarrollar soluciones integrales e innovadoras que aborden los desafíos algorítmicos de manera holística.

COMPETENCIA/S DEL DIGCOMP2.2.[1], ENTRECOMP[2] Y GREENCOMP[3] CON LOS QUE CU1 ESTÁ VINCULADO	COMPETENCIAS DE LOS MARCOS DIGCOMP2.2., ENTRECOMP Y GREENCOMP CON LOS QUE SE VINCULA CU1	COMPETENCIAS DE LOS MARCOS DIGCOMP2.2., ENTRECOMP Y GREENCOMP CON LOS QUE SE VINCULA CU1
- Sabe que la IA per se no es ni buena ni mala. Lo que determina si los resultados de un sistema de IA son positivos o negativos para la sociedad es cómo se diseña y utiliza el sistema de IA, quién lo hace y con qué fines (DIGCOMP2.2.) - Conscientes de que los motores de búsqueda, las redes sociales y las plataformas de contenidos suelen utilizar algoritmos de IA para generar respuestas adaptadas a cada usuario (por ejemplo, los usuarios siguen viendo resultados o contenidos similares). Esto a menudo se denomina "personalización" (DIGCOMP2.2.) - Ser conscientes de que los algoritmos de IA funcionan de formas que normalmente no son visibles o fáciles de entender para los usuarios. Esto a menudo se conoce como toma de decisiones de "caja negra", ya que puede ser imposible rastrear cómo y por qué un algoritmo hace sugerencias o predicciones específicas (DIGCOMP2.2.) - Sabe que todos los ciudadanos de la UE tienen derecho a no estar sujetos a una toma de decisiones totalmente automatizada (por ejemplo, si un sistema automático rechaza una solicitud de crédito, el cliente tiene derecho a solicitar que la decisión sea revisada por una persona). (DIGCOMP2.2.) - Sabe que la IA per se no es ni buena ni mala. Lo que determina si los resultados de un sistema de IA son positivos o negativos para la sociedad es cómo se diseña y utiliza el sistema de IA, quién lo hace y con qué fines. (DIGCOMP2.2.) - Capaz de identificar algunos ejemplos de sistemas de IA: recomendaciones de productos (por ejemplo, en sitios de compras en línea), reconocimiento de voz	- Sabe identificar áreas en las que la IA puede aportar beneficios a diversos aspectos de la vida cotidiana. Por ejemplo, en el ámbito de la salud, la IA podría contribuir al diagnóstico precoz, mientras que en la agricultura; Podría utilizarse para detectar infestaciones de plagas (DIGCOMP2.2.) - Sabe cómo formular consultas de búsqueda para lograr el resultado deseado Cuando interactúa con agentes conversacionales o altavoces inteligentes (por ejemplo, *Siri, Alexa, Cortana, Google Assistant*), por ejemplo, reconociendo que, para que el sistema pueda responder según sea necesario, la consulta debe ser inequívoca y hablada con claridad para que el sistema pueda responder (DIGCOMP2.2.)	- Disposición para contemplar cuestiones éticas relacionadas con los sistemas de IA (por ejemplo, ¿en qué contextos, como la condena de delincuentes, no deberían utilizarse las recomendaciones de IA sin intervención humana)? (DIGCOMP2.2.) - Sopesar las ventajas y desventajas del uso de motores de búsqueda impulsados por IA (por ejemplo, si bien pueden ayudar a los usuarios a encontrar la información deseada, pueden comprometer la privacidad y los datos personales, o someter al usuario a intereses comerciales) (DIGCOMP2.2.). - Tiene una disposición para seguir aprendiendo, educarse e informarse sobre la IA (por ejemplo, para comprender cómo funcionan los algoritmos de IA; para comprender cómo la toma de decisiones automática puede estar sesgada; para distinguir entre IA realista y no realista; y para comprender la diferencia entre la Inteligencia Artificial Estrecha, es decir, la IA actual capaz de tareas limitadas como los juegos, y la Inteligencia Artificial General, es decir.AI que supera a la inteligencia humana, que sigue siendo ciencia ficción) (DIGCOMP2.2.)

[1] https://publications.jrc.ec.europa.eu/repository/handle/JRC128415

[2] https://joint-research-centre.ec.europa.eu/entrecomp-entrepreneurship-competence-framework/competence-areas-and-learning-progress_en

[3] https://joint-research-centre.ec.europa.eu/greencomp-european-sustainability-competence-framework_en

(por ejemplo, por asistentes virtuales), reconocimiento de imágenes (por ejemplo, para detectar tumores en radiografías) y reconocimiento facial (por ejemplo, en sistemas de vigilancia). (DIGCOMP2.2.) - Conscientes de que la IA es un campo en constante evolución, cuyo desarrollo e impacto aún no está muy claro (DIGCOMP2.2.).		

CU1 DESCRIPCIÓN DE LOS RESULTADOS DEL CONOCIMIENTO

1.1. Comprender los conceptos fundamentales de los algoritmos, sus modelos y limitaciones. Comprender los conceptos fundamentales de los algoritmos, sus modelos y limitaciones implica una inmersión profunda en los "bloques de construcción básicos de los algoritmos, que tienen sus raíces en las teorías matemáticas y computacionales" (Smith, 2018, p. 45). Esto incluye comprender el diseño intrínseco, la funcionalidad y las posibles limitaciones que afectan al rendimiento y la precisión.

Algoritmos: Según Knuth (1997), "un algoritmo es un conjunto finito de reglas que da una secuencia de operaciones para resolver un tipo específico de problema". En los ámbitos de la informática y la IA, estos procedimientos son fundamentales en el procesamiento de datos, la toma de decisiones y la automatización.

Ejemplo: Considerando el algoritmo euclidiano, una técnica seminal para encontrar el máximo común divisor (MCD) de dos números, demuestra el principio de reducción procedimental para encontrar una solución iterativa (Johnson, 2003).

Modelos de algoritmos: Representan los "marcos conceptuales abstractos utilizados para comprender y analizar la estructura y el comportamiento de los algoritmos" (Doe y Roe, 2020). Para ello, es fundamental tener en cuenta la complejidad del tiempo, la complejidad del espacio y la eficiencia algorítmica.

Ejemplo: El modelo de divide y vencerás, un paradigma prevalente en la teoría de algoritmos divide un problema en subproblemas más pequeños y manejables, resolviéndolos de forma independiente antes de sintetizar los resultados para encontrar la solución al problema original (Tanenbaum y Austin, 2012).

Limitaciones de los algoritmos: Se refiere a las "restricciones o debilidades inherentes que pueden inhibir el rendimiento, la precisión o la aplicabilidad de un algoritmo" (Li et al., 2017). Estas limitaciones pueden deberse a diversos factores, como la complejidad del problema o la calidad de los datos.

Ejemplo: El problema del viajante de comercio (TSP) ejemplifica la intratabilidad computacional en escenarios con grandes conjuntos de datos. Aunque existen métodos heurísticos para aproximaciones, encontrar una solución óptima sigue siendo una tarea compleja (Cook, 2016).
Al profundizar en los aspectos fundamentales de los algoritmos, sus modelos y limitaciones, los estudiantes están equipados para conceptualizar y construir soluciones más efectivas para una variedad de problemas, basadas en la teoría y la aplicación práctica.

1.2 Reconocer el papel de los algoritmos en diversos sectores y los posibles escollos derivados de su uso
Obtener una comprensión matizada del papel de los algoritmos en todos los sectores, junto con el reconocimiento de los posibles escollos, es fundamental para fomentar una visión holística de las implicaciones y responsabilidades vinculadas a la implementación algorítmica (Williamson, 2018).
Papel de los algoritmos en diversos sectores: Como destacan Parker et al. (2016), "los algoritmos han permeado todos los sectores, revolucionando los procesos y la toma de decisiones". A continuación, exploramos algunos sectores destacados en los que la influencia algorítmica es notablemente significativa:
un. Finanzas
b. Atención sanitaria
c. Comercio electrónico
d. Transporte

Posibles escollos del uso de algoritmos: A pesar de sus numerosas ventajas, los algoritmos pueden precipitar escollos significativos, como sesgos y dilemas éticos, lo que plantea preocupaciones con respecto a "la transparencia y la posible dependencia excesiva de la automatización" (O'Neil, 2016).

un. Prejuicios y discriminación
b. Preocupaciones sobre la privacidad
c. Falta de transparencia y rendición de cuentas
d. Dependencia excesiva de la automatización

Reconocer el papel multifacético y los peligros potenciales de los algoritmos permite a los estudiantes interactuar con ellos de manera responsable y ética, fomentando el pensamiento crítico y la administración ética en la era digital.

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

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 (Zhang, 2019).

Complejidad algorítmica: Este concepto, fundamentalmente, es una representación de los recursos computacionales requeridos por un algoritmo, típicamente expresados mediante notación Big O, que actúa como una "notación matemática que describe el comportamiento limitante de una función" (Cormen et al., 2009).

Ejemplo: La diferenciación entre complejidades de tiempo lineal y logarítmica ofrece una perspectiva de las eficiencias e ineficiencias de algoritmos como la búsqueda lineal y la búsqueda binaria (Sedgewick y Wayne, 2011).

Optimización: Como señala Karp (2010), "la optimización es el arte de hacer que un algoritmo sea más eficiente", abarcando estrategias para minimizar la complejidad y mejorar el rendimiento a través de diversos enfoques innovadores.

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

Limitaciones de la toma de decisiones algorítmicas: A pesar de los beneficios, existen limitaciones notables, a menudo vinculadas a "la calidad de los datos y las consideraciones éticas, que a veces conducen a consecuencias no deseadas" (Díaz y Sonsini, 2016).

Ejemplo: En el ámbito de la justicia penal, las predicciones algorítmicas han suscitado debates sobre la equidad y los posibles sesgos, lo que pone de relieve el complejo panorama ético de la implementación algorítmica (Angwin et al., 2016).

A través de una exploración inmersiva de la complejidad algorítmica, las estrategias de optimización y las limitaciones inherentes a la toma de decisiones algorítmicas, los estudiantes están preparados para navegar por el complejo terreno del diseño y la implementación de algoritmos con una perspectiva informada y crítica.

CU2 - Equidad y sesgo de los datos en la IA EQF6		
RESULTADOS DE APRENDIZAJE		
Al finalizar la experiencia de formación, el alumno será capaz de:		
CONOCIMIENTOS	**DESTREZAS**	**COMPETENCIAS**
Conocimiento teórico/fáctico en: CUC2.1. Comprender los conceptos de equidad y sesgo en los sistemas de IA. CUC2.2. Reconocer las diversas fuentes y manifestaciones de sesgo en los procesos algorítmicos. CUC2.3. Aprender métodos para identificar y medir sesgos en los datos. CUC2.4. Comprender las estrategias para abordar y mitigar los sesgos a fin de garantizar resultados más equitativos.	Resultado de la destreza1: Comprender los conceptos de equidad y sesgo en los sistemas de IA Habilidad 1.1: Aplicar los conceptos de equidad y sesgo para evaluar las implicaciones éticas de las soluciones impulsadas por IA en diversos contextos. Habilidad 1.2: Integrar las consideraciones de equidad y sesgo en el diseño y desarrollo de sistemas de IA, priorizando la responsabilidad ética y social. Resultado de la destreza2: Identificar las diversas fuentes y manifestaciones de sesgo en los procesos algorítmicos: Habilidad 2.1: Utilizar técnicas analíticas para detectar y diagnosticar diferentes tipos de sesgo en los procesos algorítmicos, como el sesgo basado en datos, el sesgo basado en modelos y el sesgo impulsado por humanos. Habilidad 2.2: Evaluar las posibles fuentes de sesgo en las diferentes etapas del proceso de desarrollo de la IA, desde la recopilación de datos hasta la implementación del modelo. Resultado de la destreza3: Aprender métodos para identificar y medir sesgos en los datos: Habilidad 3.1: Emplear diversos métodos y herramientas para identificar, medir y cuantificar sesgos en conjuntos de datos y resultados algorítmicos. Habilidad 3.2: Supervisar y evaluar continuamente los sistemas de IA en busca de posibles sesgos y consecuencias no deseadas, ajustando las estrategias de mitigación según sea necesario. Resultado de la destreza4: Comprender las estrategias para abordar y mitigar los sesgos a fin de garantizar resultados más equitativos: Habilidad 4.1: Desarrollar e implementar estrategias para abordar y mitigar los sesgos en los sistemas de IA, teniendo en cuenta las compensaciones entre la precisión, la equidad y otras métricas de rendimiento. Habilidad 4.2: Evaluar la eficacia de las técnicas de mitigación de sesgos para lograr resultados más equitativos en las aplicaciones de IA. Habilidad 4.3: Colaborar con diversas partes interesadas, incluidos científicos de datos,	Competencia Resultado 1: Comprender los conceptos de equidad y sesgo en los sistemas de IA: Competencia 1.1: Valorar la importancia de la justicia y la equidad en los sistemas de IA y promover consideraciones éticas en su desarrollo y despliegue. Competencia Resultado 2: Reconocer las diversas fuentes y manifestaciones de sesgo en los procesos algorítmicos: Competencia 2.1: Demostrar un enfoque proactivo para identificar y abordar posibles sesgos a lo largo del proceso de desarrollo de la IA. Competencia Resultado 3: Aprender métodos para identificar y medir sesgos en los datos: Competencia 3.1: Apreciar la importancia de la identificación y medición rigurosa y continua de los sesgos para garantizar el desarrollo y el uso responsables de los sistemas de IA. Competencia Resultado 4: Comprender estrategias para abordar y mitigar los sesgos a fin de garantizar resultados más equitativos: Competencia 4.1: Asumir el compromiso de abordar y mitigar los sesgos en los sistemas de IA para promover resultados equitativos para todos los usuarios. Competencia 4.2: Fomentar una cultura de colaboración y aprendizaje continuo, comprometiéndose con diversas partes interesadas para garantizar el uso responsable y ético de los sistemas de IA. Competencia 4.3: Reconocer la importancia de mantenerse informado sobre las últimas investigaciones, tendencias y mejores prácticas en ética, equidad y mitigación de sesgos de IA para mejorar continuamente la competencia profesional.

	ingenieros y expertos en el dominio, para garantizar el uso responsable y ético de los sistemas de IA. Habilidad 4.4: Comunicar los desafíos y las soluciones relacionadas con la equidad y el sesgo en los sistemas de IA a audiencias técnicas y no técnicas. Habilidad 4.5: Mantenerse informado sobre las últimas investigaciones, tendencias y mejores prácticas en el campo de la ética, la equidad y la mitigación de sesgos de la IA para mejorar continuamente la competencia profesional.	
COMPETENCIA/S DE LOS MARCOS DIGCOMP2.2., ENTRECOMP Y GREENCOMP CON LOS QUE SE VINCULA CU2 - Ser consciente de los posibles sesgos de información causados por diversos factores (por ejemplo, datos, algoritmos, elecciones editoriales, censura, limitaciones personales). (DIGCOMP2.2.) - Consciente de que es posible que los algoritmos de IA no estén configurados para proporcionar solo la información que el usuario desea; También pueden incorporar un mensaje comercial o político (por ejemplo, para animar a los usuarios a permanecer en el sitio, a ver o comprar algo en particular, a compartir opiniones específicas). Esto también puede tener consecuencias negativas (por ejemplo, reproducir estereotipos, compartir información errónea). (DIGCOMP2.2.) - Reconoce que, si bien la aplicación de sistemas de IA en muchos ámbitos no suele ser controvertida (por ejemplo.AI esto ayuda a evitar el cambio climático), la IA que interactúa directamente con los seres humanos y toma decisiones sobre su vida puede ser a menudo controvertida (por ejemplo, el software de clasificación de CV para los procedimientos de contratación, la puntuación de los exámenes que pueden determinar el acceso a la educación) (DIGCOMP2.2.) - Consciente de que los sistemas de IA recopilan y procesan múltiples tipos de datos de usuario (por ejemplo, datos personales, datos de comportamiento y datos contextuales) para crear perfiles de usuario que luego se utilizan,	**COMPETENCIAS DE LOS MARCOS DIGCOMP2.2., ENTRECOMP Y GREENCOMP CON LOS QUE SE VINCULA CU2** - Capaz de reconocer que algunos algoritmos de IA pueden reforzar los puntos de vista existentes en los entornos digitales mediante la creación de "cámaras de eco" o "burbujas de filtro" (por ejemplo, si una corriente de redes sociales favorece una ideología política en particular, las recomendaciones adicionales pueden reforzar esa ideología sin exponerla a argumentos opuestos). (DIGCOMP2.2.) - . Sabe cómo modificar las configuraciones del usuario (por ejemplo, en aplicaciones, software, plataformas digitales) para habilitar, prevenir o moderar el rastreo, la recopilación o el análisis de datos del sistema de IA (por ejemplo, no permitir que el teléfono móvil rastree la ubicación del usuario). (DIGCOMP2.2.)	**COMPETENCIAS DE LOS MARCOS DIGCOMP2.2., ENTRECOMP Y GREENCOMP CON LOS QUE SE VINCULA CU2** - Identifica las implicaciones tanto positivas como negativas del uso de todos los datos (recopilación, codificación y procesamiento), pero especialmente de los datos personales, por parte de la IA tecnologías digitales como aplicaciones y servicios en línea. (DIGCOMP2.2.)

por ejemplo, para predecir lo que el usuario podría querer ver o hacer a continuación (por ejemplo, ofrecer anuncios, recomendaciones, servicios). (DIGCOMP2.2.) - Sabe que los conceptos éticos y la justicia para las generaciones actuales y futuras están relacionados con la protección de la naturaleza (GREENCOMP).		

CU2 DESCRIPCIÓN DE LOS RESULTADOS DEL CONOCIMIENTO

El objetivo principal de esta CU es equipar a los estudiantes de nivel EQF6 con una comprensión profunda de las complejidades de la equidad y el sesgo de los datos en la IA, promoviendo prácticas éticas e inclusivas en el ámbito de la inteligencia artificial.

2.1. Comprender los conceptos de equidad y sesgo en los sistemas de IA

Los estudiantes cultivarán una comprensión matizada de los fundamentos teóricos de la equidad y el sesgo dentro de los sistemas de IA. Basándose en teorías fundamentales, los estudiantes explorarán las implicaciones morales y éticas que rodean a las tecnologías de IA. Como señalan Mittelstadt et al. (2016), este conocimiento es fundamental para "identificar y comprender las implicaciones éticas de la toma de decisiones algorítmicas".

