



UNIVERSIDAD
DE MÁLAGA

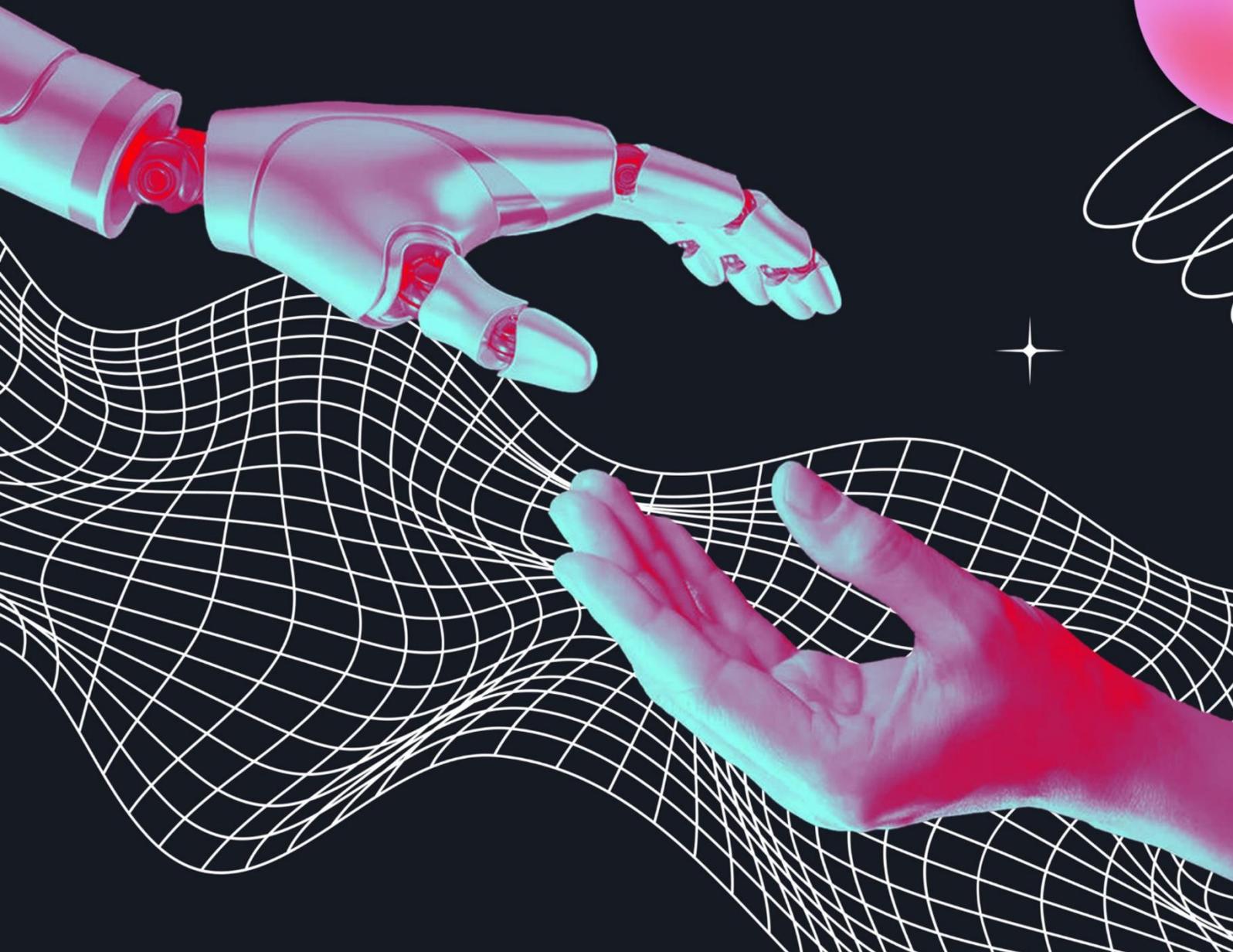


FACULTAD DE
MARKETING Y GESTIÓN

INFORME

ELABORACIÓN DE UN CÓDIGO ÉTICO PARA UNA IA RESPONSABLE

REALIZADO POR BE A LION | FEBRERO 2024



IMPULSORES:

BE A LION

COMITÉ ACADÉMICO CIENTÍFICO:

