

03/4

Interacción entre la ética y la Inteligencia Artificial.

Manuel Calvo Muñoz,

Psicólogo y Director técnico.
Residencia de Discapacidad Intelectual
de Arroyomolinos.
Complejo Asistencial Benito Menni.
Ciempozuelos (Madrid)

La relación entre la ética y la inteligencia artificial (IA) es un campo en rápida evolución que aborda las profundas implicaciones de las tecnologías de IA en la sociedad, los derechos humanos y los principios morales. A medida que los sistemas de IA se integran cada vez más en la vida cotidiana, desde la atención médica hasta la aplicación de la ley, los desafíos éticos que presentan, incluidos los problemas de sesgo, privacidad, responsabilidad y transparencia, han atraído una atención significativa entre académicos, legisladores y líderes de la industria.

La urgencia de estas discusiones surge de la naturaleza de doble uso de la IA, donde las tecnologías diseñadas para fines beneficiosos también pueden ser mal utilizadas, lo que lleva a consecuencias perjudiciales. Los patrones históricos en el desarrollo tecnológico a menudo revelan un enfoque reactivo a las consideraciones éticas, con muchos problemas recibiendo atención solo después de que han ocurrido consecuencias negativas. Esta tendencia subraya la necesidad de una **ética proactiva** que informe el diseño, desarrollo y despliegue de los sistemas de IA desde su inicio.

Diversos marcos éticos, basados en filosofías establecidas, han surgido para guiar a las partes interesadas en la navegación del complejo panorama de la IA, enfatizando **principios como la justicia, la beneficencia, la no maleficencia, la autonomía, la privacidad, la transparencia y la responsabilidad.**

Los problemas éticos clave, particularmente en relación con el sesgo y la equidad, han cobrado relevancia a medida que se demuestra que los sistemas de IA heredan prejuicios presentes en los datos de entrenamiento, lo que resulta en prácticas discriminatorias en dominios críticos como la contratación y la aplicación de la ley.

Además, las preocupaciones sobre la privacidad en relación con las vastas cantidades de datos personales que requieren los sistemas de IA destacan la necesidad de estrategias robustas de protección de datos y el cumplimiento de mar-

cos regulatorios, como el **Reglamento General de Protección de Datos (GDPR).**

A medida que estas discusiones se desarrollan, los debates en curso sobre la gobernanza y la regulación de las tecnologías de IA señalan la importancia de establecer estándares éticos que se alineen con los valores sociales y de las instituciones promoviendo una innovación responsable.

A la luz de estos desafíos, diversas organizaciones e instituciones académicas están abogando por **iniciativas educativas integrales que equipen a los futuros líderes con las herramientas necesarias para abordar las dimensiones éticas de la IA.** La interacción entre la ética y la IA seguirá siendo un área vital de enfoque, ya que las partes interesadas buscan equilibrar el avance tecnológico con la responsabilidad moral y la confianza pública en los sistemas de IA.

1/

Contexto histórico.

La relación entre la ética y la inteligencia artificial (IA) ha evolucionado significativamente a lo largo del relativo corto espacio de tiempo en el que la IA convive en nuestro día a día, reflejando tendencias históricas más amplias en tecnología y sociedad.

Así mismo la revolución digital, similar a las revoluciones agrícola e industrial, representa un período transformador que exige una reevaluación de los marcos éticos y las normas sociales, introduciendo desafíos complejos que requieren, no solo de consideraciones éticas, sino también del rediseño de herramientas conceptuales para navegar eficazmente con las implicaciones de las tecnologías en su rápida evolución.

Sin ir más lejos, históricamente, tecnologías como la energía nuclear y la automovilística han provocado extensos discursos éticos y políticos, a menudo solo después de que hayan surgido consecuencias negativas. Este patrón de **consideración ética reactiva** destaca la importancia de enfoques proactivos en el contexto de la IA que nos aplica.

Las tecnologías emergentes presentan no solo dilemas éticos, sino que también desafían los marcos morales existentes, llevando a filósofos y éticos a abogar por un enfoque más integrado que combine el diseño responsable con el uso responsable.