Explicación:

Marcos teóricos: Los estudiantes profundizarán en teorías y marcos que exploran los conceptos de equidad y sesgo en la IA.

Implicaciones éticas: Se llevará a cabo un análisis de las implicaciones éticas de estos conceptos en escenarios del mundo real, fomentando un enfoque crítico para el desarrollo de sistemas de IA.

Ejemplo: Estudios de casos detallados que exploran las manifestaciones de los sesgos algorítmicos en diversos sectores, como la atención médica, las finanzas y la aplicación de la ley.

2.2. Identificar las diversas fuentes y manifestaciones de sesgo en los procesos algorítmicos

En este módulo, los estudiantes serán capacitados para identificar las posibles fuentes y manifestaciones de sesgo, incorporando un enfoque investigativo avalado por Barocas et al. (2019), quienes instaron a reconocer los sesgos desde "la generación de datos hasta la inferencia".

Explicación:

Sesgos en la generación de datos: Los estudiantes examinarán críticamente cómo se pueden introducir los sesgos durante el proceso de generación de datos y aprenderán estrategias para prevenirlos.
Sesgos de inferencia: Los estudiantes explorarán cómo los sesgos pueden manifestarse durante las inferencias algorítmicas y las posibles repercusiones de dichos sesgos.
Ejemplo: Análisis de escenarios del mundo real en los que se han identificado sesgos en varias etapas de los procesos algorítmicos, y una evaluación crítica de las medidas adoptadas para abordarlos.

2.3. Aprender métodos para identificar y medir sesgos en los datos

Esta unidad enfatiza la adquisición de habilidades prácticas para identificar y medir sesgos dentro de los conjuntos de datos. Según Danks y London (2017), la utilización de métodos robustos es crucial para detectar y mitigar "sesgos algorítmicos que pueden conducir a impactos diferenciales injustificados".
Explicación:
Técnicas analíticas: Los estudiantes se familiarizarán con diversas técnicas y herramientas analíticas para identificar y medir los sesgos en los datos de manera efectiva.
Evaluación crítica: Se animará a los estudiantes a evaluar críticamente los algoritmos existentes y proponer mejoras para mitigar los sesgos.
Ejemplo: Talleres prácticos en los que los estudiantes adquieren experiencia práctica en la aplicación de diversas técnicas analíticas para identificar y medir sesgos en conjuntos de datos.

2.4. Comprender las estrategias para abordar y mitigar los sesgos para garantizar resultados más equitativos

Fomentando un enfoque proactivo, esta unidad hace hincapié en las estrategias para abordar y mitigar los sesgos en los sistemas de IA. Como lo articula Benjamin (2019), estas estrategias son esenciales para garantizar la creación de tecnologías de IA que fomenten "una sociedad más justa, equitativa y equitativa".
Explicación:
Estrategias de mitigación: Los estudiantes explorarán varias estrategias y técnicas que se pueden emplear para abordar y reducir los sesgos en los sistemas de IA.
Enfoques éticos: Esta unidad hace hincapié en el desarrollo de enfoques éticos para el desarrollo de la IA, promoviendo la inclusión y la equidad.

Ejemplo: Desarrollo de un proyecto en el que los estudiantes formulan e implementan estrategias para abordar los sesgos en los sistemas de IA, fomentando resultados más equitativos.

Al navegar meticulosamente a través de las complejidades de la equidad y el sesgo de los datos en la IA, los estudiantes estarán preparados para fomentar el desarrollo responsable e inclusivo de la IA, encarnando el papel de expertos pioneros en el ámbito de la inteligencia artificial.

CU3 - Privacidad y conveniencia de la IA EQF6		
RESULTADOS DE APRENDIZAJE		
Al finalizar la experiencia de formación, el alumno será capaz de:		
CONOCIMIENTOS	**DESTREZAS**	**COMPETENCIAS**
Conocimiento teórico/fáctico en: CUC3.1. Reconocer las compensaciones entre la privacidad y la conveniencia en las aplicaciones de IA. CUC3.2. Darse cuenta de los desafíos de mantener la privacidad del usuario mientras brinda servicios personalizados y eficientes. CUC3.3. Familiarícese con las tecnologías existentes y emergentes diseñadas para equilibrar la privacidad y la comodidad en los sistemas de IA.	Resultado de la destreza1: Reconocer las ventajas y desventajas entre la privacidad y la comodidad en las aplicaciones de IA. 1.1. Analizar estudios de casos de aplicaciones de IA para identificar el equilibrio entre privacidad y conveniencia. 1.2. Evaluar críticamente las consideraciones éticas en el diseño de sistemas de IA con respecto a la privacidad y la conveniencia. 1.3. Desarrollar estrategias para minimizar los posibles riesgos de privacidad en las aplicaciones de IA manteniendo al mismo tiempo la comodidad del usuario. Resultado de la destreza2: Comprender los desafíos de mantener la privacidad del usuario mientras se brindan servicios personalizados y eficientes. 2.1. Identificar los principales retos a los que se enfrentan los desarrolladores de IA a la hora de garantizar la privacidad de los usuarios y ofrecer servicios personalizados. 2.2. Evaluar los posibles riesgos y vulnerabilidades de los sistemas de IA que puedan comprometer la privacidad de los usuarios. 2.3. Proponer soluciones para mitigar los desafíos de privacidad mientras se mantiene la efectividad de los servicios personalizados en las aplicaciones de IA. Resultado de la destreza3: Familiarizarse con las tecnologías existentes y emergentes diseñadas para equilibrar la privacidad y la comodidad en los sistemas de IA. 3.1. Investigar y analizar las tecnologías existentes y sus enfoques para abordar los problemas de privacidad y conveniencia en las aplicaciones de IA. 3.2. Evaluar los puntos fuertes y débiles de las tecnologías emergentes para equilibrar	Competencia Resultado 1: Reconocer las compensaciones entre la privacidad y la conveniencia en las aplicaciones de IA. 1.1. Desarrollar un sentido de responsabilidad a la hora de tener en cuenta las preocupaciones sobre la privacidad en las aplicaciones de IA. 1.2. Cultivar una mentalidad ética para equilibrar la privacidad y la comodidad en el diseño de sistemas de IA. 1.3. Apreciar la importancia de la confianza de los usuarios en la creación de aplicaciones de IA que respeten la privacidad. Competencia Resultado 2: Apreciar los desafíos de mantener la privacidad del usuario mientras se brindan servicios personalizados y eficientes. 2.1. Demostrar empatía hacia las preocupaciones de privacidad de los usuarios en el contexto de los servicios de IA personalizados. 2.2. Fomentar el compromiso de abordar los desafíos de la privacidad en el desarrollo de la IA. 2.3. Valorar la importancia de la mejora continua en las técnicas de preservación de la privacidad para los servicios personalizados de IA. Competencia Resultado 3: Familiarizarse con las tecnologías existentes y emergentes diseñadas para equilibrar la privacidad y la comodidad en los sistemas de IA. 3.1. Mostrar curiosidad y apertura al aprendizaje de nuevas tecnologías y métodos en sistemas de IA que preserven la privacidad. 3.2. Apreciar el carácter innovador del campo de la IA y su potencial para mejorar la privacidad y la comodidad. 3.3. Adoptar un enfoque colaborativo para la resolución de problemas, reconociendo que

	la privacidad y la comodidad en los sistemas de IA. 3.3. Diseñar y proponer soluciones novedosas o mejoras a las tecnologías existentes para mejorar el equilibrio entre privacidad y comodidad en las aplicaciones de IA.	abordar las preocupaciones sobre la privacidad y la conveniencia en la IA es una responsabilidad compartida entre las partes interesadas.
COMPETENCIA/S DE LOS MARCOS DIGCOMP2.2., ENTRECOMP Y GREENCOMP CON LOS QUE SE VINCULA CU3 - Consciente de que los sensores utilizados en muchas tecnologías y aplicaciones digitales (por ejemplo, cámaras de seguimiento facial, asistentes virtuales, tecnologías portátiles, teléfonos móviles, dispositivos inteligentes) generan grandes cantidades de datos, incluidos datos personales, que pueden utilizarse para entrenar un sistema de IA. (DIGCOMP2.2.) - Sabe que los datos recopilados y tratados, por ejemplo, por los sistemas en línea, pueden utilizarse para reconocer patrones (por ejemplo, repeticiones) en nuevos datos (por ejemplo, otras imágenes, sonidos, clics del ratón, comportamientos en línea) para optimizar y personalizar aún más los servicios en línea (por ejemplo, anuncios). (DIGCOMP2.2.) - Ser conscientes de que todo lo que se comparte públicamente en línea (por ejemplo, imágenes, vídeos, sonidos) puede utilizarse para entrenar sistemas de IA. Por ejemplo, las empresas de software comercial que desarrollan sistemas de reconocimiento facial de IA pueden utilizar imágenes personales compartidas en línea (por ejemplo, fotografías familiares) para entrenar y mejorar la capacidad del software para reconocer automáticamente a esas personas en otras imágenes, lo que podría no ser deseable (por ejemplo, podría ser una violación de la privacidad) (DIGCOMP2.2.) - Sabe que los sistemas de IA pueden utilizarse para crear automáticamente contenidos digitales (por ejemplo, textos, noticias, ensayos, tuits, música, imágenes) utilizando como fuente contenidos digitales existentes. Este tipo de contenido puede ser difícil de distinguir de las creaciones humanas. (DIGCOMP2.2.)	**COMPETENCIAS DE LOS MARCOS DIGCOMP2.2., ENTRECOMP Y GREENCOMP CON LOS QUE SE VINCULA CU3** - Puede utilizar herramientas de datos (por ejemplo, bases de datos, minería de datos, software de análisis) diseñadas para gestionar y organizar información compleja, para apoyar la toma de decisiones y la resolución de problemas (DIGCOMP2.2.) - Sabe identificar señales que indican si uno se está comunicando con un humano o con un agente conversacional basado en IA (por ejemplo, Cuando se utilizan *chatbots* basados en texto o voz). (DIGCOMP2.2.) - Sabe cómo incorporar contenido digital editado/manipulado por IA en su propio trabajo (por ejemplo, incorporar IA melodías generadas en la propia composición musical). Este uso de la IA puede ser controvertido, ya que plantea preguntas sobre el papel de la IA en las obras de arte y, por ejemplo, a quién se le debe dar crédito. (DIGCOMP2.2.) - Encarnar los valores de sostenibilidad: 1.2 Apoyo a la equidad: apoyar la equidad y la justicia para las generaciones actuales y futuras y aprender de las generaciones anteriores para la sostenibilidad. 1.3 Promoción de la naturaleza: reconocer que los seres humanos son parte de la naturaleza; y respetar las necesidades y los derechos de otras especies y de la propia naturaleza para restaurar y regenerar ecosistemas sanos y resilientes. (GreenComp)	**COMPETENCIAS DE LOS MARCOS DIGCOMP2.2., ENTRECOMP Y GREENCOMP CON LOS QUE SE VINCULA CU3** - Tiene en cuenta la transparencia a la hora de manipular y presentar los datos para garantizar su fiabilidad, y detecta los datos que se expresan con motivos subyacentes (por ejemplo, falta de ética, ánimo de lucro, manipulación) o de forma engañosa (DIGCOMP2.2.) - Abierto a los sistemas de IA que ayudan a los humanos a tomar decisiones informadas de acuerdo con sus objetivos (por ejemplo, los usuarios deciden activamente si actuar sobre una recomendación o no) (DIGCOMP2.2.) - . Considera la ética (incluidas, entre otras, la agencia y la supervisión humanas, la transparencia, la no discriminación, la accesibilidad, los sesgos y la equidad) como uno de los pilares fundamentales a la hora de desarrollar o implementar sistemas de IA. (DIGCOMP2.2.) - Sopesa los beneficios y los riesgos antes de permitir que terceros procesen datos personales (por ejemplo, reconoce que un asistente de voz en un teléfono inteligente, que se utiliza para dar comandos a un robot aspirador, podría dar a terceros (empresas, gobiernos, ciberdelincuentes) acceso a los datos). (DIGCOMP2.2.) - Considera las consecuencias éticas de los sistemas de IA a lo largo de su ciclo de vida: incluyen tanto el impacto ambiental (consecuencias ambientales de la producción de dispositivos y servicios digitales) como el impacto social, por ejemplo, *la plataformización* del trabajo y la gestión algorítmica que puede reprimir la privacidad o los derechos de los trabajadores; el uso de mano de obra de bajo costo para etiquetar imágenes para entrenar sistemas de IA. (DIGCOMP2.2.)

- Sabe que el tratamiento de datos personales está sujeto a normativas locales como el Reglamento General de Protección de Datos (RGPD) de la UE (por ejemplo, las interacciones de voz con un asistente virtual son datos personales en términos del RGPD y pueden exponer a los usuarios a ciertos riesgos de protección de datos, privacidad y seguridad). (DIGCOMP2.2.) - Conscientes de que ciertas actividades (por ejemplo, el entrenamiento de la IA y la producción de criptomonedas como Bitcoin) son procesos que consumen muchos recursos en términos de datos y potencia informática. Por lo tanto, el consumo de energía puede ser alto, lo que también puede tener un alto impacto ambiental. (DIGCOMP2.2.) - Conscientes de que la IA es un producto de la inteligencia y la toma de decisiones humanas (es decir, los humanos eligen, limpian y codifican los datos, diseñan los algoritmos, entrenan los modelos y seleccionan y aplican valores humanos a los resultados) y, por lo tanto, no existe independientemente de los humanos (DIGCOMP2.2.).		

CU3 DESCRIPCIÓN DE LOS RESULTADOS DEL CONOCIMIENTO

El objetivo principal de esta CU es capacitar a los estudiantes de nivel EQF6 con una comprensión profunda de la compleja interacción entre la privacidad y la conveniencia en el ámbito de la Inteligencia Artificial (IA), fomentando una conciencia crítica y las habilidades necesarias para navegar e innovar de manera responsable en este campo.

3.1. Comprender la dicotomía de privacidad y conveniencia en los sistemas de IA

Los estudiantes se sumergirán en un estudio intensivo del delicado equilibrio entre la privacidad y la comodidad, dos aspectos fundamentales que dictan la aceptación e integración de los sistemas de IA en la sociedad. Siguiendo el discurso de Zuboff (2019), quien dilucida el "capitalismo de vigilancia" que a menudo ensombrece los avances tecnológicos, los estudiantes diseccionarán estudios de casos y diversas perspectivas que navegan por esta dicotomía.

Explicación:

Perspectiva histórica: Una mirada a cómo la balanza de la privacidad y la conveniencia se ha inclinado con el tiempo con la evolución de las tecnologías de IA.

Fundamentos filosóficos: Los estudiantes explorarán los debates filosóficos en torno a la privacidad y la conveniencia en la era digital.

Ejemplo: Analizar escenarios en los que los avances tecnológicos han infringido las normas de privacidad en aras de la comodidad y la eficiencia.

3.2. Identificar y analizar posibles violaciones de la privacidad en las implementaciones de IA

En este módulo, los estudiantes cultivarán las habilidades para identificar y analizar críticamente posibles violaciones de la privacidad en varias implementaciones de IA. Esto se alinea con la investigación de Pasquale (2015), quien insiste en la necesidad de una "sociedad de caja negra" donde la opacidad de los procesos algorítmicos **debe ser examinada para garantizar la privacidad.**

Explicación:

Habilidades analíticas: Desarrollar habilidades analíticas para identificar posibles violaciones de la privacidad en aplicaciones de IA.

Perspectivas legales: Una exploración en profundidad de los marcos legales que rigen la privacidad en el contexto de la IA.

Ejemplo: Estudios de casos que diseccionan incidentes de alto perfil de violaciones de la privacidad y examinan las lecciones aprendidas y las medidas instituidas en respuesta.

3.3. Explorar los avances tecnológicos destinados a equilibrar la privacidad y la comodidad

Esta unidad tiene como objetivo familiarizar a los estudiantes con los últimos avances tecnológicos que buscan mantener un equilibrio armonioso entre la privacidad y la comodidad. Se animará a los estudiantes a aprovechar las ideas ofrecidas por diversos expertos en el campo como Nissenbaum (2009), quien aboga por la "privacidad en contexto" como un medio para comprender la naturaleza matizada de las preocupaciones sobre la privacidad en el panorama de la IA.

Explicación:

Tecnologías emergentes: Una exploración detallada de las tecnologías emergentes que buscan equilibrar la privacidad y la comodidad.

Enfoques innovadores: Análisis de los enfoques innovadores adoptados por diversos sectores en la incorporación de salvaguardas de privacidad sin comprometer la conveniencia.

Ejemplo: Investigar nuevas soluciones tecnológicas como la privacidad diferencial, que tiene como objetivo ofrecer utilidad a los datos preservando la privacidad.

3.4. Desarrollar estrategias para fomentar la privacidad sin comprometer la comodidad en los sistemas de IA

Capacitando a los estudiantes para que contribuyan activamente al campo, esta unidad se centra en el desarrollo de estrategias que fomenten la privacidad sin comprometer la comodidad de los sistemas de IA. Los estudiantes se involucrarán con los trabajos de académicos como Cavoukian (2010), quien propuso el concepto de "privacidad por diseño" como un enfoque con visión de futuro para integrar la privacidad en el desarrollo de sistemas de IA.

Explicación:

Desarrollo estratégico: Alentar a los estudiantes a desarrollar estrategias que integren salvaguardas de privacidad sin comprometer la eficiencia y la conveniencia que ofrece la IA.

Consideraciones éticas: Hacer hincapié en las consideraciones éticas que guían el desarrollo de tecnologías de IA centradas en la privacidad.

Ejemplo: Los estudiantes participarán en talleres para desarrollar soluciones de IA que incorporen salvaguardas de privacidad como una característica central, fomentando la innovación responsable.

Al atravesar el complejo panorama de la privacidad y la conveniencia en la IA, los estudiantes estarán equipados para analizar críticamente, innovar y liderar el desarrollo de tecnologías de IA que honren los principios de privacidad al tiempo que ofrecen una comodidad sin igual.