Francisco Triguero Ruiz

Carmen Jambrino Maldonado

María del Carmen del Vivero de Porras

Álvaro Díaz Casquero

Elena Fernández Díaz

Patricia P. Iglesias Sánchez

0. CONTEXTO

La Inteligencia Artificial (IA) ha irrumpido en múltiples sectores de la sociedad, desde la salud hasta la economía, generando tanto promesas como preocupaciones debido a su potencial disruptivo y sus riesgos inherentes. Con el objetivo de promover un uso responsable de la IA y en beneficio del bien común, desde BE A LION junto con un comité de expertos de distintas áreas de la Universidad de Málaga se ha desarrollado un código ético que establece directrices claras para su diseño, desarrollo y aplicación en el ámbito de la publicidad en general y en la creación de influencers virtuales en particular.

En este contexto, se reconoce la necesidad de abordar los desafíos éticos asociados con la IA, así como de gestionar los riesgos de manera adecuada y proporcionada. Se destaca la importancia de centrar el desarrollo de la IA en el ser humano, aspirando a mejorar el bienestar y la libertad de las personas.

La Inteligencia Artificial Generativa (IAG) ha emergido como un campo especialmente disruptivo en la última década, transformando la producción de contenido creativo y la interacción humana. Sin embargo, enfrenta desafíos significativos relacionados con el **sesgo en los conjuntos** de datos utilizados para su entrenamiento, así como **preocupaciones sobre la difusión de desinformación y la manipulación de medios digitales**.

En resumen, el código ético para una IA responsable que a continuación vamos a presentar busca fomentar el compromiso de utilizar la IA en beneficio de la humanidad, gestionando sus riesgos y promoviendo la confianza en su desarrollo y aplicación. Al adoptar principios de IA digna de confianza, se aspira a desarrollar soluciones éticas y competitivas que impulsen el progreso socioeconómico de manera sostenible.

1. PUNTO DE PARTIDA

1.1 La situación de la IAG en el ámbito del marketing y la Comunicación, percepción e impacto en la audiencia.

La aplicación de la Inteligencia Artificial Generativa (IAG) en el ámbito del marketing y la publicidad ha revolucionado la manera en que las marcas interactúan con sus audiencias. La capacidad de la IAG para crear contenido altamente personalizado y relevante ha impulsado estrategias publicitarias más efectivas y dirigidas. La personalización de anuncios basada en datos, habilitada por la IAG, ha mejorado la experiencia del usuario y ha aumentado la tasa de conversión en diversas plataformas.

El uso de la IAG en la publicidad ha permitido a las marcas comprender mejor a sus consumidores, proporcionando mensajes y productos altamente personalizados. Los sistemas de recomendación basados en IAG, como los utilizados por plataformas de comercio electrónico, han mejorado la satisfacción del cliente al ofrecer productos y servicios adaptados a los gustos individuales. Esto ha llevado a una mayor fidelización de los clientes y a un mayor compromiso con las marcas.

En este contexto emergen los "influencers metahumanos" creados a través de la IAG como figuras influyentes en las redes sociales. Estos personajes digitales, generados por algoritmos, han ganado seguidores y han impactado las tendencias culturales y de consumo. Ejemplos notables incluyen a Lil Miquela, Shudu Gram y Alba Renai, quienes tienen millones de seguidores y colaboran con marcas reconocidas, desafiando la noción tradicional de influencia humana en las redes sociales.

De cara a profundizar en el impacto que supone la creación y el entrenamiento de una IA generativa en el área del marketing y de la comunicación, en cuanto a la relación con los consumidores y con la audiencia, se han realizado una serie de trabajos de campo que establecen 3 áreas de investigación y que van de lo más general a lo particular, los influencers virtuales, que son en esta ocasión el objeto de estudio:

- **Análisis de los modelos de lenguaje:** ChatGPT, Gemini (Bard) y Bing. Se ha interactuado con los tres sistemas con el fin de conocer cómo se están implementado las cuestiones éticas y morales para su interacción con los humanos. Asimismo, se ha preguntado sobre el marco de referencia para la selección de los principios fundamentales de aplicación, tanto legales como informes o instituciones que tienen autoridad y establecen recomendaciones y estándares en este sentido, como la UNESCO o la plataforma We Are Humans.

