

Unidad 5.2. Consideraciones éticas.

Introducción

“Un gran poder conlleva una gran responsabilidad”. Origen desconocido, atribuida a Damocles en el siglo I A.C., y que también se ha usado por políticos, periodistas y autores de ficción para sus personajes.



Esta cita encapsula la idea central de que la capacidad transformadora de la IA implica una obligación moral de usarla de manera que beneficie a la sociedad y evite los riesgos dañinos de un uso indebido. Al fin y al cabo se trata de una herramienta poderosa que mal utilizada puede llegar a ser muy perjudicial para la sociedad. En alguna ocasión se la ha comparado con la energía nuclear, capaz de proporcionar electricidad a gran escala con emisiones mínimas de carbono, mientras que manejada de manera incorrecta puede resultar en desastres catastróficos, como lo demuestran los accidentes de Chernobyl y Fukushima o una proliferación nuclear que puede conducir a amenazas de seguridad global.

Como hemos ido descubriendo, en la era contemporánea, la Inteligencia Artificial ha impactado en multitud de sectores, redefiniendo la forma en que operan y creando un horizonte de posibilidades inexploradas. El sector educativo no es una excepción a esta transformación digital, donde la IA

promete revolucionar la manera en que aprendemos y enseñamos. Sin embargo, junto con las promesas de eficiencia y personalización, emergen importantes cuestiones éticas que requieren una cuidadosa exploración y entendimiento.

La IA en educación tiene el potencial de personalizar y democratizar la enseñanza, optimizar los procesos administrativos liberando tiempo de los docentes para realizar tareas específicas del proceso de enseñanza-aprendizaje y la tutorización. También puede permitir disponer de un conocimiento valioso basado en el análisis de grandes conjuntos de datos. Sin embargo, la implementación de estas tecnologías conlleva responsabilidades éticas significativas. Los **dilemas éticos** relacionados con la **privacidad, el sesgo, y la autonomía**, entre otros, son cruciales y demandan una consideración seria por parte de todos los actores involucrados en el ámbito educativo.

Una serie de preguntas han de asaltarnos antes de lanzarnos a implementar sistemas basados de IA en educación: ¿Cómo asegurarnos de que los sistemas de IA sean justos y equitativos? ¿Cómo podemos evitar que se utilicen para manipular o discriminar a los estudiantes? y muchas otras.

Veremos como no estamos solos ante este nuevo horizonte y como desde la irrupción de la IA a nivel global, instituciones públicas y privadas de alcance mundial han ahondado en la necesidad de garantizar que los desarrollos de IA tanto a nivel general como especialmente a nivel educativo se realicen dentro de un **marco ético y controlado**.

El propósito de esta unidad es explorar la ética en el contexto de la IA en general, para posteriormente adentrarnos en las implicaciones éticas específicas cuando la IA se aplica en el escenario educativo. Realizaremos un repaso a los principales dilemas éticos que rodean a la IA y su traducción al contexto educativo.

Al finalizar esta unidad, tendrás una comprensión sólida de los desafíos éticos que plantea la IA en la educación, y estarás en disposición de afrontarlos de manera responsable y efectiva. Se busca fomentar una reflexión crítica y proporcionar una base sólida para la implementación ética de la IA en la educación, contribuyendo así a un **futuro educativo más inclusivo, justo y humanizado**.

Ética en la IA

Una definición de ética nos dice que es una disciplina filosófica que estudia el bien y el mal y sus relaciones con la moral y el comportamiento humano. Por tanto la ética de la IA se centra en los principios morales que deben guiar el desarrollo y el uso de la IA.

Desde un punto de vista práctico la ética de la IA ha de ser un **conjunto de directrices que asesore sobre el diseño, la implementación y los resultados que de la inteligencia artificial se quiera obtener.**

Una de los principios básicos de funcionamiento de la IA es la necesidad de contar con una gran cantidad de **datos**, esos datos son generados en gran medida por las interacciones humanas con el mundo digital, pues bien las personas tienen todo tipo de **sesgos cognitivos**, como el sesgo de confirmación, el de arrastre, efecto halo, pensamiento de grupo y muchos otros. Estos sesgos están patentes en nuestros comportamientos y, por tanto, en nuestros datos. Por tanto, siendo los datos la base de funcionamiento de los algoritmos de machine learning, es vital, intentar evitar que los sesgos terminen permeando en el sistema si no queremos que los algoritmos de IA nazcan nutridos de ellos. De otro modo la inteligencia artificial tiene altas probabilidades de escalar estos sesgos humanos a un ritmo sin precedentes.