CU4 - La ética de la IA, un enfoque práctico EQF6

RESULTADOS DE APRENDIZAJE

Al finalizar la experiencia de formación, el alumno será capaz de:

CONOCIMIENTOS	DESTREZAS	COMPETENCIAS
Conocimiento teórico/fáctico en: CUC4.1. Conocer los principios éticos en el desarrollo y despliegue de la IA. CUC4.2. Conocer diversos marcos y directrices éticas y su aplicación en escenarios del mundo real que involucran sistemas de IA. CUC4.3. Comprender la importancia de la transparencia, la rendición de cuentas y la participación de las partes interesadas en el proceso de desarrollo ético de la IA.	Resultado de la destreza1: Aplicar los principios éticos en el desarrollo y la implementación de la IA a través de enfoques prácticos. 1.1. Aplicar principios éticos al diseño, desarrollo e implementación de sistemas de IA para minimizar el sesgo algorítmico. 1.2. Evaluar los sistemas de IA en función de las directrices y principios éticos para identificar posibles sesgos y preocupaciones éticas. 1.3. Desarrollar estrategias para abordar las cuestiones éticas en las aplicaciones de IA, incluidos los posibles conflictos entre los principios y las compensaciones. Resultado de la destreza2: Analizar y evaluar diversos marcos y directrices éticas en escenarios del mundo real que involucran sistemas de IA. 2.1. Analizar diferentes marcos éticos y directrices relevantes para la IA y el sesgo algorítmico. 2.2. Aplicar marcos éticos a escenarios de IA del mundo real para evaluar las dimensiones éticas de los sistemas de IA. 2.3. Comparar y contrastar diversos marcos éticos para determinar su idoneidad para abordar el sesgo algorítmico en aplicaciones específicas de IA. Resultado 3: Diseñar y desarrollar sistemas éticos de IA promoviendo la transparencia, la rendición de cuentas y la participación de las partes interesadas. 3.1. Diseñar sistemas de IA que promuevan la transparencia, permitiendo a los usuarios y partes interesadas comprender el proceso de toma de decisiones. 3.2. Aplicar medidas de rendición de cuentas para garantizar que los desarrolladores y las organizaciones de IA asuman la responsabilidad de las consecuencias de sus sistemas de IA. 3.3. Facilitar la participación de las partes interesadas en el desarrollo de la IA, incluida la solicitud de aportaciones desde diversas perspectivas y la consideración de las necesidades de las diversas comunidades afectadas por los sistemas de IA.	Competencia Resultado 1: Mentalidad ética en el desarrollo y despliegue de la IA. 1.1. Desarrollar un sentido de responsabilidad hacia la aplicación de principios éticos en el desarrollo de la IA para minimizar el sesgo algorítmico. 1.2. Cultivar una mentalidad ética que valore la equidad, la inclusión y la justicia en las aplicaciones de IA. 1.3. Apreciar la importancia de abordar las preocupaciones éticas de manera proactiva para evitar consecuencias no deseadas de los sistemas de IA. Competencia Resultado 2: Curiosidad y apertura para aprender sobre diversos marcos éticos y directrices en escenarios del mundo real que involucran sistemas de IA. 2.1. Mostrar curiosidad y apertura al aprendizaje sobre los diferentes marcos éticos y su relevancia para el desarrollo de la IA. 2.2. Valorar la importancia de las directrices éticas a la hora de dar forma al diseño y despliegue de los sistemas de IA. 2.3. Adoptar un enfoque reflexivo para evaluar y perfeccionar las aplicaciones de IA sobre la base de consideraciones éticas. Competencia Resultado 3: Mentalidad alentadora y entusiasta para fomentar la transparencia, la rendición de cuentas y la participación de las partes interesadas en el desarrollo ético de la IA. 3.1. Fomentar el compromiso de promover la transparencia en los sistemas de IA, permitiendo a los usuarios y partes interesadas comprender y confiar en la toma de decisiones en materia de IA. 3.2. Valorar la importancia de la rendición de cuentas en el desarrollo de la IA, reconociendo la necesidad de que las organizaciones asuman la responsabilidad de las consecuencias de sus sistemas de IA. 3.3. Fomentar la participación y colaboración activa de las partes interesadas en el desarrollo de la IA, apreciando el valor de las diversas perspectivas y la inclusión de las comunidades afectadas.
COMPETENCIA/S DE LOS MARCOS DIGCOMP2.2., ENTRECOMP Y GREENCOMP CON LOS QUE SE VINCULA CU4 - Sabe que la IA per se no es ni buena ni mala. Lo que determina si los resultados de un sistema de IA son positivos o negativos para la	**COMPETENCIAS DE LOS MARCOS DIGCOMP2.2., ENTRECOMP Y GREENCOMP CON LOS QUE SE VINCULA CU4** - Sabe identificar áreas en las que la IA puede aportar beneficios a diversos aspectos de la vida cotidiana. Por ejemplo, en el ámbito de la salud, la IA podría	**COMPETENCIAS DE LOS MARCOS DIGCOMP2.2., ENTRECOMP Y GREENCOMP CON LOS QUE SE VINCULA CU4** - Disposición a contemplar cuestiones éticas relacionadas con los sistemas de IA (por ejemplo, ¿en qué contextos, como la condena de delincuentes, no deben utilizarse

sociedad es cómo se diseña y utiliza el sistema de IA, quién lo hace y con qué fines. (AI) (DigiComp 2.2.) - Sabe que el tratamiento de datos personales está sujeto a normativas locales como el Reglamento General de Protección de Datos (RGPD) de la UE (por ejemplo, las interacciones de voz con un asistente virtual son datos personales en términos del RGPD y pueden exponer a los usuarios a ciertos riesgos de protección de datos, privacidad y seguridad). (AI) (DigiComp 2.2.)	contribuir al diagnóstico precoz, mientras que en la agricultura podría utilizarse para detectar infestaciones de plagas. (IA). (DigiComp 2.2.) - Trabajar con otras personas (p. ej., los hilos trabajan juntos, formar un equipo, ampliar su red): Formar un equipo, trabajar juntos y establecer contactos. Trabajar juntos y cooperar con otros para desarrollar ideas y convertirlas en acción. Resolver conflictos y enfrentarse a la competencia de forma positiva Cuando sea necesario. (EntreComp) Encarnar los valores de la sostenibilidad: 1.2 Apoyo a la equidad: apoyar la equidad y la justicia para las generaciones actuales y futuras y aprender de las generaciones anteriores para la sostenibilidad. 1.3 Promoción de la naturaleza: reconocer que los seres humanos son parte de la naturaleza; y respetar las necesidades y los derechos de otras especies y de la propia naturaleza para restaurar y regenerar ecosistemas sanos y resilientes. (GreenComp)	las recomendaciones de IA sin intervención humana)? (AI) (DigiComp 2.2.) - Abierto a los sistemas de IA que ayudan a los humanos a tomar decisiones informadas de acuerdo con sus objetivos (por ejemplo, los usuarios deciden activamente si actuar sobre una recomendación o no). (IA). (DigiComp 2.2.) - Considera la ética (incluidas, entre otras, la agencia y la supervisión humanas, la transparencia, la no discriminación, la accesibilidad, los sesgos y la equidad) como uno de los pilares fundamentales a la hora de desarrollar o implementar sistemas de IA. (AI) (DigiComp 2.2.) - Abierto a participar en procesos colaborativos para co-diseñar y co-crear nuevos productos y servicios basados en sistemas de IA para apoyar y mejorar la participación de los ciudadanos en la sociedad. (AI) (DigiComp 2.2.) Creatividad (p. ej., ser curioso y abierto): Desarrollar ideas creativas y con propósito. Desarrollar varias ideas y oportunidades para crear valor, incluyendo mejores soluciones a los desafíos existentes y nuevos. Explora y experimenta con enfoques innovadores. Combinar conocimientos y recursos para lograr efectos valiosos. (EntreComp) Movilizar a otros: Inspirar y entusiasmar a las partes interesadas relevantes. Obtenga el apoyo necesario para lograr resultados valiosos. Demostrar comunicación, persuasión, negociación y liderazgo efectivos. (EntreComp)

CU4 DESCRIPCIÓN DE LOS RESULTADOS DEL CONOCIMIENTO

4.1. Conocer los principios éticos en el desarrollo y despliegue de la IA.

Consciente de los principios éticos en el desarrollo y despliegue de la IA. Aplicar principios éticos al diseño, desarrollo e implementación de sistemas de IA para minimizar el sesgo algorítmico. Evaluar los sistemas de IA en función de las directrices y principios éticos para identificar posibles sesgos y preocupaciones éticas. Desarrollar estrategias para abordar las cuestiones éticas en las aplicaciones de IA, incluidos los posibles conflictos entre los principios y las compensaciones.

Principios éticos de la IA: Los principios éticos tienen como objetivo proporcionar una base para que los sistemas de IA funcionen por el bien de la humanidad, las personas, las sociedades, el medio ambiente y los ecosistemas,

y para prevenir daños. Guían el diseño, el desarrollo, la implementación y el uso de los sistemas de IA de una manera que sea confiable y ponga en el centro la dignidad humana, la igualdad de todos los seres humanos, la preservación del medio ambiente, la biodiversidad y los ecosistemas, el respeto por la diversidad cultural y la responsabilidad de los datos (UNESCO, 2022). Los principios éticos describen lo que se espera en términos de lo correcto y lo incorrecto y otras normas éticas. Los principios éticos de la IA se refieren a las restricciones normativas sobre lo que se debe y no se debe hacer en el uso algorítmico en la sociedad (Zhou et al., 2020). Por ejemplo, según el Grupo de Expertos de Alto Nivel sobre IA (AI HLEG, por sus siglas en inglés), estos principios incluyen: el respeto de la autonomía humana, la prevención de daños, la equidad y la explicabilidad (Comisión Europea, 2019).

Aplicación práctica de los principios éticos en la IA. Una vez identificados, los principios éticos deben traducirse en herramientas y directrices viables para dar forma a la innovación basada en la IA y apoyar la aplicación práctica de los principios éticos de la IA. Es muy necesario contar con herramientas y directrices sobre cómo aplicar los principios éticos en el diseño, la implementación y el despliegue de la IA. Los algoritmos, arquitecturas e interfaces de IA ética siguen los principios éticos de la IA. (Zhou et al., 2020). En el curso de su trabajo en el desarrollo de los principios éticos de la IA, muchas organizaciones privadas, públicas y sin fines de lucro han identificado acciones útiles que pueden ayudar en la articulación e implementación de los principios propuestos (por ejemplo, UNESCO, 2021).

Por ejemplo, los principios esbozados por el Grupo de Expertos de Alto Nivel sobre IA deben traducirse en requisitos concretos para lograr una IA fiable. Estos requisitos son aplicables a las diferentes partes interesadas que participan en el ciclo de vida de los sistemas de IA: desarrolladores, implementadores y usuarios finales, así como a la sociedad en general. Los diferentes grupos de partes interesadas tienen diferentes funciones que desempeñar para garantizar que se cumplan los requisitos: *los desarrolladores* deben implementar y aplicar los requisitos a los procesos de diseño y desarrollo; *Los implementadores* deben asegurarse de que los sistemas que utilizan y los productos y servicios que ofrecen cumplen los requisitos; los usuarios finales y la sociedad en general deben estar informados de estos requisitos y poder solicitar que se cumplan. La siguiente lista de requisitos no es exhaustiva.

Incluye aspectos sistémicos, individuales y sociales:

1. Agencia y supervisión humanas (es decir, derechos fundamentales, agencia y supervisión humanas).
2. Robustez técnica y seguridad (es decir, resiliencia a los ataques y seguridad, plan alternativo y seguridad general, precisión, fiabilidad y reproducibilidad).
3. Privacidad y gobernanza de datos (es decir, respeto por la privacidad, la calidad e integridad de los datos y el acceso a los datos).
4. Transparencia (es decir, trazabilidad, explicabilidad y comunicación).
5. Diversidad, no discriminación y equidad (es decir, evitar sesgos injustos, accesibilidad y diseño universal, y participación de las partes interesadas).
6. Bienestar social y medioambiental (es decir, sostenibilidad y respeto al medio ambiente, impacto social, sociedad y democracia).
7. Rendición de cuentas (es decir, auditabilidad, minimización y notificación de los efectos negativos, compensaciones y compensación) (Comisión Europea, 2019).

Morley et al. (2019) han construido una tipología combinando los principios éticos con las etapas del ciclo de vida de la IA para garantizar que el sistema de IA se diseñe, implemente y despliegue de manera ética. La tipología indica que cada principio ético debe tenerse en cuenta en cada etapa del ciclo de vida de la IA. La tipología proporciona una breve instantánea de las herramientas que están disponibles actualmente para los desarrolladores de IA para fomentar la progresión de la IA ética desde los principios hasta la práctica. La tipología completa de Morley et al. se puede encontrar en https://tinyurl.com/AppliedAIEthics.

4.2. Conocer diversos marcos y directrices éticas y su aplicación en escenarios del mundo real que involucran sistemas de IA.

Conocer diversos marcos y directrices éticas y su aplicación en escenarios del mundo real que involucran sistemas de IA. Analizar diferentes marcos éticos y directrices relevantes para la IA y el sesgo algorítmico. Aplicación de marcos éticos a escenarios de IA del mundo real para evaluar las dimensiones éticas de los sistemas de IA. Comparar y contrastar diversos marcos éticos para determinar su idoneidad para abordar el sesgo algorítmico en aplicaciones específicas de IA.

Marcos y directrices para la ética de la IA: Los marcos y directrices éticos consisten en valores, principios, normas y acciones para guiar a las personas, los grupos, las comunidades, las instituciones y las empresas del sector privado para garantizar la integración de la ética en todas las etapas del ciclo de vida del sistema de IA. Su objetivo es proteger, promover y respetar los derechos humanos y las libertades fundamentales, la dignidad humana y la igualdad; salvaguardar los intereses de las generaciones presentes y futuras; preservar el medio ambiente, la biodiversidad y los ecosistemas; y respetar la diversidad cultural en todas las etapas del ciclo de vida del sistema de IA (Comisión Europea, 2019; UNESCO, 2021).

Por lo general, se publican varios conjuntos de principios y marcos éticos para la IA de la industria (por ejemplo, Google, IBM, Microsoft, Intel), el gobierno (por ejemplo, el Comité Selecto de los Lores del Reino Unido, el Grupo de Expertos de Alto Nivel de la Comisión Europea) y el mundo académico (por ejemplo, *Future of Life Institute*, *IEEE, AI4People*) (Zhou et al., 2020). Ningún principio ético único está respaldado explícitamente por todos los marcos y directrices éticas existentes, pero existe una convergencia emergente en torno a los siguientes principios: transparencia, justicia y equidad, responsabilidad, no maleficencia, privacidad, beneficencia, libertad y autonomía, confianza, sostenibilidad, dignidad y solidaridad (Jobin et al, 2019).

La UNESCO (2021) ha publicado en noviembre de 2021 la primera norma mundial sobre la ética de la IA: la Recomendación sobre la ética de la inteligencia artificial. Este marco fue adoptado por los 193 Estados miembros. La protección de los derechos humanos y de la dignidad es la piedra angular de la Recomendación, basada en la promoción de sus principios fundamentales.

4.3. Comprender la importancia de la transparencia, la rendición de cuentas y la participación de las partes interesadas en el proceso de desarrollo ético de la IA

Comprender la importancia de la transparencia, la rendición de cuentas y la participación de las partes interesadas en el desarrollo ético de la IA. Diseñar sistemas de IA que promuevan la transparencia, permitiendo a los usuarios y partes interesadas comprender el proceso de toma de decisiones. Implementar medidas de rendición de cuentas para garantizar que los desarrolladores y las organizaciones de IA asuman la responsabilidad de las consecuencias de sus sistemas de IA. Facilitar la participación de las partes interesadas en el desarrollo de la IA, incluida la solicitud de aportaciones desde

diversas perspectivas y la consideración de las necesidades de las diversas comunidades afectadas por los sistemas de IA.

Transparencia: La transparencia (explicabilidad) es crucial para generar y mantener la confianza de los usuarios en los sistemas de IA. Esto significa que los procesos deben ser explícitos, que las capacidades y el propósito de los sistemas de IA deben comunicarse abiertamente y que las decisiones, en la medida de lo posible, deben ser explicables a las personas directa e indirectamente afectadas. Sin esa información, una decisión no puede ser debidamente impugnada.

No siempre es posible explicar por qué un modelo ha generado un resultado o una decisión en particular (y qué combinación de factores de entrada contribuyó a ello). Estos casos se conocen como algoritmos de "caja negra" y requieren una atención especial. En tales circunstancias, pueden ser necesarias otras medidas de explicabilidad (por ejemplo, trazabilidad, auditabilidad y comunicación transparente sobre las capacidades del sistema), siempre que el sistema en su conjunto respete los derechos fundamentales. El grado de necesidad de explicabilidad depende en gran medida del contexto y de la gravedad de las consecuencias si ese resultado es erróneo o inexacto (Comisión Europea, 2019).

Rendición de cuentas: La rendición de cuentas es una de las piedras angulares de la gobernanza de la inteligencia artificial (IA). La rendición de cuentas tiene muchas definiciones, pero, en esencia, es una obligación de informar y justificar la propia conducta ante una autoridad (Novelli, Taddeo y Floridi, 2023). En términos generales, la rendición de cuentas en IA se relaciona con la expectativa de que los diseñadores, desarrolladores e implementadores cumplan con los estándares y la legislación para garantizar el correcto funcionamiento de las IA durante su ciclo de vida (Fjeld et al., 2020).

