Razón y Palabra
Vol. 28 - Nº 119
Enero - Abril 2024

Los problemas tecno-éticos en la inteligencia artificial generativa

The Techno-Ethical Issues in Generative Artificial Intelligence

Manuel Alejandro Gutiérrez González
Escuela de Humanidades, Universidad Anáhuac Querétaro
E-mail: **manuel.gutierrezgon@anahuac.mx**
ORCID: **https://orcid.org/0000-0002-0799-5421**

Andrés Ocádiz Amador
Escuela de Humanidades, Universidad Anáhuac Querétaro
E-mail: **andres.ocadiz@anahuac.mx**
ORCID: **https://orcid.org/0009-0007-4354-9561**

DOI: 10.26807/rp.v28i119.2106

Fecha aceptación: 16/04/2024
Fecha publicación: 30/04/2024

Resumen

La inteligencia artificial tiene un gran impacto en la actualidad, tanto en la vida personal como en la social. El objetivo principal de esta investigación es identificar los retos específicos que plantea la Inteligencia Artificial Generativa en el ámbito de la educación, así como una reflexión sobre el discernimiento ético en el uso de estas tecnología emergentes a los cuales nos están llevando a futuros digitales. Para poder llegar a este punto, la metodología que empleamos es el modelo realista con un corte naturalista metodológico moderado, esto significa que la realidad puede ser conocida por cualquier método, tiene implicaciones normativas y existe una independencia del agente moral para alcanzar la objetividad. En concreto, la perspectiva ética que emplearemos tiene, como elemento natural del ser humano, el rostro como condición de posibilidad de la moralidad gracias a una perspectiva de segunda persona. Es por ello que el rostro es considerado como un punto de partida objetivo. Encontramos que si no existe un rostro en el uso de las inteligencias artificiales generativas, esto lleva a un desconocimiento de su funcionamiento y dificulta la vivencia moral; además, impide el florecimiento a través del desarrollo de virtudes intelectuales. Por eso, consideramos que es necesaria la alfabetización digital, para lograr el desarrollo pleno del alumno.

Palabras clave: inteligencia artificial generativa, alfabetización digital, rostro, perspectiva de segunda persona, bien común

e-ISSN: 1605 -4806

Abstract

Artificial intelligence has a great impact nowadays, both in personal and social life. The main objective of this research is to identify the specific challenges posed by Generative Artificial Intelligence in the field of education, as well as a reflection on the ethical discernment in the use of these emerging technologies that are leading us to digital futures. In order to reach this point, the methodology we employ is the realist model with a moderate methodological naturalistic style, this means that reality can be known by any method, it has normative implications and there is an independence of the moral agent to reach objectivity. Specifically, the ethical perspective that we will employ has, as a natural element of the human being, the face as a condition of possibility of morality thanks to a second person perspective. That is why the face is considered as an objective starting point. We find that if there is no face in the use of generative artificial intelligences, this leads to a lack of knowledge of their functioning and hinders moral experience; moreover, it prevents the flourishing through the development of intellectual virtues. Therefore, we believe that digital literacy is necessary to achieve the full development of the student.

Keywords: generative artificial intelligence, digital literacy, face, second person relatedness, common good.

Resumo

A inteligência artificial tem atualmente um grande impacto, tanto na vida pessoal como social. O principal objetivo desta investigação é identificar os desafios específicos colocados pela Inteligência Artificial Generativa no domínio da educação, bem como uma reflexão sobre o discernimento ético na utilização destas tecnologias emergentes que nos conduzem a futuros digitais. Para chegar a este ponto, a metodologia que empregamos é o modelo realista com um estilo metodológico naturalista moderado, o que significa que a realidade pode ser conhecida por qualquer método, tem implicações normativas e há uma independência do agente moral para alcançar a objetividade. Especificamente, a perspetiva ética que iremos empregar tem, como elemento natural do ser humano, o rosto como condição de possibilidade da moralidade graças a uma perspetiva de segunda pessoa. É por isso que o rosto é considerado como um ponto de partida objetivo. Constatamos que a ausência de rosto na utilização das inteligências artificiais generativas conduz a um desconhecimento do seu funcionamento e dificulta a experiência moral; além disso, impede o florescimento através do desenvolvimento das virtudes intelectuais. Por conseguinte, consideramos que a literacia digital é necessária para alcançar o pleno desenvolvimento do aluno.