A medida que los sistemas de IA se vuelven cada vez más autónomos, los problemas éticos que rodean su despliegue ganan preponderancia. Parte de estas preocupaciones demandan una exploración y comprensión exhaustivas que incluyen:

- sesgo,
- discriminación,
- invasión de la privacidad y
- responsabilidad

La discusión en torno a la ética de la IA enfatiza así la importancia de las tecnologías de doble uso, donde las mismas aplicaciones de IA pueden ser utilizadas tanto para fines beneficiosos como perjudiciales complicando esta dualidad la toma de decisiones éticas, haciendo imperativo que los desarrolladores y legisladores consideren varios casos de uso y sus implicaciones.

Las organizaciones e instituciones académicas han comenzado a reconocer la urgencia de abordar estos desafíos éticos, enfatizando la necesidad de iniciativas educativas que informen a los futuros líderes sobre las dimensiones éticas de la tecnología, proponiéndose diversas directrices éticas para navegar las complejidades de la IA con el objetivo de asegurar que los avances tecnológicos se alineen con los valores humanos y el bienestar social.

LH n.340

Este diálogo en curso marca un punto crítico en el contexto histórico de la ética y la IA, moldeando la trayectoria del desarrollo tecnológico de una manera que prioriza la dignidad humana y la responsabilidad.

2/

Principios éticos fundamentales.

Los marcos éticos en el contexto de la inteligencia artificial (IA) son críticos para guiar el diseño, desarrollo y despliegue de sistemas de IA. Estos marcos se basan en filosofías éticas establecidas y son esenciales para abordar las complejidades y desafíos que plantean las tecnologías de IA. La aplicación de principios éticos fundamentales, como el impacto, la justicia y la autonomía, la beneficencia, la no maleficencia, la privacidad, la transparencia y la responsabilidad, se ha propuesto para guiar las discusiones en torno a las implicaciones éticas de la IA.

Estos principios se alinean bien con los utilizados en la bioética, que comparte similitudes con la ética digital al abordar nuevas formas de agentes y entornos. Esta convergencia destaca la necesidad de incorporar consideraciones éticas desde el principio en el proceso de diseño para asegurar que los sistemas de IA beneficien a la sociedad mientras preservan los derechos individuales:

1. Enfoque Basado en Principios

Un enfoque basado en principios para la ética en la IA enfatiza la importancia de la adaptabilidad y la capacidad de respuesta.

Por ejemplo, las partes interesadas deben hacer que los conocimientos sean accesibles para aquellos afectados por los procesos de toma de decisiones de la IA y estar preparados para revisar las consideraciones éticas a medida que los contextos evolucionan, como a través de nuevas regulaciones o comentarios de los usuarios. Este enfoque dinámico fomenta una cultura de responsabilidad y asegura una alineación continua con los estándares éticos.

2. Responsabilidad a lo Largo del Ciclo de Vida

Para operacionalizar eficazmente las consideraciones éticas, se aboga por una visión del ciclo de vida de la responsabilidad. Esta perspectiva requiere la integración de normas éticas desde la etapa inicial de ideación hasta la implementación posterior al mercado, abarcando auditorías éticas regulares y el establecimiento de mecanismos para monitorear la adherencia a estos principios.

Al incorporar la ética en los procesos organizacionales estándar, los desarrolladores pueden asegurar que la responsabilidad no sea simplemente un ejercicio de cumplimiento, sino un aspecto fundamental de la IA confiable.

3. Ética por Diseño

Construir sistemas de IA con consideraciones éticas desde el principio es crucial. La ética proactiva puede prevenir posibles escollos al asegurar que los problemas éticos se aborden durante la fase de diseño en lugar de como una reflexión posterior. Esto incluye buscar activamente comentarios de los usuarios para identificar y rectificar sesgos o deficiencias que puedan surgir durante el desarrollo.

4. Enfoques Mixtos

Diversas teorías éticas, incluyendo el utilitarismo y la deontología, informan la creación de marcos éticos para la IA. Los enfoques mixtos que incorporan estudios empíricos para reflejar las decisiones morales humanas pueden mejorar estos marcos, aunque se debe tener cuidado para

evitar conclusiones engañosas. Al comprender la posición moral de las máquinas y sus capacidades relacionales, los desarrolladores pueden navegar mejor el panorama ético de la IA.