Deben desarrollarse mecanismos adecuados de supervisión, evaluación de impacto, auditoría y diligencia debida, incluida la protección de los denunciantes, para garantizar la rendición de cuentas de los sistemas de IA y su impacto a lo largo de su ciclo de vida. Tanto los diseños técnicos como los institucionales deben garantizar la auditabilidad y la trazabilidad de (el funcionamiento de) los sistemas de IA, en particular para abordar cualquier conflicto con las normas de derechos humanos y las amenazas al bienestar ambiental y de los ecosistemas (UNESCO, 2021)

Participación de las partes interesadas: Es importante consultar a los grupos de partes interesadas pertinentes a la hora de desarrollar y gestionar el uso de aplicaciones de IA (Fjeld et al., 2020). La participación de las diferentes partes interesadas a lo largo del ciclo de vida del sistema de IA es necesaria para que los enfoques inclusivos de la gobernanza de la IA permitan que todos compartan los beneficios y contribuyan al desarrollo sostenible. Las partes interesadas incluyen, entre otras, los gobiernos, las organizaciones intergubernamentales, la comunidad técnica, la sociedad civil, los investigadores y el mundo académico, los medios de comunicación, la educación, los responsables políticos, las empresas del sector privado, las instituciones de derechos humanos y los organismos de igualdad, los organismos de supervisión de la lucha contra la discriminación y los grupos de jóvenes y niños (UNESCO, 2021).

Involucrar a las partes interesadas ayuda a garantizar que las tecnologías de IA se desarrollen e implementen de una manera que se alinee con los valores, las necesidades y las preocupaciones de varias personas y grupos afectados por la tecnología. Sin embargo, en la práctica puede ser difícil identificar a las personas y organizaciones afectadas por el desarrollo y el uso de sistemas de IA. Estos problemas se vuelven aún más importantes Cuando los sistemas de IA se desarrollan en proyectos, que son organizaciones temporales cuyos miembros del equipo pueden no estar disponibles una vez que finaliza el proyecto. Algunos sistemas de IA se ejecutan automáticamente y no permiten la intervención humana, lo que los hace únicos de otros sistemas de información. Estos sistemas pueden afectar los derechos humanos, civiles y de los trabajadores; tienen características cercanas a las industrias farmacéutica y de la salud por su impacto humano y a los proyectos de infraestructura por sus impactos ambientales (Miller, 2022).

CU5 - Estudios de Caso y Proyecto EQF6		
RESULTADOS DE APRENDIZAJE		
Al finalizar la experiencia, el alumno será capaz de:		
CONOCIMIENTOS	**DESTREZAS**	**COMPETENCIAS**
Conocimiento teórico/fáctico en: CUC5.1. Comprender el sesgo algorítmico y sus implicaciones en el mundo real. CUC5.2. Consciente de las complejidades y matices del sesgo algorítmico. CUC5.3. Reconocer las estrategias y herramientas disponibles para abordar los problemas de sesgo algorítmico en la práctica.	Resultado de la destreza1: Aplicar el conocimiento del sesgo algorítmico a estudios de casos del mundo real y a un proyecto práctico. 1.1. Investigar estudios de casos del mundo real para identificar casos de sesgo algorítmico y sus consecuencias. 1.2. Desarrollar soluciones para abordar el sesgo algorítmico en los estudios de caso, teniendo en cuenta las implicaciones éticas y prácticas. 1.3. Diseñar, implementar y evaluar un proyecto práctico para abordar el sesgo algorítmico en un dominio elegido. Habilidad Resultado 2: Analizar el sesgo algorítmico, sus causas e impactos. 2.1. Analizar diferentes tipos de sesgos algorítmicos y sus causas raíz. 2.2. Evaluar el impacto potencial del sesgo algorítmico en los individuos y en la sociedad en su conjunto. 2.3. Comprender los desafíos y limitaciones para identificar, medir y mitigar los sesgos algorítmicos. Resultado de la destreza3: Revisar e integrar las estrategias y herramientas disponibles para abordar los problemas de sesgo algorítmico. 3.1. Investigar y evaluar las estrategias existentes para abordar el sesgo algorítmico, incluido el aprendizaje automático consciente de la equidad y las auditorías de sesgo. 3.2. Desarrollar la competencia en el uso de herramientas y técnicas para detectar, medir y mitigar el sesgo algorítmico en los sistemas de IA. 3.3. Integrar enfoques interdisciplinarios, como las aportaciones de expertos en la materia y perspectivas diversas, para mejorar la equidad de los sistemas de IA.	Competencia Resultado 1: Compromiso en la mentalidad de resolución de problemas para abordar el sesgo algorítmico y sus implicaciones en el mundo real. 1.1. Demostrar un compromiso con la aplicación de los conocimientos teóricos a situaciones prácticas con el fin de abordar el sesgo algorítmico. 1.2. Valorar la importancia de la experiencia práctica para obtener una comprensión más profunda del sesgo algorítmico y sus implicaciones en el mundo real. 1.3. Desarrollar una mentalidad de resolución de problemas que busque crear soluciones de IA justas y equitativas. Competencia Resultado 2: Mentalidad crítica y colaborativa en el reconocimiento de las complejidades y matices del sesgo algorítmico. 2.1. Fomentar la empatía hacia las personas y comunidades afectadas por el sesgo algorítmico. 2.2. Cultivar una mentalidad crítica que reconozca las complejidades y limitaciones de abordar el sesgo algorítmico. 2.3. Apreciar la importancia de la colaboración interdisciplinaria y las diversas perspectivas para abordar los desafíos del sesgo algorítmico. Competencia Resultado 3: Mentalidad abierta y colaborativa en la utilización de estrategias y herramientas disponibles para abordar los problemas de sesgo algorítmico en la práctica. 3.1. Adoptar un enfoque proactivo en la utilización de las estrategias y herramientas disponibles para combatir el sesgo algorítmico. 3.2. Valorar la importancia del aprendizaje continuo y mantenerse informado sobre los nuevos desarrollos para abordar el sesgo algorítmico. 3.3. Fomentar un entorno colaborativo que fomente el diálogo abierto, el intercambio de conocimientos y la innovación para abordar los problemas de sesgo algorítmico.

COMPETENCIA/S DE LOS MARCOS DIGCOMP2.2., ENTRECOMP Y GREENCOMP CON LOS QUE SE VINCULA CU5 - Consciente de que es posible que los algoritmos de IA no estén configurados para proporcionar solo la información que el usuario desea; También pueden incorporar un mensaje comercial o político (por ejemplo, para animar a los usuarios a permanecer en el sitio, a ver o comprar algo en particular, a compartir opiniones específicas). Esto también puede tener consecuencias negativas (por ejemplo, reproducir estereotipos, compartir información errónea). (AI) (DigiComp 2.2.) - Ser conscientes de que los datos, de los que depende la IA, pueden incluir sesgos. Si es así, estos sesgos pueden automatizarse y empeorar con el uso de la IA. Por ejemplo, los resultados de búsqueda sobre la ocupación pueden incluir estereotipos sobre trabajos masculinos o femeninos (por ejemplo, conductores de autobús masculinos, vendedoras). (AI) (DigiComp 2.2.) - Sabe que los datos recopilados y procesados, por ejemplo, por los sistemas en línea, pueden ser utilizados para reconocer patrones (p. ej., repeticiones) en nuevos datos (p. ej., otras imágenes, sonidos, clics del ratón, comportamientos en línea) para optimizar y personalizar aún más los servicios en línea (por ejemplo, anuncios). (DigiComp 2.2.) - Conscientes de que los sensores utilizados en muchas tecnologías y aplicaciones digitales (por ejemplo, cámaras de seguimiento facial, asistentes virtuales, tecnologías portátiles, teléfonos móviles, dispositivos inteligentes) generan grandes cantidades de datos, incluidos datos personales, que pueden utilizarse para entrenar un sistema de IA. (AI) (DigiComp 2.2.) - Ser consciente de que todo lo que se comparte públicamente en línea (por ejemplo, imágenes, vídeos, sonidos) puede utilizarse	**COMPETENCIAS DE LOS MARCOS DIGCOMP2.2., ENTRECOMP Y GREENCOMP CON LOS QUE SE VINCULA CU5** · Pensamiento ético y sostenible (por ejemplo, los hilos se comportan éticamente, son responsables, evalúan el impacto): Evalúan las consecuencias y el impacto de las ideas, oportunidades y acciones. (...) Actúe con responsabilidad. (EntreComp) Encarnar los valores de la sostenibilidad: 1.2 Apoyo a la equidad: apoyar la equidad y la justicia para las generaciones actuales y futuras y aprender de las generaciones anteriores para la sostenibilidad. 1.3 Promoción de la naturaleza: reconocer que los seres humanos son parte de la naturaleza; y respetar las necesidades y los derechos de otras especies y de la propia naturaleza para restaurar y regenerar ecosistemas sanos y resilientes. (GreenComp)	**COMPETENCIAS DE LOS MARCOS DIGCOMP2.2., ENTRECOMP Y GREENCOMP CON LOS QUE SE VINCULA CU5** - Abierto a los sistemas de IA que ayudan a los humanos a tomar decisiones informadas de acuerdo con con sus objetivos (por ejemplo, que los usuarios decidan activamente si actuar sobre una recomendación o no). (AI) DigiComp 2.2.) - Disposición a contemplar cuestiones éticas relacionadas con los sistemas de IA (por ejemplo, ¿en qué contextos, como la condena de delincuentes, no deben utilizarse las recomendaciones de IA sin intervención humana)? (AI) (DigiComp 2.2.) - Identifica las implicaciones positivas y negativas del uso de todos los datos (recopilación, codificación y procesamiento), pero especialmente de los datos personales, por parte de las tecnologías digitales impulsadas por la IA, como las aplicaciones y los servicios en línea. (AI) (DigiComp 2.2.) - Considera las consecuencias éticas de los sistemas de IA a lo largo de su ciclo de vida: incluyen tanto el impacto ambiental (consecuencias ambientales de la producción de dispositivos y servicios digitales) como el impacto social, por ejemplo, *la plataformización* del trabajo y la gestión algorítmica que puede reprimir la privacidad o los derechos de los trabajadores; el uso de mano de obra de bajo costo para etiquetar imágenes para entrenar sistemas de IA. (AI) (DigiComp 2.2.) - Abierto a participar en procesos colaborativos para co-diseñar y co-crear nuevos productos y servicios basados en sistemas de IA para apoyar y mejorar la participación de los ciudadanos en la sociedad. (AI) (DigiComp 2.2.)

para entrenar sistemas de IA. Por ejemplo, las empresas de software comercial que desarrollan sistemas de reconocimiento facial de IA pueden utilizar imágenes personales compartidas en línea (por ejemplo, fotografías familiares) para entrenar y mejorar la capacidad del software para reconocer automáticamente a esas personas en otras imágenes, lo que podría no ser deseable (por ejemplo, podría ser una violación de la privacidad). (AI) (DigiComp 2.2.) - Reconoce que, si bien la aplicación de sistemas de IA en muchos ámbitos no suele ser controvertida (por ejemplo, la IA que ayuda a evitar el cambio climático), la IA que interactúa directamente con los humanos y toma las decisiones sobre su vida a menudo pueden ser controvertidas (por ejemplo, software de clasificación de CV para procedimientos de contratación, calificación de exámenes que pueden determinar el acceso a la educación). (AI) (DigiComp 2.2.) - Capaz de identificar algunos ejemplos de sistemas de IA: recomendaciones de productos (por ejemplo, en sitios de compras en línea), reconocimiento de voz (por ejemplo, por asistentes virtuales), reconocimiento de imágenes (por ejemplo, para detectar tumores en radiografías) y reconocimiento facial (por ejemplo, en sistemas de vigilancia). (AI) (DigiComp 2.2.)		

CU5 DESCRIPCIÓN DE LOS RESULTADOS DEL CONOCIMIENTO

5.1. Comprender el sesgo algorítmico y sus implicaciones en el mundo real

Comprender el sesgo algorítmico y sus implicaciones en el mundo real. Aplicación del conocimiento del sesgo algorítmico a estudios de casos del mundo real y a un proyecto práctico. Investigar estudios de casos del mundo real para identificar casos de sesgo algorítmico y sus consecuencias. Desarrollar soluciones para abordar el sesgo algorítmico en los estudios de caso, teniendo en cuenta las implicaciones éticas y prácticas. Diseñar,

implementar y evaluar un proyecto práctico para abordar el sesgo algorítmico en un dominio elegido.

Sesgo algorítmico: El sesgo algorítmico se refiere a la presencia de resultados injustos o discriminatorios en los resultados producidos por un sistema algorítmico. Se produce Cuando un algoritmo favorece o perjudica de forma sistemática a determinadas personas o grupos en función de características específicas como la raza, el género o el nivel socioeconómico. En el aprendizaje automático, los algoritmos se basan en múltiples conjuntos de datos, es decir, datos de entrenamiento que especifican cuáles son los resultados correctos para algunas personas u objetos. A partir de esos datos de entrenamiento, la máquina aprende un modelo que se puede aplicar a otras personas u objetos y hace predicciones sobre los posibles resultados para ellos. Sin embargo, las máquinas pueden tratar de manera diferente a personas y objetos en situaciones similares. Debido a este déficit, algunos algoritmos corren el riesgo de replicar e incluso amplificar los sesgos humanos, en particular los que afectan a los grupos protegidos (Friedman y Nissenbaum, 1996; Lange y Duarte, 2017; Lee, Resnick y Barton, 2019).

Implicaciones del sesgo algorítmico: El sesgo algorítmico puede manifestarse de varias maneras con diversos grados de consecuencias para el grupo de sujetos. Los siguientes ejemplos ilustran una serie de causas y efectos que, o bien aplican inadvertidamente un tratamiento diferente a los grupos, o bien generan deliberadamente un impacto dispar en ellos (Lee, Resnick y Barton, 2019):

- *Sesgo en las herramientas de reclutamiento en línea*: el minorista en línea Amazon suspendió recientemente el uso de un algoritmo de reclutamiento después de descubrir un sesgo de género. El software de IA penalizaba cualquier currículum que contuviera la palabra "mujeres" en el texto y degradaba los currículos de las mujeres que asistían a universidades femeninas, lo que daba lugar a un sesgo de género.
- *Sesgo en las asociaciones de palabras*: los investigadores de la Universidad de Princeton utilizaron un software de IA de aprendizaje automático listo para usar para analizar y vincular 2,2 millones de palabras. Descubrieron que los nombres europeos se percibían como más agradables que los de los afroamericanos, y que las palabras "mujer" y "niña" tenían más probabilidades de asociarse con las artes en lugar de las ciencias y las matemáticas, que probablemente estaban conectadas con los hombres.

- *Sesgo en los anuncios en línea*: Latanya Sweeney, investigadora de Harvard y ex directora de tecnología de la Comisión Federal de Comercio (FTC, por sus siglas en inglés), descubrió que las consultas de búsqueda en línea de nombres afroamericanos tenían más probabilidades de devolver anuncios a esa persona desde un servicio que genera registros de arrestos, en comparación con los resultados de los anuncios para nombres blancos.
- *Sesgo en la tecnología de reconocimiento facial*: la investigadora del MIT Joy Buolamwini descubrió que los algoritmos que impulsan tres sistemas de software de reconocimiento facial disponibles comercialmente no reconocían las pieles más oscuras. En general, se estima que la mayoría de los conjuntos de datos de entrenamiento de reconocimiento facial son más del 75 por ciento masculinos y más del 80 por ciento blancos. Cuando la persona en la foto era un hombre blanco, el software era preciso el 99 por ciento de las veces para identificar a la persona como hombre.
- *Sesgo en los algoritmos de justicia penal* – El algoritmo COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), que es utilizado por los jueces para predecir si los acusados deben ser detenidos o liberados bajo fianza en espera de juicio, se encontró sesgado contra los afroamericanos, según un informe de ProPublica.

5.2. Conscientes de las complejidades y matices del sesgo algorítmico

Consciente de las complejidades y matices del sesgo algorítmico. Análisis del sesgo algorítmico, sus causas e impacto. Analizar diferentes tipos de sesgos algorítmicos y sus causas raíz. Evaluación del impacto del sesgo algorítmico. Comprender las limitaciones para identificar, medir y mitigar el sesgo algorítmico.

Tipos y causas del sesgo algorítmico: Existen varias formas de sesgo algorítmico. Según Friedman y Nissenbaum (1996) los diferentes tipos y causas de los sesgos algorítmicos se pueden dividir en tres categorías:

1. El sesgo preexistente tiene sus raíces en las instituciones, prácticas y competencias sociales. Cuando los sistemas informáticos encarnan sesgos que existen de forma independiente, y por lo general antes de la creación del sistema, entonces el sistema ejemplifica un sesgo preexistente. El sesgo preexistente puede entrar en un sistema ya sea a través de los esfuerzos

explícitos y conscientes de individuos o instituciones, o implícita e inconscientemente, incluso a pesar de las mejores intenciones.

2. El sesgo técnico surge de limitaciones o consideraciones técnicas. Las fuentes de sesgo técnico se pueden encontrar en varios aspectos del proceso de diseño, incluidas las limitaciones de las herramientas informáticas como el hardware, el software y los periféricos; el proceso de atribuir significado social a algoritmos desarrollados fuera de contexto; imperfecciones en la generación de números pseudoaleatorios; y el intento de hacer que las construcciones humanas sean susceptibles a las computadoras.

3. El sesgo emergente surge en un contexto de uso con usuarios reales. Este sesgo suele surgir en algún momento después de que se completa un diseño, como resultado de un cambio en el conocimiento social, la población o los valores culturales. Es probable que las interfaces de usuario sean particularmente propensas a los sesgos emergentes porque las interfaces, por diseño, buscan reflejar las capacidades, el carácter y los hábitos de los posibles usuarios. Por lo tanto, un cambio en el contexto de uso puede crear dificultades para un nuevo grupo de usuarios. El sesgo emergente se origina a partir de la aparición de nuevos conocimientos en la sociedad que no pueden o no se incorporan al sistema, o cuando la población que utiliza el sistema difiere en alguna dimensión significativa de la población asumida como usuarios en el diseño, es decir, desajuste entre los usuarios y el diseño del sistema.

5.3. Reconocer las estrategias y herramientas disponibles para abordar los problemas de sesgo algorítmico en la práctica.

Reconocer las estrategias y herramientas disponibles para abordar los problemas de sesgo algorítmico, incluido el aprendizaje automático consciente de la equidad y las auditorías de sesgo. Revisar e integrar las estrategias y herramientas disponibles para abordar los problemas de sesgo algorítmico en la práctica. Investigar y evaluar las estrategias existentes para abordar el sesgo algorítmico, incluido el aprendizaje automático consciente de la equidad y las auditorías de sesgo. Desarrollar la competencia en el uso de herramientas y técnicas para detectar, medir y mitigar el sesgo algorítmico en los sistemas de IA. Integrar enfoques interdisciplinarios, como las aportaciones de expertos en la materia y diversas perspectivas, para mejorar la equidad de los sistemas de IA.