Palavras-chave: inteligência artificial generativa, literacia digital, rosto, relação de segunda pessoa, bem comum.

1. Introducción

La tecnología y su uso no escapan de las regulaciones éticas. En 2004, la Unión Europea concebía que el uso de la tecnología podría hacer una reingeniería del cuerpo y de la mente humana; sin embargo, este tipo de implementación generaba problemas legales y filosóficos, pues significaba una pérdida de conocimiento, la rendición de la autonomía y la responsabilidad (European

Commission, Directorate-General for Research and Innovation, 2004, p. 32). En su informe final de 2016, la visión de la Unión Europa para el 2050 incluye el uso de la inteligencia artificial (IA) para reemplazar las actividades rutinarias del ser humano; sin embargo, menciona en esta visión futurista, que se eligirá al primer *cyborg* resolviendo un problema de justicia social (European Commission, Directorate-General for Research and Innovation, 2016).

A inicios de este siglo XXI ya encontramos algunas personas que tienen algunos implantes tecnológicos en su cuerpo para biomejorarlos, como son los casos de Neil Harbisson y el bio-hacker Ludo Disco Gamma Meow-Meow; sin embargo, todavía no existe un ser humano con biomejoramiento de las potencias intelectuales (mentales) gracias al uso de la IA. El argumento de la Unión Europea sobre el problema de la justicia social con los *cyborgs* no es un tema a futuro, sino presente. Llamas-Covarrubias (2020) menciona que la sociedad debe ir evolucionando y aceptar a esta "nueva especie humana" (conocida como transhumanismo), así como sus derechos transhumanos.

La IA tiene un gran impacto actualmente, tanto en la vida personal como en la social; sin embargo, en nuestra preocupación, le atribuimos grandes avances que todavía no tiene ni se ha desarrollado. Creemos que la IA piensa por sí misma, pero los desarrollos de esta tecnología apenas logran hacer lo que un bebé de 3 meses puede realizar (Dehaene, 2018).

Por estas razones nos proponemos, en este escrito, lo siguiente: en primer lugar, repasar los retos que supone el uso de la IA en general; en segundo lugar, identificar los retos específicos que plantea la Inteligencia Artificial Generativa en el ámbito de la educación, porque su irrupción en esta área no ha sido ni será neutral; en tercer lugar, hablaremos sobre una perspectiva moral sobre el uso de la IA en educación, pues como se ha hecho explícito con los informes de la Unión Europea, parece ser que los futuros digitales nos deben llevar a reflexionar sobre el discernimiento ético del uso de estas tecnologías emergentes. En concreto, la perspectiva ética que emplearemos tiene un elemento natural del ser humano como condición de posibilidad de la moralidad, el cual considera un punto de partida objetivo de ésta.

2. Los problemas que está generando la Inteligencia Artificial

La ética como disciplina filosófica y ciencia moral estudia y analiza los actos humanos conforme a su fin último. Si bien existen muchos modelos metaéticos, nuestro análisis parte de la metodología del modelo realista con un corte naturalista metodológico moderado. Esto significa que la realidad puede ser conocida a través de diferentes métodos (específicamente nos referimos a la realidad moral o los hechos morales), la cual tiene ciertas características que tienen algún tipo de normatividad; además, otra nota característica de estos hechos morales es que son independientes de la mente del agente moral (Lariguet, Yuan y Alles, 2023). Esto último, es de suma importancia, pues permite que existan verdades morales que no caen en un subjetivismo, sino en un objetivismo y que la verdad tiene un sentido de correspondencia (hecho moral y mente del agente moral) (David, 2022).