El big data, ha sido un acicate para que las empresas desarrollen sistemas cada vez mas automatizados capaces de tomar decisiones basadas en datos. Su intención primigenia, siendo generalmente entidades con ánimo de lucro, ha sido un incremento en los resultados, sin embargo en ocasiones están experimentando situaciones imprevistas, como consecuencia de una falta de investigación inicial, posterior control o a los sesgos subyacentes de los conjuntos de datos. Los sesgos, es uno de los aspectos mas preocupantes cuando se habla de la etica en la IA, pero no es el único como veremos mas adelante.

Obviamente, los errores conducen a correcciones y han surgido nuevas pautas de trabajo e investigaciones desde el campo de la ciencia de datos. Las empresas líderes en IA están interesadas también en establecer pautas éticas para su uso, ya que han experimentado directamente las consecuencias de no hacerlo. La falta de diligencia en esta área puede tener implicaciones legales y financieras, así como daños a la reputación. Como con cualquier avance tecnológico, la innovación en IA suele ir por delante de la regulación gubernamental. Si bien, la reacción no se ha hecho esperar demasiado y también diversas instituciones públicas a nivel mundial han emitido ya informes y pautas de recomendación sobre un uso ético de la IA.

Las preocupaciones manifestadas y las recomendaciones y principios comunes de actuación, tanto si provienen de instituciones privadas como públicas se agrupan mayoritariamente en los siguientes aspectos:

Preocupaciones de la IA en general

Privacidad

Antes de la existencia de la IA, ya existía esta preocupación debido a la tecnología digital existente (internet, telefonía móvil, hiperconectividad...). Sin embargo, la IA introduce nuevas dimensiones a

estos riesgos o los amplifica. Estas nuevas dimensiones en cuanto a riesgos en la privacidad surgen de las supuestas bondades que la IA nos ofrece como:

Análisis predictivo y perfilación. La IA puede analizar grandes volúmenes de datos para predecir comportamientos y preferencias individuales con un alto grado de precisión. Esto permite una perfilación detallada que puede ser utilizada de manera invasiva o discriminatoria.

Reidentificación de datos anonimizados. Técnicas avanzadas de IA pueden desanonimizar datos que inicialmente fueron anonimizados, permitiendo la reidentificación de individuos, lo cual era mucho más difícil sin la IA.

Tecnologías de reconocimiento. La IA potencia tecnologías como el reconocimiento facial, de voz o de patrones de comportamiento, lo cual permite una vigilancia más invasiva y precisa en tiempo real.

Automatización de decisiones. La IA permite la toma de decisiones automatizadas basadas en datos personales, lo que puede resultar en discriminación o injusticia si los sistemas están sesgados o mal configurados, y esto puede ser más difícil de detectar o impugnar en comparación con decisiones tomadas por humanos.

Si bien, es en estos campos donde los creadores de políticas han podido hacer más avances en los últimos años. Por ejemplo, en 2016 se instauró el Reglamento General de Protección de Datos (RGPD) con el objetivo de resguardar la información personal de los ciudadanos en la Unión Europea y el Espacio Económico Europeo, brindándoles más control sobre sus propios datos. En Estados Unidos, varios estados han estado formulando políticas, como la Ley de Privacidad del Consumidor de California (CCPA), que exige a las empresas ser transparentes sobre cómo recopilan la información del consumidor. Estas nuevas legislaciones han impulsado a las empresas a reconsiderar sus prácticas de almacenamiento y uso de los datos de identificación personal (PII). Como consecuencia, la inversión en seguridad se ha catapultado como una prioridad para las empresas en su esfuerzo por mitigar cualquier vulnerabilidad y prevenir la vigilancia, hackeos y ciberataques.