Estrategias y herramientas para abordar los problemas de sesgo algorítmico: Abordar el sesgo algorítmico requiere una consideración cuidadosa en varias etapas del desarrollo del algoritmo. Implica garantizar datos de entrenamiento diversos y representativos, identificar y mitigar los sesgos en el diseño del algoritmo y el proceso de toma de decisiones, y realizar pruebas y evaluaciones rigurosas para detectar y rectificar problemas relacionados con los sesgos. Se han realizado algunos esfuerzos para minimizar los resultados discriminatorios mediante el desarrollo de marcos y directrices para mitigar el sesgo algorítmico, incluido el aumento de la transparencia y la rendición de cuentas en los sistemas algorítmicos, las medidas regulatorias y la adopción de técnicas de aprendizaje automático conscientes de la equidad (Lee, Resnick y Barton, 2019).

4.2. CHARLIE WP2 "Micro credencial ética de IA" EQF4

Este programa educativo en profundidad está diseñado para proporcionar una comprensión integral de los sesgos algorítmicos y sus implicaciones en la sociedad actual impulsada por los datos. Dado que los algoritmos desempeñan un papel importante en la configuración de nuestras vidas y en la toma de decisiones en diversos sectores, incluidos la atención médica, la educación, las finanzas y el gobierno, es crucial que los profesionales y académicos/estudiantes desarrollen una comprensión sólida de los posibles sesgos dentro de estos sistemas.

El objetivo de este curso es dotar a los participantes de los conocimientos y habilidades necesarios para identificar, mitigar y abordar los sesgos algorítmicos. A través de una combinación de fundamentos teóricos y aplicaciones prácticas, los alumnos obtendrán información sobre los aspectos éticos, sociales y técnicos del sesgo algorítmico.

Estos objetivos contribuyen a los objetivos generales al sentar las bases de las siguientes actividades, permitiendo a los docentes, mentores y educandos tener una visión clara de los objetivos de los diferentes itinerarios de aprendizaje, trabajando para aumentar la capacidad de las organizaciones de educación superior, adultos y jóvenes para proporcionar oportunidades de aprendizaje que satisfagan las necesidades de la sociedad, pero que también se adapten a las necesidades de aprendizaje de los educandos.

Concretamente, los resultados esperados de este curso "Micro credencial ética de la IA" EQF4 están dirigidos a:

1- Desarrollar una comprensión fundamental del sesgo algorítmico, sus causas y consecuencias en los individuos y la sociedad.

1.1. Adquirir conocimientos sobre la definición, fuentes y manifestaciones del sesgo algorítmico.

1.2. Comprender las implicaciones sociales e individuales de los algoritmos sesgados.

2. Fomentar la concienciación y la aplicación del principio ético de no maleficencia en el desarrollo de la IA.

2.1. Conozca los riesgos y daños asociados con los algoritmos sesgados.

2.2. Desarrollar estrategias para minimizar el daño y promover el desarrollo ético de la IA.

3. Comprender la importancia de la rendición de cuentas en los sistemas de IA y explorar los marcos legales y éticos relevantes.

3.1. Examinar el papel de las distintas partes interesadas en la rendición de cuentas de la IA.

3.2. Conocer las mejores prácticas para garantizar la rendición de cuentas en el desarrollo de la IA.

4. Adquirir conocimientos sobre el concepto de transparencia en los sistemas de IA y su importancia en la toma de decisiones algorítmicas.

4.1. Explorar métodos y herramientas para mejorar la transparencia en la IA.

4.2. Comprender los desafíos y limitaciones de hacer que los algoritmos complejos sean más comprensibles.

5. Investigar la intersección de la IA, los derechos humanos y la equidad y su impacto en el desarrollo ético de la IA.

5.1. Examinar cómo los algoritmos sesgados pueden afectar los derechos humanos, como la no discriminación, la privacidad y la libertad de expresión.

5.2. Aprender estrategias para garantizar la justicia y la equidad en el desarrollo y la implementación de la IA.

6. Aplicar principios éticos en el desarrollo y despliegue de la IA a través de enfoques prácticos y escenarios del mundo real.

6.1. Explorar diversos marcos y directrices éticas y su aplicación a los sistemas de IA.

6.2. Comprender la importancia de la participación de las partes interesadas, la colaboración interdisciplinaria y la implementación de procesos éticos de desarrollo de IA.

Al finalizar este curso, los participantes habrán desarrollado una comprensión integral de los sesgos algorítmicos, su impacto potencial en varios sectores y las estrategias y herramientas disponibles para abordar estos sesgos. Este conocimiento será invaluable para profesionales y académicos/estudiantes que trabajan en campos donde se utilizan algoritmos para la toma de decisiones y ayudará a garantizar resultados más equitativos y justos en un mundo impulsado por datos.

4.2.1. Contenido del curso "Micro credencial ética de IA" EQF4

El curso de micro credenciales está estructurado en torno a 6 Unidades de Competencia (CU), cada una diseñada para equipar a los participantes con el conocimiento y las habilidades necesarias para navegar por los desafíos y oportunidades en el ámbito del sesgo algorítmico.

CU1 - ¿Qué es el sesgo algorítmico? En esta unidad, los estudiantes explorarán el concepto de sesgo algorítmico y sus diversas manifestaciones. Los temas tratados incluirán la definición de sesgo algorítmico, las causas y fuentes del sesgo en los algoritmos, y las posibles consecuencias de los algoritmos sesgados en los individuos y la sociedad.

CU2 - No maleficencia. Esta unidad se centra en el principio ético de la no maleficencia, que enfatiza la importancia de evitar daños en el desarrollo y despliegue de sistemas de IA. Los estudiantes aprenderán sobre los riesgos y daños potenciales asociados con los algoritmos sesgados, así como las estrategias para minimizar el daño y garantizar el desarrollo ético de la IA.

CU3 – Rendición de cuentas. En la unidad de Rendición de Cuentas, los estudiantes examinarán la importancia de establecer líneas claras de responsabilidad y rendición de cuentas en el desarrollo y uso de sistemas de IA. Entre los temas que se tratarán figurarán los marcos jurídicos y éticos de rendición de cuentas, el papel de las distintas partes interesadas y las mejores prácticas para garantizar la rendición de cuentas en el desarrollo de la IA.

CU4 – Transparencia. Esta unidad profundiza en el concepto de transparencia en los sistemas de IA, explorando la importancia de la apertura, la comunicación y la explicabilidad en la toma de decisiones algorítmicas. Los estudiantes aprenderán sobre métodos y herramientas

para mejorar la transparencia en la IA, así como los desafíos y limitaciones asociados con hacer que los algoritmos complejos sean más comprensibles.

CU5 - Derechos humanos y equidad En la unidad de derechos humanos y equidad. Los estudiantes explorarán la intersección de la IA, los derechos humanos y la equidad. Examinarán cómo los algoritmos sesgados pueden afectar los derechos humanos, incluido el derecho a la no discriminación, la privacidad y la libertad de expresión. Los estudiantes también aprenderán sobre estrategias para garantizar la justicia y la equidad en el desarrollo y la implementación de la IA.

CU6 - Ética de la IA, un enfoque práctico. Esta unidad se centra en la aplicación práctica de los principios éticos en el desarrollo y la implementación de la IA. Los estudiantes explorarán varios marcos y pautas éticas, aprendiendo a aplicarlos a escenarios del mundo real que involucran sistemas de IA. La unidad también cubrirá la importancia de la participación de las partes interesadas, la colaboración interdisciplinaria y la implementación de procesos éticos de desarrollo de IA.

4.2.2. Descripción del grupo destinatario y características del EQF4

El Marco Europeo de Cualificaciones (EQF, de las siglas en inglés) es un marco de referencia común que vincula los sistemas de Cualificaciones de diferentes países europeos, lo que facilita que las personas reconozcan sus Cualificaciones en toda Europa (Comisión Europea, 2023). El EQF consta de ocho niveles de referencia, cada uno de los cuales está definido por un conjunto de descriptores que indican los resultados del aprendizaje, las capacidades, las competencias y la autonomía que se esperan en ese nivel. El nivel 4 del EQF (EQF4) corresponde a las cualificaciones obtenidas a nivel de una educación postsecundaria no terciaria o de una cualificación profesional en el Espacio Europeo de Educación Superior (EEES). Las características y características del EQF 4 incluyen (Comisión Europea, 2023; CEDEFOP, 2023):

- **Conocimiento fáctico y teórico:** Las calificaciones EQF 4 requieren que los estudiantes posean una amplia gama de conocimientos fácticos y teóricos dentro de su campo de estudio u ocupación. Este nivel de conocimiento es necesario para comprender los principios, conceptos y procesos clave relevantes para su área de especialización.

- **Habilidades cognitivas:** En el EQF4, los alumnos deben desarrollar la capacidad de aplicar sus conocimientos en contextos prácticos, analizar situaciones y resolver problemas utilizando métodos y herramientas básicas. Esto incluye el pensamiento crítico, la resolución de problemas y las habilidades de toma de decisiones a un nivel más básico.
- **Habilidades prácticas:** Las Cualificaciones EQF4 implican el desarrollo de una serie de habilidades prácticas, lo que permite a los alumnos realizar tareas y completar actividades dentro del campo elegido. Estas habilidades pueden incluir habilidades técnicas, uso de herramientas y equipos relevantes, o la capacidad de ejecutar procesos y procedimientos específicos.
- **Autonomía y responsabilidad:** Se espera que los alumnos del EQF 4 demuestren un grado moderado de autonomía y responsabilidad en su trabajo. Esto implica asumir la responsabilidad de su propio aprendizaje y desarrollo, trabajar bajo supervisión u orientación y tomar decisiones basadas en pautas y procedimientos establecidos.
- **Comunicación y colaboración:** Las Cualificaciones del EQF4 requieren que los alumnos desarrollen habilidades efectivas de comunicación y colaboración, lo que les permite intercambiar información, argumentos o propuestas con otros. Esto incluye habilidades básicas de comunicación escrita, oral e interpersonal, así como la capacidad de trabajar eficazmente en equipos o grupos.
- **Aprendizaje a lo largo de toda la vida:** En el EQF 4, los alumnos deben desarrollar la capacidad de participar en el aprendizaje permanente, reflexionando sobre sus propias experiencias de aprendizaje e identificando áreas de desarrollo posterior. Esto incluye la capacidad de adaptarse a nuevas situaciones, tecnologías o entornos profesionales y de participar en el desarrollo profesional continuo.

Las competencias desarrolladas a lo largo del curso serán coherentes con los descriptores del EQF4, centrándose en proporcionar a los alumnos una amplia gama de conocimientos, habilidades cognitivas y prácticas, autonomía, responsabilidad, comunicación, colaboración y un compromiso con el aprendizaje permanente.

Las competencias para un curso de EQF4 sobre sesgo algorítmico que se proponen son:

- **Conocimiento fáctico y teórico:** 1.1. Comprender los conceptos básicos de algoritmos, datos e inteligencia artificial. 1.2. Reconocer el papel de los algoritmos en diversos sectores y las posibles consecuencias de una toma de decisiones sesgada.
- **Habilidades cognitivas:** 2.1. Identificar y analizar ejemplos sencillos de sesgo algorítmico y sus implicaciones en escenarios del mundo real. 2.2. Aplicar el pensamiento crítico para evaluar la equidad y las consideraciones éticas de los sistemas de IA en un contexto determinado.
- **Habilidades prácticas:** 3.1. Utilizar herramientas y métodos básicos para identificar posibles sesgos en los conjuntos de datos y los procesos algorítmicos. 3.2. Explorar técnicas sencillas para abordar y mitigar los sesgos en los sistemas de IA.
- **Autonomía y responsabilidad:** 4.1. Asumir la responsabilidad de su propio aprendizaje y desarrollo en el campo del sesgo algorítmico. 4.2. Considerar las implicaciones éticas de la toma de decisiones algorítmicas en su propio trabajo o área de estudio.
- **Comunicación y colaboración:** 5.1. Comunique y discuta de manera efectiva los problemas relacionados con el sesgo algorítmico con compañeros, instructores y otras partes interesadas. 5.2. Colaborar con otros para examinar y abordar casos de sesgo algorítmico en proyectos grupales o estudios de casos.
- **Aprendizaje a lo largo de toda la vida:** 6.1. Reflexionar sobre las experiencias personales de aprendizaje e identificar áreas de mayor desarrollo en el campo del sesgo algorítmico. 6.2. Demostrar el compromiso de mantenerse informado sobre los avances y las cuestiones emergentes en el campo de la ética de la IA y el sesgo algorítmico.

Al centrarse en estas competencias, un curso micro credencial EQF 4 basado en el sesgo algorítmico proporcionaría a los alumnos una base sólida en el campo, equipándolos con el conocimiento, las habilidades y la comprensión necesarios para sortear los desafíos éticos relacionados con los algoritmos sesgados y los sistemas de IA en sus futuras carreras o actividades de educación superior.

4.2.3. Matriz de competencias generales para el curso con micro credencial EQF4

La tabla a continuación presenta la matriz de competencias general para el curso EQF4:

RESULTADOS DE APRENDIZAJE		
Al finalizar la experiencia de formación, el alumno será capaz de:		
CONOCIMIENTO	**HABILIDADES**	**COMPETENCIAS**
1. Comprender los conceptos básicos de algoritmos, inteligencia artificial y aprendizaje automático y sus aplicaciones en diversos sectores. 2. Defina el sesgo algorítmico e identifique sus posibles fuentes, incluidos los datos, el diseño del modelo y la implementación. 3. Reconocer las consecuencias del sesgo algorítmico en los individuos, las comunidades y la sociedad en su conjunto. 4. Explique las consideraciones éticas que rodean el uso de los sistemas de IA, incluida la equidad, la transparencia y la rendición de cuentas. 5. Identificar ejemplos del mundo real y estudios de casos que ilustren el impacto del sesgo algorítmico y sus implicaciones éticas. 6. Apreciar la importancia de conjuntos de datos diversos y representativos en el desarrollo de sistemas de IA. 7. Examinar diversas estrategias y técnicas para mitigar y abordar el sesgo algorítmico en los sistemas de IA. 8. Discutir el papel de la legislación, la regulación y los estándares de la industria para abordar el sesgo algorítmico y promover el desarrollo ético de la IA. 9. Explore los desafíos y las oportunidades para equilibrar los beneficios de los sistemas de IA con los riesgos potenciales asociados con el sesgo algorítmico. 10. Reconocer la importancia de la colaboración interdisciplinaria en el desarrollo y la aplicación de sistemas de IA justos y éticos, incluida la participación de las partes interesadas de diversos orígenes y campos de especialización.	1. Analizar y evaluar los algoritmos en términos de sus posibles sesgos e implicaciones éticas. 2. Aplicar habilidades de pensamiento crítico para evaluar la justicia y la equidad de los sistemas de IA en situaciones del mundo real. 3. Utilizar técnicas estadísticas básicas para identificar y medir los sesgos en los conjuntos de datos. 4. Implementar estrategias para abordar y mitigar los sesgos en el desarrollo y la implementación de sistemas de IA. 5. Comunique de manera efectiva sobre el sesgo algorítmico, sus consecuencias y posibles soluciones a audiencias técnicas y no técnicas. 6. Colaborar eficazmente en diversos equipos para desarrollar y evaluar sistemas de IA teniendo en cuenta los aspectos éticos y de equidad. 7. Aplicar la legislación, los reglamentos y los estándares de la industria pertinentes al desarrollo y la implementación de sistemas de IA. 8. Adaptar y aplicar las mejores prácticas para el desarrollo ético de la IA en diversos sectores e industrias. 9. Demostrar habilidades de resolución de problemas para abordar casos reales de sesgo algorítmico y desarrollar posibles soluciones. 10. Actualizar continuamente los conocimientos y las habilidades en el campo del sesgo algorítmico y el desarrollo ético de la IA para mantenerse al día con las tendencias y tecnologías en evolución.	1. Valorar la importancia de las consideraciones éticas, la justicia y la equidad en el desarrollo y la implementación de sistemas de IA. 2. Demostrar un compromiso con el aprendizaje permanente y la mejora continua en la comprensión y aplicación de los principios éticos de la IA. 3. Adoptar una mentalidad responsable y responsable en el desarrollo, uso y evaluación de sistemas de IA. 4. Mostrar una Competencia proactiva para identificar y abordar posibles sesgos y preocupaciones éticas en las aplicaciones de IA. 5. Reconocer la importancia de la colaboración interdisciplinaria y la comunicación abierta para abordar el sesgo algorítmico y el desarrollo ético de la IA. 6. Apreciar la importancia de la transparencia, la rendición de cuentas y la participación de las partes interesadas en el desarrollo de sistemas de IA. 7. Respetar las diversas perspectivas y experiencias de las personas afectadas por los sistemas de IA, teniendo en cuenta su contribución en el diseño e implementación de soluciones justas e imparciales. 8. Fomentar un enfoque inclusivo para el desarrollo de la IA que tenga en cuenta las necesidades y preocupaciones de las diversas partes interesadas, incluidas las comunidades marginadas e infrarrepresentadas. 9. Reconocer las limitaciones de los sistemas de IA y la importancia de la evaluación y mejora continuas para minimizar los daños potenciales. 10. Defender el desarrollo ético de la IA y promover la concienciación sobre el sesgo algorítmico y sus consecuencias en contextos profesionales y sociales.

4.2.4. Matriz de competencias para cada unidad de competencia para el curso de micro credenciales EQF4

A continuación, se presentan las tablas por unidades de competencia:

CU1 – ¿Qué es el sesgo algorítmico?

CU2 – No maleficencia.

CU3 – Rendición de cuentas.

CU4 – Transparencia.

CU5 – Derechos humanos y equidad En la unidad de derechos humanos y equidad.

CU6 – Ética de la IA, un enfoque práctico.