Respecto al análisis que se ha llevado a cabo interactuando con Gemini, ChatGPT y Bing, ante las preguntas sobre sus principios éticos, y realizando una comparativa de estos tres sistemas en base a sus respuestas, el único principio ético que coincide en los tres es el de "Seguridad". ChatGPT y Gemini son los que más coinciden en sus principios éticos, concretamente en: "Seguridad", "Limitaciones", "Precisión" y "Legibilidad".

- **Investigación sobre los principios éticos que guían las actuaciones tanto de influencers humanos como metahumanos.** Para llevar a cabo esta investigación se ha preguntado a una serie de influencers humanos y virtuales, con entrevistas personales a los primeros y preguntas a través de sus perfiles de Instagram a los segundos. Las entrevistas abordaban consideraciones morales, criterios para evaluar colaboraciones y casos de conflictos.

Entre las principales conclusiones obtenidas de estas entrevistas destacamos que todos los *influencers* coincidieron en destacar que cuidan y protegen que su labor de difusión en diferentes campañas nunca suponga un conflicto con sus valores que han ido desarrollando a lo largo de su vida por la educación que han recibido por parte de su familia, amistades y entorno. *Asimismo, añaden* no contar con herramientas ni conocimientos suficientes para en ocasiones evaluar las diferentes propuestas de campañas de comunicación que reciben y que por ello suelen confiar principalmente en el filtro que le realizan su **agencia de representación**, su **propia experiencia** con la marca a promocionar y en alguna ocasión pequeñas **investigaciones** que realizan en medios digitales para conocer más sobre el producto o empresa a recomendar.

- Por último y con el fin de comprender mejor a nuestra audiencia, se lleva a cabo un **estudio para investigar las preocupaciones, intereses y motivaciones tanto de la Generación X como de la Generación Z en relación con la familiarización con influencers, el conocimiento de influencers virtuales**, así como sus percepciones sobre la autenticidad, la confianza y la transmisión de valores y el compromiso generado por los contenidos creados por influencers virtuales. El estudio revela que el 93,6% de los participantes sigue influencers en redes sociales, pero solo el 68% conoce a influencers virtuales. Sin embargo, hay confusión entre influencers virtuales y humanos. Se encontraron motivos para no seguir a influencers virtuales, como la falta de credibilidad o el temor a la confusión entre la realidad y la ficción. Algunos participantes encuentran fascinante la existencia de influencers virtuales, pero otros expresan preocupación por la falta de autenticidad.

2. MARCO REGULATORIO

A medida que la IAG se integra en diversas industrias y su impacto se hace más evidente, los esfuerzos por establecer marcos regulatorios y normativos han ganado relevancia para abordar preocupaciones éticas, legales y de seguridad. Actualmente no existe un marco regulatorio específico para la actividad de los influencers virtuales, es por ello que se establecen una serie de recomendaciones *soft law*, una normativa aplicable de interés o *hard law* (es decir aquellas normativas existentes y aplicables a dicha actividad) y unas implicaciones jurídicas destacadas.

a. Principios éticos en el desarrollo y uso de la IA:

En el ámbito de las recomendaciones, los principios éticos en el desarrollo y uso de la Inteligencia Artificial (IA) incluyen:

- **Beneficio humano:** La IA debe ser utilizada para beneficiar a las personas y la sociedad en general.
- **Transparencia y explicabilidad:** Los sistemas de IA deben ser transparentes y explicables para los usuarios y las partes afectadas.
- **Justicia y equidad:** La IA debe desarrollarse de manera justa y equitativa, evitando sesgos y discriminación.
- **Privacidad y seguridad de los datos:** Deben establecerse medidas de seguridad adecuadas para proteger la privacidad y confidencialidad de la información.
- **Responsabilidad y rendición de cuentas:** Los desarrolladores, proveedores y usuarios de sistemas de IA deben asumir la responsabilidad de sus acciones y decisiones.
- **Evaluaciones de impacto ético:** Procesos para identificar y evaluar posibles impactos éticos y sociales de los sistemas de IA durante su desarrollo e implementación.

b. Condición de Alba Renai como IA generativa:

La condición de Alba Renai como IA generativa implica que, según la legislación europea, debe cumplir con ciertos criterios:

- **Definición de IA:** Se considera IA a los sistemas capaces de generar contenidos con cierto grado de autonomía, basándose en objetivos definidos por humanos.
- **Tipos de sistemas de IA:** Se distinguen entre sistemas basados en máquinas, sistemas de uso general y sistemas generativos, estos últimos dedicados a generar contenido complejo.
- **Obligaciones de transparencia:** Los proveedores de sistemas de IA generativos deben publicar resúmenes detallados de los datos de entrenamiento protegidos por derechos de autor.