Sesgo y discriminación

Los casos de sesgo y discriminación detectados en varios sistemas inteligentes han planteado muchas preguntas éticas sobre el uso de la inteligencia artificial. **¿Cómo podemos protegernos contra el sesgo y la discriminación cuando los propios datos de entrenamiento pueden estar sesgados?** En principio, las empresas ponen sus mejores intenciones en sus iniciativas de automatización, pero hay diversos informes y casos concretos que evidencian que no siempre tienen éxito. Por ejemplo [Reuters](#) destaca algunas de las consecuencias imprevistas de incorporar la IA en las prácticas de contratación. En su esfuerzo por automatizar y simplificar un proceso, Amazon involuntariamente sesgó los posibles candidatos a un trabajo por género para puestos

técnicos libres y, en última instancia, tuvieron que desechar el proyecto. A medida que se descubren sucesos como este, [Harvard Business Review](#) ha planteado otras cuestiones sobre el uso de la IA en las prácticas de contratación como, por ejemplo, **qué datos se deberían poder utilizar al evaluar a un candidato para un puesto.**

El sesgo y la discriminación no se encuentran únicamente en los sistemas de contratación, también están presentes otras aplicaciones de la IA, como reconocimiento facial o los algoritmos de redes sociales.

Con el creciente entendimiento de los riesgos asociados con la IA, las corporaciones están participando más activamente en las discusiones relativas a la ética y los valores de la IA. Por ejemplo, Arvind Krishna, el director ejecutivo de IBM, declaró que la empresa ha retirado sus productos de análisis y reconocimiento facial de uso general, subrayando que:

“ IBM se opone firmemente y no tolerará los usos de ninguna tecnología, incluida la tecnología de reconocimiento facial ofrecida por otros proveedores, para la vigilancia masiva, la creación de perfiles raciales, las violaciones de derechos humanos y otras libertades básicas, o cualquier propósito que no concuerde con nuestros valores y Principios de confianza y transparencia”. Arvind Krishna, CEO IBM

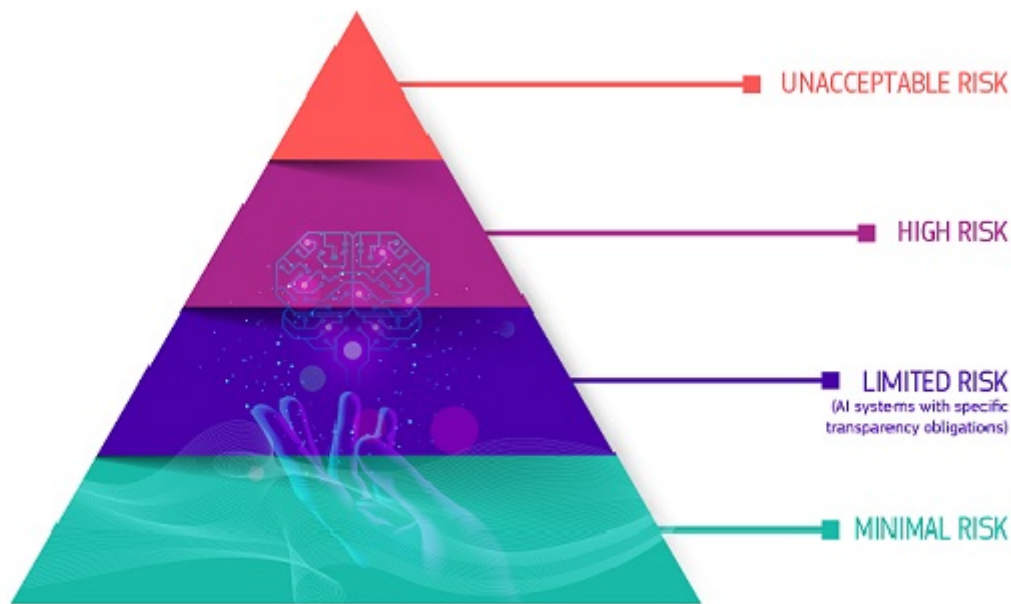
Impacto de la IA en el empleo

La inquietud pública sobre la IA suele enfocarse en la pérdida de empleos, pero hay motivos para pensar que debería replantearse el enfoque. Cada avance tecnológico altera la demanda laboral en ciertos sectores. Por ejemplo, en la industria automotriz, las principales marcas se están inclinando hacia los vehículos eléctricos acorde a tendencias ecológicas. La industria energética no va a desaparecer, pero la fuente de energía está cambiando de una economía basada en el combustible a otra eléctrica. La inteligencia artificial puede mirarse bajo el mismo prisma. La IA reconfigurará la demanda laboral hacia otras áreas, necesitando por ejemplo trabajadores para gestionar estos sistemas y resolver problemas complejos en sectores como el servicio al cliente, adaptándose así a las nuevas exigencias del mercado.