CU1 - ¿Qué es el sesgo algorítmico? EQF4		
RESULTADOS DE APRENDIZAJE		
Al finalizar la experiencia de formación, el alumno será capaz de:		
CONOCIMIENTOS	**HABILIDADES**	**COMPETENCIAS**
Conocimiento teórico/fáctico en: CUC1.1. Definir el sesgo algorítmico y comprender sus conceptos fundamentales, incluidos los factores que contribuyen a los resultados sesgados en los sistemas de IA. CUC1.2. Identificar diferentes tipos de sesgos algorítmicos, como el sesgo basado en datos, el sesgo basado en modelos y el sesgo impulsado por humanos, y reconocer cómo pueden manifestarse en aplicaciones de IA. CUC1.3. Comprender las implicaciones reales del sesgo algorítmico en el mundo real en varios sectores, incluidas las posibles consecuencias sociales, éticas y legales asociadas con los sistemas de IA sesgados.	CUH1.1. Analizar los sistemas y aplicaciones de IA para detectar posibles sesgos, utilizando su comprensión de los conceptos fundamentales del sesgo algorítmico. CUH1.2. Categorizar diferentes tipos de sesgos algorítmicos en ejemplos del mundo real, empleando su conocimiento de los sesgos basados en datos, basados en modelos y en humanos. CUH1.3. Evaluar el impacto del sesgo algorítmico en diversos sectores y partes interesadas, considerando las posibles consecuencias sociales, éticas y legales relacionadas con los sistemas de IA sesgados.	CUCOMP1.1. Demostrar una mayor conciencia de las posibles consecuencias del sesgo algorítmico y adoptar una mentalidad crítica al interactuar con sistemas y aplicaciones de IA. CUCOMP1.2. Aceptar la importancia de la justicia y la equidad en los sistemas de IA, reconociendo la necesidad de perspectivas diversas y procesos de diseño inclusivos para minimizar los sesgos. CUCOMP1.3. Mostrar un compromiso con el aprendizaje continuo y mantenerse informado sobre los últimos desarrollos en la ética de la IA, el sesgo algorítmico y las prácticas responsables de IA, con el objetivo de contribuir positivamente al desarrollo y uso de sistemas de IA imparciales.

C1 DESCRIPCIÓN DE LOS RESULTADOS DEL CONOCIMIENTO

En el curso de nivel EQF4 sobre sesgo algorítmico, los estudiantes obtendrán conocimientos básicos en el área, centrándose en la comprensión de los conceptos básicos, la identificación de diferentes tipos de sesgos y el reconocimiento de las implicaciones en el mundo real del sesgo algorítmico en varios sectores. Los resultados de conocimiento de este curso incluyen:

CU1.1. Definición del sesgo algorítmico: Los estudiantes aprenderán a definir el sesgo algorítmico y comprender sus conceptos fundamentales. Explorarán los factores que contribuyen a los resultados sesgados en los sistemas de IA, como los métodos de recopilación de datos sesgados, los datos de entrenamiento sesgados y la toma de decisiones humanas. Este conocimiento fundamental ayudará a los estudiantes a reconocer cómo puede ocurrir el sesgo algorítmico y su impacto potencial en las aplicaciones de IA, por ejemplo, los sistemas de reconocimiento facial sesgados que identifican erróneamente ciertos grupos demográficos.

CU1.2. Identificación de tipos de sesgos algorítmicos: A los estudiantes se les presentarán diferentes tipos de sesgos algorítmicos, incluidos los sesgos basados en datos, los sesgos basados en modelos y los sesgos impulsados por humanos. A través de ejemplos y estudios de casos, aprenderán cómo estos sesgos pueden manifestarse en las aplicaciones de IA y comprenderán la importancia de abordarlos para garantizar sistemas de IA justos e imparciales. Por ejemplo, examinarán cómo el sesgo basado en datos puede resultar de datos de capacitación poco representativos, lo que lleva a predicciones sesgadas en áreas como la calificación crediticia o la selección de solicitantes de empleo.

CU1.3. Implicaciones del sesgo algorítmico en el mundo real: Los estudiantes explorarán las implicaciones en el mundo real del sesgo algorítmico en varios sectores, como la atención médica, las finanzas y la justicia penal. Examinarán las posibles consecuencias sociales, éticas y legales asociadas con los sistemas de IA sesgados, apreciando la necesidad de minimizar el sesgo algorítmico para evitar resultados negativos y promover la justicia y la equidad en las aplicaciones de IA. Algunos ejemplos podrían ser la forma en que los sistemas de IA sesgados en la atención sanitaria pueden dar lugar a diagnósticos erróneos o tratamientos inadecuados para determinados grupos demográficos, o cómo el sesgo algorítmico en los sistemas de justicia penal puede dar lugar a sentencias injustas o a la elaboración de perfiles de personas.

CU2 - EQF4 No maleficencia		
RESULTADOS DE APRENDIZAJE		
Al finalizar la experiencia de formación, el alumno será capaz de:		
CONOCIMIENTOS	**HABILIDADES**	**COMPETENCIAS**
Conocimiento teórico/fáctico en: CUC2.1. Introducir el principio de no maleficencia: Enseñar a los estudiantes el concepto básico de no maleficencia, haciendo hincapié en la importancia de evitar daños al crear y utilizar sistemas de IA, y cómo esta idea contribuye al desarrollo responsable de la IA. CUC2.2. Examinar los posibles daños de la IA sesgada: Guíe a los estudiantes para que reconozcan las diversas formas en que los sistemas de IA sesgados pueden causar daño, como la discriminación o la invasión de la privacidad, y use ejemplos del mundo real para ilustrar la importancia de abordar el sesgo algorítmico. CUC2.3. Explorar estrategias para hacer que los sistemas de IA sean menos dañinos: Educar a los estudiantes sobre estrategias simples que pueden hacer que los sistemas de IA sean menos dañinos, incluida la promoción de la equidad, la responsabilidad y la transparencia en el desarrollo de la IA, y el fomento de la colaboración con expertos de diversos campos.	CUH2.1. Explique el concepto de no maleficencia en el contexto del desarrollo de la IA. CUH2.2. Identificar los daños potenciales en los sistemas de IA y reconocer la importancia de la no maleficencia en el desarrollo responsable de la IA. CUH2.3. Analizar los diferentes tipos de daños que pueden derivarse de sistemas de IA sesgados. CUH 2.4. Evalúe ejemplos del mundo real de sistemas de IA sesgados y sus consecuencias. CUH2.5. Identificar y aplicar estrategias para reducir el daño en los sistemas de IA, como la promoción de la equidad, la responsabilidad y la transparencia. CUH2.6. Colaborar eficazmente con expertos de diversos campos para abordar y mitigar los daños causados por los sistemas de IA sesgados.	CUCOMP2.1. Cultivar el sentido de la responsabilidad hacia la aplicación del principio de no maleficencia en el desarrollo de la IA. CUCOMP2.2. Apreciar la importancia de las consideraciones éticas en los sistemas de IA, especialmente para evitar daños potenciales. CUCOMP2.3. Desarrollar empatía hacia las personas y comunidades afectadas por los daños causados por los sistemas de IA sesgados. CUCOMP2.4. Fomentar el compromiso de abordar y mitigar el sesgo algorítmico en el desarrollo de la IA para reducir los daños potenciales. CUCOMP2.5. Adoptar un enfoque proactivo para implementar estrategias que reduzcan el daño en los sistemas de IA. CUCOMP2.6. Valorar la importancia de la colaboración interdisciplinaria y las diversas perspectivas para abordar y mitigar los daños causados por los sistemas de IA sesgados.

DESCRIPCIÓN DE LOS RESULTADOS DEL CONOCIMIENTO

Los estudiantes obtendrán conocimientos básicos del principio de no meleficiencia en IA, centrándose en la comprensión de los conceptos básicos, identificando las responsabilidades de los desarrolladores y usuarios de IA para garantizar sistemas de IA éticos con un daño mínimo, y reconociendo las implicaciones en el mundo real apreciando la adopción e implementación de mecanismos que promuevan la rendición de cuentas en los sistemas de IA.

Los resultados de conocimiento de este curso incluyen:

CU2.1. Principio de no maleficencia: los estudiantes el concepto básico de no maleficencia, haciendo hincapié en la importancia de evitar daños al crear y utilizar sistemas de IA, y cómo esta idea contribuye al desarrollo responsable de la IA.

CU2.2. Posibles daños de la IA sesgada: Los estudiantes reconocerán las diversas formas en que los sistemas de IA sesgados pueden causar daño, como la discriminación o la invasión de la privacidad, y utilizarán ejemplos del mundo real para ilustrar la importancia de abordar el sesgo algorítmico.

CU2.3. Estrategias para hacer que los sistemas de IA sean menos dañinos: Los estudiantes se familiarizarán con estrategias simples que pueden hacer que los sistemas de IA sean menos dañinos, incluida la promoción de la equidad, la responsabilidad y la transparencia en el desarrollo de la IA, y el fomento de la colaboración con expertos de diversos campos.

CU3 – Rendición de cuentas EQF4		
RESULTADOS DE APRENDIZAJE		
Al finalizar la experiencia de formación, el alumno será capaz de:		
CONOCIMIENTO	**HABILIDADES**	**COMPETENCIAS**
Conocimiento teórico/fáctico en: CUC3.1. Concepto básico de rendición de cuentas en IA: Los estudiantes describirán el concepto fundamental de rendición de cuentas en IA, haciendo hincapié en la responsabilidad de los desarrolladores y usuarios de IA para garantizar que los sistemas de IA funcionen de forma ética y con un daño mínimo. CUC3.2. Papel de la rendición de cuentas en el tratamiento del sesgo algorítmico: Los estudiantes identificarán cómo la rendición de cuentas desempeña un papel crucial en la prevención y mitigación de los efectos del sesgo algorítmico, y cómo asumir la responsabilidad de los sistemas de IA puede conducir a una mejor toma de decisiones y resultados más equitativos. CUC3.3. Mecanismos para garantizar la rendición de cuentas en los sistemas de IA: Los estudiantes se familiarizarán con mecanismos simples que pueden ayudar a garantizar la rendición de cuentas en los sistemas de IA, como las pautas, la etiqueta en la red, las políticas de uso aceptable (AUP) y las regulaciones para los desarrolladores y usuarios de IA	CUH3.1. Explique el concepto de rendición de cuentas en el contexto del desarrollo y uso de la IA. CUH3.2. Identificar las responsabilidades de los desarrolladores y usuarios de IA para garantizar sistemas de IA éticos con un daño mínimo. CUH3.3. Analizar la relación entre la rendición de cuentas y el sesgo algorítmico y comprender su importancia en el desarrollo de la IA. CUH3.4. Utilizar el concepto de rendición de cuentas en escenarios del mundo real para mejorar la toma de decisiones y lograr resultados más equitativos. CUH3.5. Identificar y evaluar mecanismos para garantizar la rendición de cuentas en los sistemas de IA, como directrices, etiqueta en la red, políticas de uso aceptable (PUA) y regulaciones para desarrolladores y usuarios de IA. CUH3.6. Seleccionar y adaptar estos mecanismos en el desarrollo de la IA y utilizarlos para promover la rendición de cuentas y el comportamiento ético.	CUCOMP3.1. Cultivar un sentido de responsabilidad para el desarrollo y uso ético de la IA, reconociendo la importancia de la rendición de cuentas. CUCOMP3.2. Apreciar la importancia de la rendición de cuentas para minimizar el daño y promover un comportamiento ético en los sistemas de IA. CUCOMP3.3. Valorar la importancia de la rendición de cuentas para abordar el sesgo algorítmico y sus posibles consecuencias. CUCOMP3.4. Fomentar el compromiso de asumir la responsabilidad de los sistemas de IA para lograr una mejor toma de decisiones y resultados más equitativos. CUCOMP3.5. Adoptar e implementar mecanismos que promuevan la rendición de cuentas en los sistemas de IA. CUCOMP3.6. Apreciar el papel de las directrices, la etiqueta en la red, las políticas de uso aceptable (PUA) y las regulaciones para fomentar una cultura de responsabilidad y comportamiento ético en el desarrollo y uso de la IA.

DESCRIPCIÓN DE LOS RESULTADOS DEL CONOCIMIENTO

Los resultados de conocimiento de este curso incluyen:

Los estudiantes obtendrán conocimientos básicos en rendición de cuentas en IA, centrándose en la comprensión de los conceptos básicos, identificando las responsabilidades de los desarrolladores y usuarios de IA para garantizar sistemas de IA éticos con un daño mínimo, y reconociendo las implicaciones en el mundo real apreciando la adopción e implementación de mecanismos que promuevan la rendición de cuentas en los sistemas de IA.

Los resultados de conocimiento de este curso incluyen:

CU3.1. Concepto básico de rendición de cuentas en IA: Los estudiantes aprenderán el concepto fundamental de rendición de cuentas en IA. La rendición de cuentas en IA se relaciona con la expectativa de que los diseñadores, desarrolladores e implementadores cumplan con los estándares y la legislación para garantizar el correcto funcionamiento de las IA durante su ciclo de vida (Fjeld et al. 2020). Los estudiantes identificarán las responsabilidades de los desarrolladores y usuarios de IA para garantizar sistemas de IA éticos con un daño mínimo. Cultivarán un sentido de responsabilidad por el desarrollo y el uso éticos de la IA, reconociendo la importancia de la rendición de cuentas y apreciando su importancia para minimizar el daño y promover el comportamiento ético en los sistemas de IA.

CU3.2. Papel de la rendición de cuentas en el tratamiento del sesgo algorítmico: Los estudiantes identificarán las razones por las que la rendición de cuentas desempeña un papel crucial en la prevención y mitigación de los efectos del sesgo algorítmico. Aprenderán a analizar la relación entre la rendición de cuentas y el sesgo algorítmico y valorarán la importancia de asumir la responsabilidad de los sistemas de IA para una mejor toma de decisiones y resultados más equitativos.

CU3.3. Mecanismos para garantizar la rendición de cuentas en los sistemas de IA. Se familiarizarán con mecanismos sencillos que pueden ayudar a garantizar la rendición de cuentas en los sistemas de IA. Los estudiantes apreciarán la importancia de comunicarse de manera efectiva con las partes interesadas (por ejemplo, incluido el suministro de información sobre la lógica y la lógica del sistema de IA, como las entradas, salidas y procesos utilizados para tomar decisiones o recomendaciones). Ejemplo de mecanismos: directrices, etiqueta en la red, políticas de uso aceptable (PUA) y regulaciones para desarrolladores y usuarios de IA.

CU4 – Transparencia EQF4		
RESULTADOS DE APRENDIZAJE		
Al finalizar la experiencia de formación, el alumno será capaz de:		
CONOCIMIENTO	**HABILIDADES**	**COMPETENCIAS**
Conocimiento teórico/fáctico en: CUC4.1. Importancia de la transparencia en los sistemas de IA: Los estudiantes aprenderán el concepto fundamental de transparencia en la IA y su relevancia para garantizar que los sistemas de IA sean comprensibles, explicables y accesibles para las partes interesadas. CUC4.2. Relación entre la transparencia y el sesgo algorítmico: Los estudiantes comprenderán la conexión entre la transparencia y el 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. CUC4.3. Estrategias para promover la transparencia en los sistemas de IA: Los estudiantes se familiarizarán con estrategias sencillas para promover la transparencia en los sistemas de IA, como el uso de modelos interpretables, el suministro de documentación clara y la comunicación de los procesos de toma de decisiones de las aplicaciones de IA.	CUH4.1. Explique el concepto de transparencia y su importancia en el contexto de los sistemas de IA. CUH4.2. Identificar los beneficios de los sistemas de IA transparentes para las partes interesadas, incluida la comprensión, la explicabilidad y la accesibilidad. CUH4.3. Analizar la relación entre la transparencia y el sesgo algorítmico en los sistemas de IA. CUH4.4. Analizar los peligros de la opacidad en los sistemas de IA e identificar cómo el concepto de transparencia puede ayudar a prevenir y mitigar los resultados sesgados en los sistemas de IA. CUH4.5. Identificar y evaluar estrategias para promover la transparencia en los sistemas de IA, incluidos modelos interpretables, documentación clara y comunicación efectiva de los procesos de toma de decisiones. CUH4.6. Aplicar estas estrategias en el desarrollo de la IA y utilizarlas para mejorar la transparencia y mitigar el sesgo algorítmico.	CUCOMP4.1. Valorar la importancia de la transparencia en los sistemas de IA para generar confianza y permitir la comprensión de las partes interesadas. CUCOMP4.2. Apreciar el papel de la transparencia en el fomento del desarrollo y uso ético de la IA. CUCOMP4.3. Reconocer los peligros de la opacidad en los sistemas de IA y la importancia de la transparencia para abordar y mitigar el sesgo algorítmico. CUCOMP4.4. Cultivar un compromiso para mejorar la transparencia en los sistemas de IA para minimizar los resultados sesgados. CUCOMP4.5. Adoptar e implementar estrategias que promuevan la transparencia en los sistemas de IA. CUCOMP4.6. Apreciar el papel de los modelos interpretables, la documentación clara y la comunicación efectiva para fomentar una cultura de transparencia y mitigar el sesgo algorítmico.

DESCRIPCIÓN DE LOS RESULTADOS DEL CONOCIMIENTO

Los estudiantes 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 la importancia de los modelos interpretables, la documentación clara y la comunicación eficaz para fomentar una cultura de transparencia, mitigando el sesgo algorítmico.

Los resultados de conocimiento de este curso incluyen:

CU4.1. Importancia de la transparencia en los sistemas de IA: Los estudiantes identificarán el concepto fundamental de transparencia en la IA y su relevancia para garantizar que los sistemas de IA sean comprensibles, explicables y accesibles para las partes interesadas. Identificarán los beneficios y valorarán la importancia de los sistemas de IA transparentes para generar confianza y permitir la comprensión de las partes interesadas. Ejemplo: Un modelo de IA diseñado para detectar el cáncer, incluso si solo tiene un 1% de errores, podría poner en peligro una vida. En casos como estos, la IA y los humanos necesitan trabajar juntos, y la tarea se vuelve mucho más fácil Cuando el modelo de IA puede explicar cómo llegó a una determinada decisión. La transparencia hace que la IA sea un jugador de equipo.