c. Usuario de especial relevancia:

En el contexto de influencers, Alba Renai podría ser considerada un usuario de especial relevancia si cumple con ciertos requisitos:

- **Ingresos significativos:** Debe obtener ingresos significativos de su actividad en servicios de intercambio de vídeo.
- **Responsabilidad editorial:** Es responsable de los contenidos audiovisuales que comparte.
- **Audiencia significativa:** Debe tener un gran número de seguidores y publicar regularmente contenido.

d. Responsabilidad civil e Inteligencia Artificial:

En caso de daños causados por la IA, la responsabilidad recae en la persona o entidad titular del sistema IA causante del daño, siempre que exista una relación de causalidad y el causante sea culpable o responsable del daño.

e. Propiedad intelectual:

La normativa en desarrollo busca clarificar si las creaciones realizadas con IA son susceptibles de derechos de propiedad intelectual. Se trabaja en establecer obligaciones para los desarrolladores de sistemas de IA respecto a la documentación técnica y etiquetado de contenidos generados por IA.

f. Obligación de transparencia en el uso de la IA y los datos:

Se reconoce la importancia de la transparencia en el uso de la IA, incluyendo la divulgación de métodos de uso, protección de datos y naturaleza de la información recopilada. También se requiere informar a los usuarios cuando interactúan con contenido generado por IA y gestionar los datos personales de manera transparente y conforme al RGPD.

3. LOS CÓDIGOS DE CONDUCTA DE UNA INFLUENCER METAHUMANA.

La ética desempeña un papel fundamental en el desarrollo y la aplicación de la Inteligencia Artificial Generativa (IAG). Establecer principios éticos sólidos es crucial para garantizar que la IAG se utilice de manera responsable, equitativa y segura, abordando preocupaciones relacionadas con la transparencia, la equidad y la responsabilidad en su desarrollo y aplicación:

3.1. Consideraciones Generales:

Como cuestiones generales a considerar en la elaboración de un Código Ético para Alba Renai, es conveniente resaltar la ética en las distintas fases de entrenamiento de Alba.

a) Aspectos Básicos:

- o **Ética en la Creación:** Establecer estándares éticos al crear metahumanos para garantizar una representación justa y respetuosa de diversas identidades y culturas.

- **Responsabilidad Social:** Promover un compromiso social positivo a través de influencers metahumanos, evitando contenido perjudicial o controvertido.
- **Transparencia y Autenticidad:** Informar claramente cuando un influencer metahumano es creado digitalmente y no representa a una persona real.
- **Cumplimiento de Normativas Éticas:** Adherirse a normativas éticas y legales en la industria de influencers metahumanos.
- **Honestidad en la Promoción:** Promocionar productos o servicios de manera honesta y transparente.
- **Respeto a la Privacidad:** Proteger la privacidad de las personas y evitar la divulgación de información personal sin consentimiento explícito.
- **Promoción de Valores Positivos:** Fomentar valores como la inclusión, el respeto y la empatía en todas las interacciones.
- **Evitar Contenido Sensacionalista:** Evitar la creación de contenido sensacionalista que pueda generar polémica o contribuir a la desinformación.
- **Interacción Auténtica:** Fomentar interacciones auténticas con el público, respondiendo a comentarios y mensajes de manera genuina.
- **Sensibilidad Cultural:** Mostrar sensibilidad hacia diversas culturas, evitando estereotipos y representaciones inexactas.

b) Aspectos Adicionales:

- **Verificación de Fuentes:** Incorporar un sistema de verificación de fuentes automatizado para evaluar la credibilidad y confiabilidad de la información antes de ser utilizada en respuestas.
- **Referencias Transparentes:** Indicar claramente las fuentes utilizadas al proporcionar respuestas.
- **Actualización Continua:** Mantener actualizadas las fuentes utilizadas para brindar respuestas precisas.

c) Códigos Éticos:

- Códigos Éticos:
 - Transparencia y Explicabilidad.
 - Equidad y Ausencia de Sesgo.
 - Responsabilidad y Rendición de Cuentas.
 - Privacidad y Protección de Datos.
- Cuestiones de Moralidad:
 - Respeto por los Valores Humanos.
 - Neutralidad Moral.
 - Transparencia en la Toma de Decisiones Morales.
 - Personalización Ética.
 - Límites Éticos Claros.