Responsabilidad

La nueva ley de la UE, aprobada por el Parlamento Europeo el **13 de marzo de 2024** y que esta destinada a regular la IA según un enfoque basado en el riesgo, tiene como objetivos principales garantizar que los sistemas de IA utilizados en la Unión Europea e introducidos en el mercado sean seguros y respeten los derechos de los ciudadanos, y estimular la inversión y la innovación en el

ámbito de la IA. Se aplicará basándose en el nivel de riesgo de los sistemas de IA, estableciendo normas más estrictas para aquellos considerados de alto riesgo y prohibiendo prácticas inaceptables como la manipulación cognitiva. Además, se creará la Oficina de la IA para supervisar el cumplimiento de esta normativa. Está previsto que, si no hay imprevistos, la ley entre totalmente en vigor hacia finales de 2026, aunque algunas partes se implementarán antes. [Más detalle sobre la ley.](#)



Hasta la puesta en vigor de esta ley los únicos incentivos para que las empresas sigan unas directrices éticas son las repercusiones negativas que un sistema de IA poco ético puede tener en el resultado final. Como sustituto de esta carencia, se han realizado tentativas de encauzar el desarrollo de la IA mediante diversos comunicados y cartas abiertas de algunos de los pioneros y desarrolladores destacados en este campo, como la carta: [Pause Giant AI Experiments: An Open Letter](#) en la que varios de los principales actores en el campo de la IA y mas de un millar de expertos solicitaban frenar los desarrollo de IA hasta que existiera un consenso sobre el desarrollo y el uso de la misma.

También han surgido marcos éticos como parte de una colaboración entre expertos en ética e investigadores para gobernar la construcción y distribución de modelos de IA dentro de la sociedad. No obstante, por ahora, solo sirven como guía, como ejemplo este estudio de la Universidad de Tubinga en Alemania (Centro Internacional para la Ética en Ciencias y Humanidades): [The Ethics of AI Ethics](#). El estudio aborda la proliferación de guías éticas en respuesta al rápido desarrollo y aplicación de sistemas de inteligencia artificial. Sin embargo, según el mismo, estas guías, solo sirven de momento como recomendaciones. El estudio no proporciona un análisis optimista sobre la eficacia de las guías éticas analizadas para gobernar el

desarrollo de modelos de IA en la sociedad. Por lo tanto, concluye, aunque se han establecido marcos éticos, su implementación práctica y su capacidad para mitigar los riesgos asociados con la IA aún están en una fase inicial, sirviendo más como una guía que como una regulación efectiva.

Recomendaciones éticas para la IA en general

Las preocupaciones expuestas han llevado, como se ha adelantado a la emisión de comunicados e informes por parte de organismos e instituciones encaminadas a recomendar pautas éticas en el desarrollo de modelos de IA así como a guiar el uso ético de la misma.

Muchos de estos informes vienen a concluir en una serie de puntos clave. Extraemos como ejemplo los cuatro aspectos que del informe: [Directrices éticas para una IA fiable](#), publicado en abril del 2019 por la Comisión Europea y desarrollado por un grupo de expertos de alto nivel. En este estudio se concreta la visión europea sobre la IA. De acuerdo con estas directrices, Europa tiene que investigar y desarrollar una IA fiable, es decir, **lícita, ética y robusta**.