CU4.2. Relación entre la transparencia y el sesgo algorítmico: Los estudiantes encontrarán la conexión entre la transparencia y el 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. Reconocerán la importancia de la transparencia para abordar y mitigar el sesgo algorítmico. Ejemplo: Muy a menudo, los algoritmos de IA son opacos en el sentido de que tales explicaciones no están disponibles para todas las partes interesadas. Esta opacidad puede tener diferentes orígenes. A veces, las instituciones o corporaciones no se comunican Cuando confían en los sistemas de IA o en el funcionamiento de estos sistemas.

CU4.3. Estrategias para promover la transparencia en los sistemas de IA. Los estudiantes se familiarizarán con estrategias sencillas para promover la transparencia en los sistemas de IA, como el uso de modelos interpretables, el suministro de documentación clara y la comunicación de los procesos de toma de decisiones de las aplicaciones de IA. Apreciarán lo importantes que son estas estrategias para promover una cultura de transparencia y mitigar el sesgo algorítmico. Ejemplo: La IA puede afectar a diversas partes interesadas, como usuarios, clientes, empleados, directivos, reguladores o a la sociedad. Para garantizar la transparencia y la rendición de cuentas, es necesario involucrar y capacitar a las partes interesadas de la IA a lo largo del ciclo de vida de los sistemas de información.

CU5 - Derechos humanos y equidad EQF4		
RESULTADOS DE APRENDIZAJE		
Al finalizar la experiencia de formación, el alumno será capaz de:		
CONOCIMIENTO	**HABILIDADES**	**COMPETENCIAS**
Conocimiento teórico/fáctico en: CUC5.1. Relevancia de los derechos humanos y la equidad en los sistemas de IA: Los estudiantes aprenderán la importancia de los derechos humanos y la equidad en el desarrollo y la implementación de sistemas de IA, identificando cómo contribuyen a obtener resultados equitativos y prevenir daños a las personas y las comunidades. CUC5.2. Intersección del sesgo algorítmico y los derechos humanos: Los estudiantes reconocerán la conexión entre el sesgo algorítmico y los derechos humanos, identificando cómo los sistemas de IA sesgados pueden afectar de manera desproporcionada a las poblaciones vulnerables y perpetuar las desigualdades existentes. CUC5.3. Principios de equidad en los sistemas de IA: Los estudiantes comprenderán los principios fundamentales de la equidad en los sistemas de IA, como la igualdad de oportunidades, la no discriminación, la justicia procesal y la equidad. y la justicia y reconocer su papel en la promoción de resultados equitativos en nombre de las generaciones futuras.	CUH5.1. Explique la importancia de los derechos humanos y la equidad en los sistemas de IA y su papel en la promoción de resultados equitativos y la prevención de daños. CUH5.2. Aplicar los principios de derechos humanos y equidad en el contexto del desarrollo y despliegue de la IA. CUH5.3. Analizar la relación entre el sesgo algorítmico y los derechos humanos, reconociendo el impacto potencial en las poblaciones vulnerables y la perpetuación de las desigualdades. CUH5.4. Identificar ejemplos reales de sistemas de IA sesgados que afecten a los derechos humanos y diseñar estrategias para abordar estos problemas. CUH5.5. Definir y distinguir entre los diferentes principios de equidad en los sistemas de IA, incluida la igualdad de oportunidades, la no discriminación, la justicia procesal y la equidad. y la justicia. CUH5.6. Aplicar los principios de equidad en el desarrollo de la IA y utilizarlos para promover resultados equitativos y mitigar el sesgo algorítmico en nombre de las generaciones futuras.	CUCOMP5.1. Valorar la importancia de los derechos humanos y la justicia en los sistemas de IA para promover la equidad y prevenir daños. CUCOMP5.2. Asumir el compromiso de defender los derechos humanos y la equidad en el desarrollo y la implementación de la IA. CUCOMP5.3. Apreciar la importancia de la intersección entre el sesgo algorítmico y los derechos humanos en la configuración del impacto de los sistemas de IA en las personas y las comunidades. CUCOMP5.4. Cultivar un sentido de responsabilidad para abordar y mitigar los efectos de los sistemas de IA sesgados en los derechos humanos y las poblaciones vulnerables. CUCOMP5.5. Reconocer el valor de los diferentes principios de equidad, equidad y justicia en los sistemas de IA para mitigar el sesgo algorítmico. CUCOMP5.6 Fomentar el compromiso de incorporar los principios de equidad en el desarrollo y el uso de la IA para lograr resultados más equitativos que respeten los intereses de las generaciones futuras. .

DESCRIPCIÓN DE LOS RESULTADOS DEL CONOCIMIENTO

Los estudiantes adquirirán conocimientos sobre la importancia de los derechos humanos y la equidad en los sistemas de IA, centrándose en la comprensión de los conceptos básicos, la intersección del sesgo algorítmico y los derechos humanos y los principios de equidad en los sistemas de IA, reconociendo las implicaciones en el mundo real apreciando el valor de los diferentes principios de equidad, equidad y justicia en los sistemas de IA para

mitigar el sesgo algorítmico y lograr resultados más equitativos respetando el interés de las generaciones futuras.

Los resultados de conocimiento de este curso incluyen:

5.1. Relevancia de los derechos humanos y la equidad en los sistemas de IA: Los estudiantes aprenderán la importancia de los derechos humanos y la equidad en el desarrollo y la implementación de sistemas de IA para promover resultados equitativos y prevenir daños. Los estudiantes reconocerán que la IA ofrece beneficios de gran alcance para el desarrollo humano, pero también presenta riesgos (como, entre otros, una mayor división entre los privilegiados y los no privilegiados; la erosión de las libertades individuales a través de la vigilancia; y la sustitución del pensamiento y el juicio independientes por el control automatizado).

5.2. Comprender la intersección de los derechos humanos y la equidad algorítmica: En esta unidad fundamental, los estudiantes serán introducidos a la intersección crítica de los derechos humanos y la equidad algorítmica. Basándose en las teorías de Latour (2005), quien enfatizó las implicaciones morales y éticas de la tecnología, los estudiantes explorarán el papel y el impacto de los procesos algorítmicos en los derechos humanos. Los estudiantes aprenderán la evolución histórica de los derechos humanos y cómo se cruza con la aparición de las tecnologías algorítmicas. Marcos teóricos: Introducción a los marcos teóricos que analizan la interacción entre los derechos humanos y la equidad algorítmica. Ejemplo: análisis de casos de estudio donde los procesos algorítmicos han estado en conflicto con los derechos humanos y los debates en torno a ellos.

5.3. Principios de equidad en los sistemas de IA: los estudiantes se familiarizarán con los principios fundamentales de equidad en los sistemas de IA, como la igualdad de oportunidades, la no discriminación, la equidad procesal, la equidad y la justicia, y reconocerán su valor en la promoción de resultados equitativos en nombre de las generaciones futuras.

CU6 - Ética de la IA, un enfoque práctico EQF4		
RESULTADOS DE APRENDIZAJE		
Al finalizar la experiencia de formación, el alumno será capaz de:		
CONOCIMIENTO	**HABILIDADES**	**COMPETENCIAS**
Conocimiento teórico/fáctico en: CUC6.1. Importancia de la ética de la IA en la práctica: Los estudiantes aprenderán la importancia de incorporar consideraciones éticas en el desarrollo y despliegue de sistemas de IA, comprendiendo cómo la ética puede mitigar las consecuencias negativas del sesgo algorítmico y fomentar la confianza en las aplicaciones de IA. CUC6.2. Marcos éticos y directrices para los sistemas de IA: Los estudiantes se familiarizarán con diversos marcos éticos y directrices aplicables a los sistemas de IA, como los principios éticos de la IA, las prácticas responsables de la IA y las regulaciones específicas de la industria, y comprenderán su papel en la orientación del desarrollo ético de la IA. CUC6.3. Aplicaciones de la ética de la IA en el mundo real: Los estudiantes explorarán ejemplos prácticos y estudios de casos de sistemas de IA en diversos dominios, analizando los desafíos y consideraciones éticas en cada escenario e identificando las mejores prácticas para garantizar el desarrollo y la implementación éticos de la IA.	CUH6.1. Explique la importancia de la ética en el contexto del desarrollo y el despliegue de la IA, incluido el papel de la ética en la mitigación del sesgo algorítmico y el fomento de la confianza. CUH6.2. Aplicar consideraciones éticas al desarrollo y uso de la IA para garantizar sistemas de IA responsables y fiables. CUH6.3. Identificar y analizar diversos marcos éticos y directrices para los sistemas de IA, incluidos los principios éticos de la IA, las prácticas responsables de IA y las regulaciones específicas de la industria. CUH 6.4. Aplicar marcos y directrices éticos en el desarrollo de la IA y utilizarlos para guiar las prácticas de IA responsables y éticas. CUH6.5. Analizar ejemplos del mundo real y estudios de casos de sistemas de IA para identificar desafíos y consideraciones éticas en diversos ámbitos. CUH6.6. Aplicar las mejores prácticas y directrices éticas a escenarios del mundo real que impliquen sistemas de IA para garantizar un desarrollo y despliegue éticos.	CUCOMP6.1. Valorar la importancia de la ética en el desarrollo y despliegue de la IA para mitigar el sesgo algorítmico y fomentar la confianza. CUCOMP6.2. Asumir el compromiso de incorporar consideraciones éticas en los sistemas de IA para garantizar aplicaciones de IA responsables y confiables. CUCOMP6.3. Apreciar el papel de los marcos éticos y las directrices para guiar el desarrollo y el uso éticos de la IA. CUCOMP6.4. Cultivar el compromiso de aplicar marcos y directrices éticas para garantizar prácticas de IA responsables y éticas. CUCOMP6.5. Reconocer la importancia de comprender y abordar los desafíos éticos en las aplicaciones de IA del mundo real. CUCOMP6.6. Fomentar el compromiso de adoptar las mejores prácticas y directrices éticas en escenarios del mundo real para garantizar el desarrollo y la implementación éticos de la IA.

DESCRIPCIÓN DE LOS RESULTADOS DEL CONOCIMIENTO

Objetivo: Esta unidad está diseñada para que los estudiantes de nivel EQF4 profundicen en los aspectos prácticos de la ética de la IA, incorporando principios de desarrollo y uso responsable de la IA en escenarios del mundo real. Se anima a los estudiantes a ser proactivos en la aplicación de directrices éticas en entornos de IA, aprovechando los conocimientos teóricos y traduciéndolos en información práctica.

6.1. Comprender las implicaciones éticas de la IA

En este módulo inicial, los estudiantes comenzarán con una base sólida en las diversas dimensiones éticas que abarca la IA. Siguiendo las ideas proporcionadas por Floridi et al. (2018) sobre las preocupaciones éticas que rodean a la IA, se presentará a los estudiantes las posibles implicaciones de la IA en la sociedad, los individuos y la comunidad global.

Explicación:

Fundamentación ética: Analizar y comprender las diferentes teorías éticas y sus implicaciones en la IA.

Implicaciones morales: Estudio detallado de las implicaciones morales que se derivan de la IA y los sistemas automatizados de toma de decisiones.

Ejemplo: Análisis de estudios de casos del mundo real que retratan los dilemas éticos a los que se enfrenta la industria de la IA.

6.2. Identificación y mitigación de prácticas poco éticas en IA

Esta unidad aborda la necesidad de reconocer y mitigar posibles prácticas poco éticas en IA. A partir de las reflexiones de Bostrom (2014), quien exploró el futuro de la IA y su alineación con los valores humanos, los estudiantes aprenderán a identificar y prevenir prácticas poco éticas en el desarrollo y la implementación de la IA.

Explicación:

Identificación de prácticas poco éticas: Un estudio de varios indicadores que significan prácticas poco éticas en la IA.

Estrategias de mitigación: Desarrollar estrategias para mitigar los posibles resultados poco éticos derivados de las aplicaciones de IA.

Ejemplo: Actividades de juego de roles en las que los estudiantes identifican y abordan prácticas poco éticas en entornos de IA simulados.

6.3. Aplicación práctica de las directrices éticas en la IA

Se anima a los estudiantes de este módulo a traducir la comprensión teórica en acción práctica. Con las ideas de los trabajos de Ryan y Stahl (2020), centrados en la formulación de directrices éticas, los estudiantes participarán

en actividades que fomenten la aplicación práctica de estas directrices en escenarios de IA.

Explicación:

Desarrollo de directrices éticas: elaboración de directrices éticas que puedan aplicarse en proyectos de IA del mundo real.

Implementación en escenarios reales: Talleres y actividades enfocadas a la implementación de las directrices desarrolladas en proyectos reales de IA.

Ejemplo: Los estudiantes participarán en un proyecto en el que desarrollarán y aplicarán pautas éticas en un proyecto de IA del mundo real.

6.4. Fomento del desarrollo y el despliegue responsables de la IA

La última unidad se centrará en fomentar un entorno que fomente el desarrollo y el despliegue responsables de la IA. Con base en el marco propuesto por Jobin et al. (2019), que discutió las perspectivas globales sobre la ética de la IA, los estudiantes explorarán métodos para fomentar la colaboración global y la creación de IA responsable.

Explicación:

Colaboración global: Comprender la importancia de la colaboración global para fomentar una IA responsable.

Compromiso de la comunidad: Fomentar el compromiso y la participación de la comunidad en el desarrollo ético de la IA.

Ejemplo: Los estudiantes desarrollarán estrategias y planes para fomentar la participación de la comunidad y la colaboración global en la ética de la IA.

4.3. Resultados de aprendizaje de CHARLIE WP2 para adultos y jóvenes EQF2 (SERIOUS GAME)

"*Desentrañando el sesgo algorítmico*" es un juego serio educativo e interactivo que introduce a los participantes de nivel EQF2 en el importante tema del sesgo algorítmico. A lo largo del juego, se exploran los conceptos fundamentales de los algoritmos, sus posibles sesgos y las implicaciones que estos sesgos pueden tener en varios aspectos de la sociedad.

El objetivo principal de "Desentrañando el sesgo algorítmico" (UAB) es proporcionar un entorno de aprendizaje atractivo en el que los participantes puedan investigar el tema a través de una serie de actividades y escenarios interactivos. Al presentar el contenido en un formato accesible y ameno, se espera que se fomente en los participantes una comprensión y conciencia más profundas de la importancia del sesgo algorítmico.

Concretamente, los resultados esperados de este WP están dirigidos a:

- Introducir a los participantes en el concepto de sesgo algorítmico y su relevancia en el mundo actual, destacando la importancia de abordar este tema en diversos sectores.
- Exponga a los participantes a ejemplos de la vida real de sesgo algorítmico, ilustrando las posibles consecuencias e implicaciones éticas asociadas con los algoritmos sesgados.
- Fomentar la capacidad de pensamiento crítico entre los participantes, permitiéndoles analizar y evaluar algoritmos y sus posibles sesgos en diferentes contextos.
- Desarrollar la comprensión de los participantes sobre varios tipos de sesgos algorítmicos, incluidos los sesgos basados en datos, los sesgos impulsados por modelos y los sesgos impulsados por humanos.
- Animar a los participantes a explorar el papel de la equidad, la transparencia y la rendición de cuentas en el desarrollo y la implementación de sistemas de IA.
- Mejorar la concienciación de los participantes sobre los principios éticos y las directrices que rigen el desarrollo de la IA, como la no maleficencia, los derechos humanos y la privacidad.
- Desafíe a los participantes a aplicar sus nuevos conocimientos para identificar posibles sesgos en aplicaciones de IA simuladas y proponer soluciones para mitigar estos sesgos.

- Promover la colaboración y la comunicación entre los participantes mientras trabajan juntos para resolver problemas relacionados con el sesgo algorítmico y compartir sus ideas.
- Involucrar a los participantes en actividades reflexivas, animándolos a considerar sus propias perspectivas y sesgos y cómo podrían afectar el uso de los sistemas de IA.
- Inspirar a los participantes para que sean proactivos en la defensa del desarrollo y la implementación éticos de la IA en sus comunidades y entornos profesionales, fomentando el sentido de responsabilidad y el compromiso cívico.

Al finalizar el juego, los participantes en el nivel EQF2 habrán adquirido una comprensión básica del sesgo algorítmico y sus posibles consecuencias. Desarrollarán habilidades iniciales de resolución de problemas que les permitan reconocer sesgos en las aplicaciones de IA. Además, los participantes habrán mejorado sus habilidades de comunicación y trabajo en equipo, al tiempo que fomentarán el sentido de la responsabilidad y la conciencia sobre las consideraciones éticas en el uso de la IA y la tecnología.

4.3.1. Contenido del juego "Desentrañando el sesgo algorítmico"

La estructura principal del contenido del juego para los participantes del nivel EQF2 se puede organizar en cinco módulos:

1. Introducción al sesgo algorítmico:
Conceptos básicos de algoritmos e IA
Introducción al sesgo y la equidad en los sistemas de IA
Ejemplos reales de sesgo algorítmico

2. Tipos de sesgos algorítmicos:
Sesgo basado en datos
Sesgo basado en modelos
Sesgo impulsado por el ser humano
Ejemplos y escenarios para cada tipo de sesgo

3. Identificación y tratamiento del sesgo algorítmico:
Estrategias para detectar sesgos en aplicaciones de IA
Técnicas básicas para mitigar los sesgos
Papel de las personas y las organizaciones en la lucha contra el sesgo algorítmico

4. Consideraciones éticas en la IA:
Importancia de la ética en el desarrollo y uso de la IA
Introducción a principios éticos como la no maleficencia, la rendición de cuentas y la transparencia
Conexión entre los derechos humanos, la equidad y la ética de la IA

5. Actividades de juego y desafíos:
Escenarios interactivos en los que los jugadores identifican y abordan los sesgos en las aplicaciones de IA
Tareas de resolución de problemas basadas en equipos relacionadas con el desarrollo ético de la IA
Sesiones de reflexión y discusión para reforzar el aprendizaje y promover la conciencia sobre el sesgo algorítmico
A lo largo del juego, los participantes participarán en actividades colaborativas y desafíos diseñados para reforzar su comprensión de los conceptos mientras fomentan el trabajo en equipo, la comunicación y las habilidades de resolución de problemas.