3/

Problemas éticos clave en la IA.

El rápido avance de la inteligencia artificial (IA) ha puesto de relieve una serie de problemas éticos que son críticos para su despliegue responsable. A medida que los sistemas de IA se vuelven cada vez más autónomos e integrados en varios sectores, surgen varias preocupaciones éticas clave, incluyendo el sesgo y la equidad, la privacidad, la responsabilidad y la transparencia:

1. Sesgo y Equidad

Los sistemas de IA pueden heredar, al igual que los humanos en el proceso de su educación, y amplificar involuntariamente los sesgos presentes en sus datos de entrenamiento, lo que lleva a resultados discriminatorios en aplicaciones como la contratación, los préstamos y la aplicación de la ley. Garantizar la equidad en los algoritmos de IA es primordial, ya que los juicios injustos pueden surgir de sesgos cognitivos que a menudo no son reconocidos por sus creadores humanos. Abordar estos sesgos requiere una **monitorización y validación continuas** de los modelos de IA para mantener su confiabilidad.

2. Privacidad

La dependencia de los sistemas de IA en grandes cantidades de datos personales plantea importantes preocupaciones sobre la privacidad. **Proteger la información sensible del uso indebido y el acceso no autorizado es fundamental**, lo

que requiere estrategias robustas de protección de datos, como la anonimización, el cifrado y controles de acceso estrictos. Además, el cumplimiento de regulaciones como el **Reglamento General de Protección de Datos (GDPR)** y la **Ley de Privacidad del Consumidor de California (CCPA)** es esencial para la gobernanza ética de la IA, requiriendo auditorías regulares y avisos de privacidad claros para los usuarios.

3. Responsabilidad

A medida que las tecnologías de IA evolucionan, la cuestión de la responsabilidad se vuelve cada vez más compleja. La opacidad de muchos modelos de IA complica los esfuerzos para garantizar la transparencia y la equidad en sus procesos de toma de decisiones. Esto plantea preocupaciones sobre quién es responsable de las acciones de los sistemas autónomos, particularmente en áreas de alto riesgo como la atención médica y los vehículos autónomos, donde la seguridad y la confiabilidad son primordiales. **Los marcos éticos que definen la responsabilidad y la rendición de cuentas en la IA son necesarios para navegar estos desafíos.**

4. Transparencia

La transparencia en la IA implica hacer que los procesos de toma de decisiones de los sistemas de IA sean comprensibles para diversas partes interesadas. **Desarrollar políticas accionables que promuevan la transparencia es crucial para fomentar la confianza** y mitigar las preocupaciones éticas asociadas con el uso de la IA. Participar en investigaciones y consultas exhaustivas puede ayudar a las organizaciones a diseñar tecnologías de IA que respeten los valores humanos y promuevan estándares éticos.

4/

Gobernanza y regulación.

La gobernanza y regulación de la inteligencia artificial (IA) son áreas de preocupación cada vez más reconocidas a medida que la tecnología continúa avanzando e integrándose en varios sectores.

Los esfuerzos regulatorios se están implementando a nivel mundial, con iniciativas notables en regiones como la Unión Europea (UE), Brasil y Japón.

Estos esfuerzos tienen como objetivo aclarar y unificar las diversas corrientes de gobernanza de la IA, aunque su efectividad depende en gran medida de los mecanismos prácticos de implementación y aplicación:

1. Desafíos en la Aplicación

Uno de los principales desafíos en la aplicación de la regulación de la IA radica en identificar las entidades legales responsables y probar las reclamaciones en su contra. Las víctimas de posibles violaciones deben navegar por sistemas legales complejos para establecer la **responsabilidad, la intención y encontrar tribunales competentes para hacer cumplir las decisiones.**

En muchos casos, las protecciones legales bien establecidas, como los derechos del consumidor y los derechos de propiedad intelectual, se aplican de manera inadecuada en el ámbito digital, lo que lleva a una situación en la que las empresas tecnológicas pueden operar con una sensación de inmunidad mientras defienden robustamente su propiedad intelectual.