Los cuatro principios éticos identificados en las directrices son:

1. **Principio de respeto a la autonomía humana:** Este principio valora y quiere preservar la capacidad de las personas para tomar decisiones independientes, asegurando que la IA no manipule, engañe o reduzca la libertad de elección de los individuos.
2. **Principio de prevención de los daños:** Prioriza la identificación y mitigación de los riesgos y daños potenciales que pueden surgir del uso de la IA, promoviendo el bienestar y la seguridad de las personas y comunidades.
3. **Principio de equidad:** Promueve la justicia y la igualdad en el acceso y los beneficios de la IA, además de abordar y corregir los sesgos sistémicos para evitar la discriminación.
4. **Principio de explicabilidad:** Asegura que las decisiones tomadas por sistemas de IA sean comprensibles y rastreables para los usuarios y partes interesadas, permitiendo una mejor comprensión y la clara identificación de responsabilidades.

En relación con recomendaciones éticas en la aplicación de la IA a nivel general puedes leer también el informe emitido por la UNESCO: [Recomendación sobre la ética de la inteligencia artificial](#)



También de la UNESCO, dejamos aquí un ilustrativo artículo: [Inteligencia artificial: ejemplos de dilemas éticos](#), con casos concretos sobre dilemas éticos aplicables a la IA. Entre ellos, este clarificador video:

<https://www.youtube.com/embed/HzYG56HLxbl>

Ética en la IA educativa

Hemos visto preocupaciones y recomendaciones aplicables al desarrollo y aplicación de la IA en general. Estos principios son también aplicables a la educación. Si bien podemos establecer matices debido a las posibilidades y peligros específicos que la IA tiene en el ámbito educativo.

Se han realizado publicaciones específicas al respecto de las consideraciones éticas en la educación de las que se pueden extraer una serie de puntos clave comunes, tanto a la hora de plantear las preocupaciones como en las recomendaciones de los principios de actuación.

Preocupaciones en la IA educativa

Privacidad

El uso de la IA en el ámbito educativo supone recopilar y analizar grandes cantidades de datos de los estudiantes. La recopilación y el uso de datos personales supone un **riesgo para su privacidad**, si los datos no se gestionan o protegen adecuadamente.

Deshumanización del aprendizaje

Un incremento paulatino del uso de la IA en la educación, puede traducirse en una pérdida en la conexión interpersonal entre profesores y estudiantes. El exceso de automatización y el enfoque en la eficiencia pueden restar valor a la experiencia educativa, **dejando de lado** aspectos importantes como la **empatía, la creatividad y el pensamiento crítico**.

Sesgos y discriminación

Como ya hemos visto, los algoritmos de IA pueden estar sesgados debido a los datos con los que son entrenados. Esto puede llevar a decisiones discriminatorias en la educación, afectando la calidad de la enseñanza, la orientación profesional y las oportunidades académicas de los estudiantes. Los sesgos implícitos en la IA pueden **agravar las desigualdades** existentes y generar situaciones injustas para ciertos grupos de estudiantes.

Dependencia excesiva

Una utilización indiscriminada de sistemas de IA en la esfera educativa, existe la posibilidad de que los alumnos lleguen a depender excesivamente de la tecnología, lo que podría menguar sus aptitudes para pensar de forma autónoma y resolver problemas por sí mismos. El abuso de la IA en la educación podría **comprometer la independencia y el crecimiento intelectual** de los estudiantes.

Falta de supervisión y control humano

A pesar de su considerable potencia, la IA aún presenta restricciones y puede cometer fallos. La ausencia de una supervisión y control humanos apropiados en la aplicación de la IA en entornos educativos podría resultar en **imprecisiones o efectos negativos en el proceso de enseñanza-aprendizaje**.

Distracción tecnológica

La presencia de tecnología de IA en el entorno escolar podría generar distracciones para los alumnos o provocar una dependencia excesiva de dispositivos y sistemas, lo que podría **mermar su habilidad para enfocarse** en el contenido educativo **y para interactuar de manera efectiva con sus compañeros y docentes**.