- Supervisión Humana.
- Aprendizaje y Mejora Continua.
- Responsabilidad.

El desarrollo ético de la IAG requiere la colaboración entre diversos actores y la implementación de principios éticos mediante políticas y directrices concretas en todo el ciclo de vida de los sistemas de IAG. Esto incluye la creación de marcos normativos claros, la promoción de la educación y concienciación sobre ética en IAG, así como la realización de evaluaciones periódicas para garantizar el cumplimiento de los principios éticos establecidos.

3.2. Códigos de Conducta Generales:

- **Transparencia en la Generación de Contenido:** Garantizar la transparencia en la creación y origen de cada publicación metahumana. Origen de la Generación, Identificación de Contenido Generado por IA, Colaboraciones y Patrocinios, Desglose de Elementos Generados por IA.
- **Enfoque en la Privacidad:** Medidas para asegurar la privacidad de los usuarios al interactuar con influencers metahumanos.
- **Responsabilidad Social:** Integración de la responsabilidad social en las acciones y decisiones de influencers metahumanos.
- **Veracidad de la Información:** Protocolos para verificar y garantizar la veracidad de la información compartida por influencers.
- **No Discriminación:** Medidas para evitar la discriminación en el contenido generado y fomentar la diversidad e inclusión.

Cada sección plantea preguntas importantes sobre cómo se abordan estos temas en el contexto de los influencers metahumanos, desde la transparencia en la generación de contenido hasta la promoción de la diversidad y la inclusión, pasando por la protección de la privacidad de los usuarios y la verificación de la información compartida. El feedback de la comunidad se destaca como una parte integral del proceso de mejora continua en todos estos aspectos.

4. LOS PRINCIPIOS, PREVENCIÓN DE PRÁCTICAS INDIRECTAS Y RIESGOS PARA UNA INFLUENCER METAHUMANA.

Características específicas:

- **Influencer:**
 - **Ética de Marketing:** Características éticas en la promoción de productos y colaboraciones.
 - **Autenticidad y Confianza:** Importancia de la autenticidad y confianza como factores distintivos en influencers.
 - **Transparencia:** Definición de influencers como socio-técnicos con énfasis en la transparencia.
 - **Efectividad de Influencers Humanos:** Comparativa de la efectividad entre influencers humanos y virtuales.
 - **Virtudes Éticas:** Análisis de virtudes éticas en actores virtuales en plataformas sociales.

GLOSARIO

Prompt: Los prompts son las frases que debemos usar para comunicarnos, preguntar, ordenar o proponer cosas a la IA.

IA: Simulación de la inteligencia humana a través de la programación de máquinas que la imitan, reproduciendo de esta manera acciones propias como el aprendizaje o la resolución de problemas. Ej.: LuzIA.

Influencer virtual: Personaje digital creado en software de gráficos por computadora, al que se dota de una personalidad definida por una visión del mundo en primera persona, la cual comparte en las plataformas sociales, y donde desarrolla su actividad, como si de un ser humano se tratase.

Vtuber: Influencer virtual – mayoritariamente, personajes de anime – que transmite su vida en vivo en plataformas de vídeo populares como Twitch o YouTube, gracias a la tecnología de captura de movimiento.

Avatar: Icono o figura que representa a una persona en particular en universos digitales (videojuegos, foros de Internet, RR.SS., eventos, etc.).

Metahumano: Representación virtual de una persona ya sea real o imaginaria.

IA Generativa: Sistema de inteligencia artificial capaz de generar textos, imágenes, vídeo u otros materiales a través de un comando o prompt.

Algoritmo: Conjunto de instrucciones predefinidas que se utilizan para realizar una tarea de manera determinada.

Modelo Transformer: Red neuronal que aprende contexto y, por lo tanto, significado mediante el seguimiento de relaciones en datos secuenciales como las palabras de una oración.



UNIVERSIDAD
DE MÁLAGA



FACULTAD DE
MARKETING Y GESTIÓN

INFORME

ELABORACIÓN DE UN CÓDIGO ÉTICO PARA UNA IA RESPONSABLE

REALIZADO POR BE A LION | FEBRERO 2024