Esto ha fomentado una cultura a menudo descrita como “**Libertarismo de Internet**”, donde prevalece la creencia de que las soluciones tecnológicas pueden resolver de manera independiente los problemas sociales.

2. Implementación de la Regulación

La implementación real de las regulaciones de la IA plantea dificultades adicionales, ya que puede desafiar los modelos de negocio y prácticas existentes. Existe una tendencia entre ciertas partes interesadas, como las empresas y las organizaciones militares, a participar en “**lavado de ética**”, lo que les permite mantener una imagen pública favorable mientras continúan con prácticas potencialmente dañinas. En consecuencia, los desarrollos de políticas reales pueden enfrentar resistencia que prefieren el **statu quo** sobre las restricciones legalmente vinculantes.

En la UE, por ejemplo, la propuesta de la Ley de IA de la UE ha provocado llamadas a refinamientos para asegurar su viabilidad y aplicabilidad. La ley está diseñada para establecer un marco legal que enfatice las consideraciones éticas en el despliegue de la IA, pero también subraya las complejidades involucradas en traducir teorías éticas en políticas accionables.

3. Direcciones Futuras e Involucramiento de las Partes Interesadas

Existe un reconocimiento creciente de que **la gobernanza efectiva requiere más que solo marcos legales**; necesita un compromiso de todas las partes interesadas para priorizar las consideraciones éticas en el desarrollo y despliegue de la IA.

La regulación gubernamental se ve como un factor clave para fomentar este compromiso, ya que puede incentivar a las empresas a priorizar los problemas éticos y apoyar a los profesionales de la ética de la IA en su trabajo.

Además, los esfuerzos legislativos, como la Ley de Divulgación de Inteligencia Artificial Política (PAID) en los Estados Unidos, tienen como objetivo aumentar la transparencia sobre el uso de la IA en las comunicaciones políticas, destacando la necesidad de regulaciones específicas de dominio que aborden los desafíos únicos que plantean las tecnologías de IA.

A medida que el panorama de la IA continúa evolucionando, la interacción entre la ética y la gobernanza seguirá siendo un área vital de enfoque. Será crucial para los reguladores no solo reaccionar a los desafíos emergentes, sino también establecer proactivamente marcos que fomenten la innovación responsable y mitiguen los posibles riesgos asociados con la IA.

5/

Casos de estudio.

La exploración de consideraciones éticas en la inteligencia artificial se informa significativamente a través de estudios de caso del mundo real que ilustran las complejas interacciones entre la tecnología y los valores sociales.

Los Diálogos de Princeton sobre IA y Ética han desarrollado una serie de estudios de caso ficticios diseñados para provocar reflexión y discusión sobre varios dilemas éticos que surgen de la integración de la IA en la sociedad.

Estos estudios de caso fueron elaborados a través de una colaboración entre el **Centro Universitario de Valores Humanos (UCHV)** y el **Centro de Política de Tecnología de la Información (CITP)** en la Universidad de Princeton, comenzando en 2017-18.

1. Visión General de los Estudios de Caso

Los estudios de caso publicados por los Diálogos de Princeton abarcan seis análisis exhaustivos que abordan cuestiones apremiantes en la intersección de la IA, la ética y la sociedad.

Cada estudio aborda un conjunto distinto de desafíos éticos, promoviendo una comprensión de los compromisos morales y prácticos que son inherentes al despliegue de sistemas de IA.

2. Resúmenes de los Estudios de Caso

- **Aplicación de Salud Automatizada:** Este estudio de caso plantea preocupaciones sobre la legitimidad, el paternalismo, la transparencia, la censura y la desigualdad en el contexto de la tecnología de la salud.
- **Identificación Dinámica de Sonidos:** Centrado en cuestiones como los derechos, los daños representacionales, la neutralidad y la responsabilidad descendente, este estudio examina las implicaciones de las tecnologías de identificación de sonidos.
- **Optimización de Escuelas:** Abordando la privacidad, la autonomía, el consecuencialismo y la retórica, este caso profundiza en el uso de la IA en entornos educativos y su impacto en la dinámica entre estudiantes y profesores.
- **Chatbots de Aplicación de la Ley:** Este estudio discute las implicaciones éticas que rodean la automatización, la ética de la investigación y la soberanía en las aplicaciones de la ley.
- **Contratación por Máquina:** Destacando los desafíos del sesgo y la discriminación, este caso explora el impacto de la IA en los procesos de reclutamiento y las preocupaciones éticas asociadas.