Recomendaciones éticas para la IA educativa

También en el ámbito educativo las organismos e instituciones publicas y privadas han emitido informes guía con el propósito de establecer unos principios de actuación para el uso de la IA educativa. La UNESCO, como consecuencia del [Consenso de Beijing](#), emitió el informe [Inteligencia artificial: guía para las personas a cargo de formular políticas](#), en cuyo resumen inicial se cita :

“ La inteligencia artificial (IA) tiene la capacidad de hacer frente a algunos de los mayores desafíos que afronta, hoy en día, el ámbito de la educación, de desarrollar prácticas de enseñanza y aprendizaje innovadoras y, finalmente, de acelerar el progreso en la consecución del Objetivo de Desarrollo Sostenible (ODS) 4. No obstante, estos avances tecnológicos rápidos implican inevitablemente numerosos riesgos y retos, que los debates sobre las políticas y los marcos reglamentarios tienen aún dificultades para superar. Esta publicación ofrece orientación a las personas a cargo de formular políticas sobre la mejor manera de aprovechar las oportunidades y enfrentar los riesgos que presenta la creciente conexión entre la IA y la educación.

Todo el informe merece una lectura, si bien, en el ámbito de recomendaciones para el uso de la IA en educación en el que se enfoca el apartado de esta unidad, es el capítulo 6 del informe el que las detalla. La base de tales recomendaciones son **cuatro objetivos estratégicos**, si bien habrían de ser interpretados en función del contexto local:

- **Garantizar el uso inclusivo y equitativo de la IA en la educación.**
- **Hacer uso de la IA para mejorar la educación y el aprendizaje.**
- **Promover el desarrollo de habilidades para la vida en la era de la IA, incluyendo la enseñanza del funcionamiento de la IA y sus implicancias para la humanidad.**
- **Salvaguardar el uso transparente y auditable de los datos educativos.**

Estos cuatro principios estratégicos son desarrollados en el informe en las siguientes **recomendaciones**:

- **Planificación interdisciplinaria y gobernanza intersectorial.**
- **Políticas sobre el uso equitativo, inclusivo y ético de la IA.**
- **Desarrollo de un plan maestro para el uso de la IA en educación, en la gestión, enseñanza, aprendizaje y evaluación.**

- **Pruebas piloto, seguimiento y evaluación, así como la creación de una base de pruebas.**
- **Promoción de las innovaciones locales de IA para la educación.**

Todos estos puntos están ampliamente desarrollados en el informe, incluyen ejemplos de iniciativas en curso y su lectura es recomendable para comprender las eventuales hojas de ruta a implantar por las instituciones educativas.

La **Unión Europea**, dentro de su Plan de Acción en Educación Digital 2021-2027 ha incluido una guía ética para el uso de la IA. en el que indica que hay cuatro **consideraciones esenciales** que respaldan el **uso ético** de la IA en la enseñanza, el aprendizaje y la evaluación:

- **Capacidad de acción humana:** se refiere a la habilidad de las personas para tomar decisiones informadas y ser responsables de sus acciones. Incluye conceptos como la autonomía y la autodeterminación.
- **Equidad:** implica tratar con justicia a todas las personas, asegurando que todos tengan acceso igualitario a oportunidades y recursos. Esto abarca la inclusión, la no discriminación, etc.
- **Humanidad:** considera la identidad, integridad y dignidad de las personas. Resalta la importancia del bienestar, la seguridad, el tratar a las personas con respeto y no como simples objetos de datos.
- **Elección justificada:** hace referencia al uso de conocimientos, experiencia y datos para respaldar elecciones en entornos educativos. Esto requiere transparencia, modelos participativos y explicabilidad en la toma de decisiones.

Haz click en el siguiente enlace para ver el video explicativo de la UE sobre la guía ética en el uso de la IA en educación:

[Ethical guidelines on the use of artificial intelligence \(AI\) and data in teaching and learning, por European Union](#)

A otro nivel, es necesario que los educadores tengan claras determinadas pautas que les permitan evitar o mitigar los riesgos inherentes a las preocupaciones antes detalladas. En este sentido, el siguiente video realizado por *Code.org*®, una organización sin fines de lucro dedicada a expandir la participación en ciencias de la computación, desarrolla **recomendaciones prácticas expresamente dirigidas a los docentes; Asegurando una IA responsable.**



<https://www.youtube.com/embed/P14XpLzaEac?t=41s>

Revision #17

Created 2023-10-07 11:18:42 CEST by Pedro López

Updated 2024-03-17 20:00:20 CET by Pedro López