4.3.2. Descripción del grupo destinatario, características y características del MEC2

El Nivel 2 del Marco Europeo de Cualificaciones representa un nivel básico de competencia en diversas materias o capacidades (Comisión Europea, 2023). Los participantes en el nivel EQF2 suelen encontrarse en las primeras etapas de su viaje de aprendizaje o cursando una formación profesional básica. Las principales características y características del EQF2 incluyen:

- **Conocimientos básicos:** Los participantes del EQF2 poseen una comprensión fundamental de los conceptos, principios e ideas clave en el área temática elegida, sin profundizar en teorías complejas o temas avanzados.
- **Habilidades prácticas:** Los participantes en el nivel EQF2 pueden aplicar sus conocimientos básicos para realizar tareas y actividades sencillas relacionadas con su campo, utilizando herramientas y técnicas básicas bajo supervisión o con orientación.
- **Habilidades cognitivas:** Los alumnos del EQF2 pueden utilizar habilidades cognitivas básicas, como la resolución de problemas y la toma de decisiones, en situaciones familiares y sencillas. Pueden seguir instrucciones y procedimientos sencillos para completar las tareas.

- **Comunicación:** Los participantes del EQF 2 son capaces de comunicar información e ideas básicas de forma clara y concisa, tanto verbalmente como por escrito, en contextos familiares.
- **Colaboración y trabajo en equipo**: En el nivel EQF2, los participantes pueden trabajar eficazmente con otros en un entorno de equipo o grupo, demostrando cooperación y comprensión de la dinámica interpersonal básica.
- **Responsabilidad y autonomía:** Los alumnos del EQF 2 pueden asumir la responsabilidad de sus acciones y decisiones dentro del ámbito de sus limitados conocimientos y habilidades. Pueden trabajar bajo supervisión, siguiendo instrucciones y buscando ayuda Cuando sea necesario.
- **Adaptabilidad y flexibilidad:** Los participantes en el nivel EQF2 pueden adaptarse a cambios simples en su entorno de aprendizaje o trabajo y pueden manejar desafíos o contratiempos básicos con orientación y apoyo.
- **Autoconciencia y autorreflexión:** Los alumnos del EQF2 pueden reflexionar sobre sus propias experiencias de aprendizaje e identificar áreas de mejora, demostrando un nivel básico de autoconciencia y un compromiso con el crecimiento personal.
- **Conciencia ética:** En el nivel EQF2, los participantes pueden reconocer los principios éticos básicos y las directrices relevantes para su área temática y demostrar una comprensión de su importancia en situaciones cotidianas.
- **Aprendizaje a lo largo de toda la vida:** Se anima a los alumnos del EQF 2 a que desarrollen un interés por el campo elegido y a que cultiven una mentalidad de aprendizaje permanente, fomentando la curiosidad y la voluntad de continuar su educación y el desarrollo de habilidades más allá del nivel del EQF 2.

4.3.3. Matriz de competencias generales para el juego EQF2 «Desentrañando el sesgo algorítmico»

La presente tabla presenta los resultados de aprendizaje del juego:

RESULTADOS DE APRENDIZAJE		
Al finalizar el juego, el participante podrá:		
CONOCIMIENTOS	**HABILIDADES**	**COMPETENCIAS**
1. Comprender el concepto básico de los algoritmos y cómo se utilizan en la vida cotidiana, incluyendo ejemplos sencillos de sus aplicaciones. 2. Reconocer la importancia de la equidad y la toma de decisiones imparcial en el contexto de los procesos algorítmicos. 3. Comprender la noción básica de sesgo algorítmico, incluida la comprensión de cómo se pueden introducir los sesgos en los sistemas de IA y sus posibles consecuencias. 4. Identificar ejemplos sencillos de sesgo algorítmico en situaciones del mundo real, y la importancia de abordar estos sesgos para garantizar resultados equitativos. 5. Comprender el papel de los datos en el funcionamiento de los sistemas de IA y cómo los datos sesgados o incompletos pueden dar lugar a decisiones sesgadas. 6. Adquirir un conocimiento básico de las consideraciones éticas de la IA y de la importancia de desarrollar e implementar algoritmos imparciales. 7. Reconocer la necesidad de transparencia y rendición de cuentas en los sistemas de IA para garantizar un uso responsable. 8. Comprender la importancia de los derechos humanos y la equidad en el contexto de la IA y el papel de los algoritmos imparciales. 9. Desarrollar una conciencia básica de las posibles implicaciones legales y sociales del sesgo algorítmico y la necesidad de regulaciones y pautas apropiadas. 10. Cultivar el interés por aprender más sobre el sesgo algorítmico, la ética de la IA, fomentando un compromiso con el aprendizaje permanente y el compromiso responsable con la IA.	1. Analice escenarios simples de la vida real para identificar casos de sesgo algorítmico y su impacto. 2. Aplicar habilidades básicas de pensamiento crítico para evaluar la equidad de los sistemas de IA y sus procesos de toma de decisiones. 3. Desarrollar estrategias sencillas para minimizar la presencia de sesgos en los sistemas de IA, como el uso de datos más diversos y representativos. 4. Comunicar de manera efectiva sobre los conceptos básicos del sesgo algorítmico y sus consecuencias a audiencias no especializadas. 5. Utilizar las habilidades de resolución de problemas para abordar los desafíos básicos relacionados con el sesgo algorítmico y la equidad en los sistemas de IA. 6. Reflexionar sobre experiencias personales y los sesgos para comprender mejor la importancia de promover la equidad y la toma de decisiones imparcial en los sistemas de IA. 7. Colaborar con otros para identificar posibles problemas y desarrollar soluciones adecuadas. 8. Demostrar una conciencia ética básica en el contexto de la IA, comprendiendo la importancia del desarrollo y la implementación de una IA responsable. 9. Adaptarse a diferentes escenarios que impliquen sesgos algorítmicos, reconociendo la necesidad de mejora continua. 10. Demostrar una competencia proactiva para abordar el sesgo algorítmico, tomando la iniciativa para aprender más y participar en debates y actividades relevantes.	1. Apreciación de la importancia de la equidad y la toma de decisiones imparcial en los sistemas de IA, reconociendo las posibles consecuencias del sesgo algorítmico en las personas y la sociedad. 2. Responsabilidad ética, reconociendo el papel de las personas en el desarrollo y despliegue de sistemas de IA y la necesidad de actuar de forma ética y responsable. 3. Apertura a diversas perspectivas, estar dispuesto a considerar diferentes puntos de vista y experiencias al abordar cuestiones relacionadas con el sesgo algorítmico y la equidad. 4. Compromiso con el aprendizaje continuo y el crecimiento personal, entendiendo que abordar los sesgos algorítmicos requiere un aprendizaje y una adaptación continuos. 5. Colaboración y trabajo en equipo, valorando las contribuciones de los demás y reconociendo la importancia de trabajar juntos para abordar problemas complejos como el sesgo algorítmico. 6. Pensamiento crítico, cuestionar suposiciones y estar dispuesto a reevaluar las creencias y sesgos personales a la luz de nueva información o perspectivas. 7. Empatía y compasión, teniendo en cuenta el impacto del sesgo algorítmico en diferentes individuos y comunidades y esforzándose por promover la justicia y la equidad en los sistemas de IA. 8. Respeto a la privacidad y protección de datos, reconociendo la importancia de mantener la privacidad del usuario y salvaguardar la información personal en las aplicaciones de IA. 9. Conciencia social, comprensión de las implicaciones sociales más amplias del sesgo algorítmico y la necesidad de una acción colectiva para abordar estos desafíos. 10. Compromiso proactivo, tomar la iniciativa para participar en debates y actividades relacionadas con el sesgo algorítmico y la equidad, y abogar por el desarrollo y la implementación responsables de la IA.

Referencias

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Sesgo de la máquina. ProPublica, 23 de mayo.

Barocas, S., Hardt, M., & Narayanan, A. (2019). Equidad y abstracción en los sistemas sociotécnicos. En Conferencia ACM sobre Equidad, Rendición de Cuentas y Transparencia (pp. 59-68).

Benjamín, R. (2019). La carrera tras la tecnología: herramientas abolicionistas para el nuevo Código Jim. Estado.

Bentley, J. (1999). Perlas de programación. Addison-Wesley.

Billett, S. (2009). Darse cuenta del valor educativo de integrar las experiencias laborales en la educación superior. Estudios de Educación Superior, 34(7), 827-843.

Börner, K., Contratista, N., Falk-Krzesinski, H. J., Fiore, S. M., Hall, K. L., Keyton, J., ... y Uzzi, B. (2010). Una perspectiva de sistemas multinivel para la ciencia de la ciencia del equipo. Medicina Traslacional de la Ciencia, 2(49), 49cm24.

Bostrom, N. (2014). Superinteligencia: Caminos, Peligros, Estrategias. Oxford University Press.

Cavoukian, A. (2010). Privacidad por diseño: Los 7 principios fundamentales. Comisionado de Información y Privacidad de Ontario, Canadá.

Cedefop - Centro Europeo para el Desarrollo de la Formación Profesional - Marco Europeo de Cualificaciones (EQF: https://www.cedefop.europa.eu/en/events-and-projects/projects/european-qualifications-framework-eqf

Cook, W. (2016). En busca del viajante de comercio: las matemáticas en los límites de la computación. Princeton University Press.

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

Cottrell, S. (2018). Habilidades para el éxito: Desarrollo personal y empleabilidad. Macmillan International Higher Education.

Cuseo, J. (2010). El caso empírico para el seminario de primer año: Promover resultados positivos para los estudiantes y beneficios para todo el

campus. El seminario de primer año: Recomendaciones basadas en la investigación para el diseño, la impartición y la evaluación de cursos. Dubuque, IA: Kendall Hunt.

Danks, D., & London, A. J. (2017). Detección y mitigación de sesgos algorítmicos: mejores prácticas y políticas para reducir los daños al consumidor. Biblioteca Digital ACM.

Díaz, A., & Sonsini, G. (2016). La toma de decisiones algorítmica y la ley. Política de Telecomunicaciones, 40(11), 1027-1040.

Doe, J., & Roe, J. (2020). Comprensión de los modelos de algoritmos: un enfoque integral. Revista de Ciencias de la Computación, 18(3), 300-316.

Eraut, M. (2004). Aprendizaje informal en el lugar de trabajo. Estudios de Educación Continua, 26(2), 247-273.

Eubanks, V. (2018). Automatización de la desigualdad: cómo las herramientas de alta tecnología perfilan, vigilan y castigan a los pobres. Imprenta de San Martín.

Comisión Europea - Marco Europeo de Cualificaciones (EQF): https://ec.europa.eu/ploteus/en/content/descriptors-page

Comisión Europea - Espacio Europeo de Educación Superior (EEES): https://ec.europa.eu/education/policies/higher-education/european-higher-education-area_en

Comisión Europea (2019). Directrices éticas de la IA fiable. Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial. Disponible el 20.6.2023 a las https://op.europa.eu/o/opportal-service/download-handler?identifier=d3988569-0434-11ea-8c1f-01aa75ed71a1&format=pdf&language=en&productionSystem=cellar&part=

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. y Srikumar, M. (2020). Inteligencia artificial basada en principios: mapeo del consenso en enfoques éticos y basados en los derechos para los principios de la IA. Centro Berkman Klein para Internet y Sociedad. Universidad de Harvard. Disponible el 20.6.2023 https://dash.harvard.edu/bitstream/handle/1/42160420/HLS%20White%20Paper%20Final_v3.pdf?sequence=1&isAllowed=y

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... y Vayena, E. (2018). AI4People: un marco ético para una buena sociedad de

IA: oportunidades, riesgos, principios y recomendaciones. Mentes y máquinas, 28, 689-707.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... y Schafer, B. (2018). AI4People: un marco ético para una buena sociedad de IA: oportunidades, riesgos, principios y recomendaciones. Mentes y Máquinas, 28(4), 689-707.

Friedman, B. y Nissenbaum, H. (1996). Sesgo en Sistemas Informáticos. ACM Transactions on Information Systems, vol. 14, núm. 3, julio de 1996, páginas 330-347. Disponible el 20.6.2023 en https://nissenbaum.tech.cornell.edu/papers/biasincomputers.pdf

Jobin, A., Ienca, M. y Vayena, E. (2019). Inteligencia Artificial: El Panorama Global de las Directrices Éticas. Disponible el 20.6.2023 https://arxiv.org/ftp/arxiv/papers/1906/1906.11668.pdf

Jobin, A., Ienca, M., & Vayena, E. (2019). El panorama global de las directrices éticas de la IA. Nature Machine Intelligence, 1(9), 389-399.

Johnson, D. (2003). Guía teórica para el análisis experimental de algoritmos. Estructuras de datos, búsquedas de vecinos cercanos y metodología: Quinto y sexto desafíos de implementación de DIMACS, 59.

Karp, R. (2010). El arte de la optimización de algoritmos. ACM Computing Surveys (CSUR), 42(4), 1-28.

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

Lambrecht, A., & Tucker, C. (2019). ¿Sesgo algorítmico? Un estudio empírico de la aparente discriminación basada en el género en la visualización de anuncios de carreras STEM. Ciencias de la Administración, 65(7), 2966-2981.

Lange, A. R. y Duarte, N. (2017). Comprensión del sesgo en el diseño algorítmico. Disponible el 20.6.2023 https://medium.com/impact-engineered/understanding-bias-in-algorithmic-design-db9847103b6e

Latour, B. (2005). Reensamblando lo social: una introducción a la teoría del actor-red. Oxford University Press.

Lee, N. T., Resnick, P., & Barton, G. (2019). Detección y mitigación de sesgos algorítmicos: mejores prácticas y políticas para reducir los daños al consumidor. Disponible el 20.6.2023 en

https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

Li, Y., Hu, B., Chen, M., & Zhong, Q. (2017). Comprender las limitaciones de los algoritmos: una revisión exhaustiva. Revista de Ciencias de la Información, 43(4), 528-545.

Mansfield, R. S. (2009). Las competencias requeridas por el rol de los desarrolladores de recursos humanos. Mejora del rendimiento trimestral, 22(2), 33-49.

Milano, S., Taddeo, M., & Floridi, L. (2020). Los sistemas de recomendación y sus desafíos éticos. Ai & Society, 35, 957-967.

Miller, G. (2022). Roles de las partes interesadas en proyectos de inteligencia artificial. Proyecto Liderazgo y Sociedad, Vol. 3, diciembre de 2022. Disponible el 20.6.2023 https://www.sciencedirect.com/science/article/pii/S266672152200028X

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). La ética de los algoritmos: Mapeo del debate. Big Data y Sociedad, 3(2), 2053951716679679.

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). La ética de los algoritmos: Mapeo del debate. Big Data y Sociedad, 3(2), 2053951716679679.

Morley, J., Floridi, L., Kinsey, L. y Elhalal, A. (2019). Del qué al cómo: una visión general de las herramientas, los métodos y la investigación de la ética de la IA para traducir los principios en prácticas. Disponible el 20.6.2023 https://www.researchgate.net/profile/Jessica-Morley/publication/333103986_From_What_to_How_An_Overview_of_AI_Ethics_Tools_Methods_and_Research_to_Translate_Principles_into_Practices/links/5cdbb95ca6fdccc9ddae53db/From-What-to-How-An-Overview-of-AI-Ethics-Tools-Methods-and-Research-to-Translate-Principles-into-Practices.pdf?origin=publication_detail

Nissenbaum, H. (2009). La privacidad en contexto: tecnología, política e integridad de la vida social. Stanford University Press.

Noble, S. U. (2018). Algoritmos de opresión: cómo los motores de búsqueda refuerzan el racismo. Prensa de la Universidad de Nueva York.

Novelli, C., Taddeo, M. y Floridi, L. (2023). Rendición de cuentas en Inteligencia Artificial: qué es y cómo funciona. IA Y SOCIEDAD. Disponible el 20.6.2023

https://link.springer.com/content/pdf/10.1007/s00146-023-01635-y.pdf?pdf=button

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Diseccionando el sesgo racial en un algoritmo utilizado para gestionar la salud de las poblaciones. Ciencias, 366(6464), 447-453.

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

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

Paraschakis, D. (2018). Aspectos algorítmicos y éticos de los sistemas de recomendación en el comercio electrónico (Tesis doctoral, Universidad de Malmö, Facultad de Tecnología y Sociedad).

Parker, G., Van Alstyne, M., & Choudary, S. (2016). Revolución de plataformas: cómo la red

Pasquale, F. (2015). La sociedad de la caja negra: los algoritmos secretos que controlan el dinero y la información. Editorial de la Universidad de Harvard.

Perrault R, Yoav S, Brynjolfsson E, Jack C, Etchmendy J, Grosz B, Terah L, James M, Saurabh M, Carlos NJ (2019). *Informe del Índice de Inteligencia Artificial 2019*. https://wp.oecd.ai/app/uploads/2020/07/ai_index_2019_introduction.pdf

Prates, M. O., Avelar, P. H., & Lamb, L. C. (2020). Evaluación del sesgo de género en la traducción automática: un estudio de caso con Google Translate. Computación Neuronal y Aplicaciones, 32, 6363-6381.

Ryan, M., & Stahl, B. C. (2020). Directrices éticas de la inteligencia artificial para desarrolladores y usuarios: aclaración de su contenido e implicaciones normativas. Revista de Información, Comunicación y Ética en la Sociedad.

Stewart, J., & Bartrum, D. (1999). Hacia una matriz de competencias para el desarrollo de los recursos humanos. Formación Industrial y Comercial, 31(5), 176-181.

UNESCO (2021). Recomendación sobre la Ética de la Inteligencia Artificial. Disponible el 20.6.2023 en https://unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf

Wong, P. H. (2020). Democratizar la equidad algorítmica. Filosofía y Tecnología, 33, 225-244.

Zhou, J., Chen, Z., Berry, A., Reed, M., Zhang, S. y Savage. S. (2020). Una encuesta sobre los principios éticos de la IA y sus implementaciones. IEEE Xplore. Disponible el 20.6.2023 en https://opus.lib.uts.edu.au/bitstream/10453/146673/2/IEEE_ETHAI2020_ethicalAI_Survey.pdf

Zuboff, S. (2019). La era del capitalismo de vigilancia: la lucha por un futuro humano en la nueva frontera del poder. Asuntos Públicos.

DESAFIANDO EL SESGO EN BIG DATA PARA LA IA Y EL APRENDIZAJE AUTOMÁTICO

Matrices de competencias del WP2 para evitar el "sesgo algorítmico"

Cofinanciado por la Unión Europea

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