6/

Direcciones futuras.

Las consideraciones éticas que rodean la inteligencia artificial (IA) están evolucionando rápidamente, lo que requiere un enfoque prospectivo para abordar los desafíos y oportunidades futuros. A medida que la tecnología continúa avanzando, los académicos y legisladores enfatizan la importancia de medidas proactivas para

LH n.340

anticipar y mitigar los posibles riesgos asociados con los sistemas de IA.

1. Marcos Éticos y Legislación

Un aspecto clave de las direcciones futuras en la ética de la IA implica el desarrollo de marcos integrales que puedan guiar las prácticas éticas en el despliegue de tecnologías de IA. Estos marcos deben abordar las implicaciones multifacéticas de la IA, incluyendo la equidad, la responsabilidad y la transparencia. Luciano Floridi, una figura destacada en el discurso, aboga por un trabajo pionero que informe de mejores políticas y regulaciones, con el objetivo de abordar preventivamente los dilemas éticos que plantean las tecnologías emergentes. Existe un consenso creciente entre los líderes de la industria y los legisladores de que definir objetivos políticos claros desde el principio es esencial para evitar los escollos experimentados con la regulación de las redes sociales. Este enfoque proactivo es crítico para asegurar que el desarrollo de la IA se alinee con los valores sociales y los estándares éticos, lo que puede ayudar a construir la confianza pública en los sistemas de IA.

2. Incorporación de Perspectivas de las Partes Interesadas

Involucrar a diversas partes interesadas es crucial para el desarrollo ético de la IA. Es esencial hacer que los conocimientos sean accesibles para aquellos que pueden verse afectados por las decisiones de la IA, incluidas las comunidades marginadas y las poblaciones vulnerables. Además, incorporar los comentarios de los usuarios en las fases de diseño e implementación puede ayudar a identificar sesgos y preocupaciones éticas que los desarrolle.

3. Los desafíos éticos de la inteligencia artificial (IA)

Los desafíos éticos de la inteligencia artificial (IA) son numerosos y complejos. Se añade un resumen de los principales desafíos éticos que enfrenta la IA:

- **Sesgo y Equidad:** Los sistemas de IA pueden heredar y amplificar los sesgos presentes en los datos de entrenamiento, lo que lleva a resultados discriminatorios en aplicaciones como la contratación, los préstamos y la aplicación de la ley. Es crucial garantizar la equidad en los algoritmos de IA, ya que los juicios injustos pueden surgir de sesgos cognitivos que a menudo no son reconocidos por sus creadores humanos.
- **Privacidad:** La dependencia de los sistemas de IA en grandes cantidades de datos personales plantea importantes preocupaciones sobre la privacidad. Proteger la información sensible del uso indebido y el acceso no autorizado es fundamental, lo que requiere estrategias robustas de protección de datos, como la anonimización, el cifrado y controles de acceso estrictos.
- **Responsabilidad:** A medida que las tecnologías de IA evolucionan, la cuestión de la responsabilidad se vuelve cada vez más compleja. La opacidad de muchos modelos de IA complica los esfuerzos para garantizar la transparencia y la equidad en sus procesos de toma de decisiones. Esto plantea preocupaciones sobre quién es responsable de las acciones de los sistemas autónomos, especialmente en áreas de alto riesgo como la atención médica y los vehículos autónomos.
- **Transparencia:** La transparencia en la IA implica hacer que los procesos de toma de decisiones de los sistemas de IA sean comprensibles para diversas partes interesadas. Desarrollar políticas accionables que promuevan la transparencia es crucial para fomentar la confianza y mitigar las preocupaciones éticas asociadas con el uso de la IA.

Estos desafíos subrayan la necesidad de un enfoque ético proactivo que informe el diseño, desarrollo y despliegue de los sistemas de IA desde su inicio.

Referencias

- ▶ **Aggarwal, V. (2023, Octubre 12)**
Senate Leaders Propose New Bipartisan Framework for AI Regulation
<https://jolt.law.harvard.edu/digest/senate-leaders-propose-new-bipartisan-framework-for-ai-regulation>
- ▶ **Ali, S.J., Christin, A., Smart, A., Katila, R. (2023, Diciembre 7)**
Policy Brief Walking the Walk of AI Ethics in Technology Companies.
Stanford University Human-Centered Artificial Intelligence <https://hai.stanford.edu/policy-brief-walking-walk-ai-ethics-technology-companies>
- ▶ **Cave, B.**
US state-by-state AI legislation snapshot.
BCLP <https://www.bclplaw.com/en-US/events-insights-news/us-state-by-state-artificial-intelligence-legislation-snapshot.html>
- ▶ **Cummings, M. (2024, Febrero 21)**
Uncovered, unknown, and uncertain: Guiding ethics in the age of AI.
Yale News.
<https://news.yale.edu/2024/02/21/uncovered-unknown-and-uncertain-guiding-ethics-age-ai>
- ▶ **Gelson Singh, M. (2023, Octubre 14)**
Case Studies in Ethical AI: Real-World Bias and Fairness. Medium.
<https://medium.com/@gelsonm/case-studies-in-ethical-ai-real-world-bias-and-fairness-d274c5c57fb5>
- ▶ **Gordon, J. S. y Nyholm, S. (2021)**
Ethics of Artificial Intelligence.
Internet Encyclopedia of Philosophy.
<https://philpapers.org/rec/AUTEOA>
<https://iep.utm.edu/ethics-of-artificial-intelligence/>
- ▶ **Guszcza, J. et al. (2020, Enero 28)**
Designing an AI ethics framework.
Deloitte Insights.
<https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/design-principles-ethical-artificial-intelligence.html>
- ▶ **Hohma, E. (2023, Octubre 2)**
A Practical Organizational Framework for AI Accountability.
European Commission.
<https://futurium.ec.europa.eu/en/european-ai-alliance/best-practices/practical-organizational-framework-ai-accountability>
- ▶ **IBM.** *What is AI Ethics?*
<https://www.ibm.com/topics/ai-ethics>
- ▶ **ISO.**
Building a responsible AI: How to manage the AI ethics debate.
<https://www.iso.org/artificial-intelligence/responsible-ai-ethics>
- ▶ **Kearns, M. y Roth, A. (2020, Enero 13)**
Ethical algorithm design should guide technology regulation.
Brookings. <https://www.brookings.edu/articles/ethical-algorithm-design-should-guide-technology-regulation/>
- ▶ **Müller, Vincent C. (2023, Otoño)**
Ethics of Artificial Intelligence and Robotics.
The Stanford Encyclopedia of Philosophy.
<https://plato.stanford.edu/archives/fall2023/entries/ethics-ai/>

LH n.340

▶ **Pazzanese, C. (2020, Octubre 26)**
Ethical concerns mount as AI takes bigger decision-making role.
 The Harvard Gazette.
<https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/>

▶ **Princeton Dialogues on AI and Ethics.**
Case Study PDFs
<https://aiethics.princeton.edu/case-studies/case-study-pdfs>

▶ *Stanford University.*
<https://storm.genie.stanford.edu/>

▶ **Stewart, K. (2024, Marzo 21)**
The ethical dilemmas of AI.
 USC Annenberg School for Communication and Journalism.
<https://annenberg.usc.edu/research/center-public-relations/usc-annenberg-relevance-report/ethical-dilemmas-ai>

▶ **Sullivan, M. (2023, Octubre 20)**
Key principles for ethical AI development.
 Transcend. <https://transcend.io/blog/ai-ethics>

▶ **The White House.**
Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence
<https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>

▶ **Tuzhilin, A. (2024, Agosto 31)**
The Ethical Challenges of AI.
 University of the People.
<https://www.uopeople.edu/blog/ethical-challenges-of-ai/>

▶ **West, D. M. (2018, Febrero 13)**
The role of corporations in addressing AI's ethical dilemmas.
 Brookings. <https://www.brookings.edu/articles/how-to-address-ai-ethical-dilemmas/>