Si bien, los modelos éticos se dan en abundancia, Teixidó-Durán (2023) afirma que existen tres conjuntos de indagaciones éticas en los cuales se puede estudiar el problema de la IA, a saber, la nomoética, la bioética y la tecnoética

(p. 155). Los dos primeros conjuntos son más conocidos; por un lado, la nomoética se podría definir como una concepción ética legalista del deber o kantiana (Massini-Correas, 2019); por otro lado, la bioética es un estudio interdisciplinar (donde convergen la ética, la epistemología, la antropología, el derecho, la medicina, entre otros) para analizar y estudiar los problemas científicos relacionados con la vida humana (Lucas-Lucas, 2016). Podemos decir, a grandes rasgos, que la tecnoética es la reflexión moral que se hace sobre las aplicaciones tecnológicas, ya sea en el mismo ser humano o en las diferentes actividades humanas, como puede ser el caso de la comunicación, de la educación, de la economía, etc.

Los problemas que actualmente estamos presenciando con el uso de la IA se encuentran enmarcadas en estos tres conjuntos éticos; para el caso de la tecnoética queda claro que la IA es una tecnología que está irrumpiendo en nuestras actividades cotidianas, aunque algunos usuarios no sean conscientes de que están usándola o de algunas consecuencias de usarla. Sin embargo, para los otros dos conjuntos no queda tan claro el por qué se pueden enmarcar, pero esto será objeto de estudio en otros escritos. No obstante, podemos afirmar que empiezan a surgir cierto tipo de preocupaciones de regulaciones morales sobre cierto manejo de información privada y, todavía más haciendo énfasis en la bioética, en el uso de los datos biométricos.

Si bien, uno podría pensar que los problemas que se pueden generar con el uso de la IA son netamente individuales; empero, también encontramos problemas a nivel social. Encontramos varios usos maliciosos que se pueden realizar con el uso de la IA y que llevan a diferentes dilemas éticos: 1) sufrir un ataque insertando datos que llevan a cometer errores en procesos de aprendizaje; 2) clasificar erróneamente debido a los sistemas de aprendizaje automático; 3) generación de textos, imágenes, videos y audios para hacerse pasar por otras personas; 4) seguridad física; 5) seguridad digital; 6) seguridad política; 7) robo de identidad; 8) imposibilitar la empatía a la hora de la toma de decisiones (González-Arencibia y Martínez-Cardero, 2020). Es por ello que algunas investigaciones van encaminadas a generar una gobernanza de estas tecnologías que introdujo la cuarta revolución industrial (Landa-Arroyo, 2021).

Estos dilemas éticos tienen una repercusión a nivel global; sin embargo, ciertas derivaciones de la IA llevan a dilemas morales más específicos, como es el caso de la IA Generativa y sus usos en la comunidad educativa y en las comunicaciones.

3. Inteligencia Artificial Generativa

La IA cuenta actualmente con diversas ramas, una de ellas es la Inteligencia Artificial Generativa (IAG). Suele ser definida como la IA capaz de generar contenido original de forma autónoma a partir de instrucciones o *prompts* definidos por el usuario (Gutiérrez, 2023). La IAG tiene su base en una cantidad masiva de información y datos que fueron utilizados para entrenarla y, de esa manera, poder producir textos, imágenes, códigos o videos, entre otras cosas (Fleckenstein, Meyer, Jansen, Keller, Köller y Möller, 2024). Algunos ejemplos de IAG son ChatGPT, Midjourney o Microsoft Copilot (Chiu, 2024). La IAG más popular, al momento de redactar este texto, es ChatGPT, una IAG del ramo de los modelos de lenguaje (*Aligning Language Models to Follow Instructions*, s/f) que se caracteriza por su capacidad de generar textos originales imitando el lenguaje humano. De este modo, es capaz de redactar textos según las

instrucciones dadas, responder preguntas, generar códigos de programación o realizar traducciones entre idiomas (Fleckenstein *et al*., 2024).

Más que hablar de una introducción de las IAG en el ámbito educativo, se debería hablar de una irrupción (Chiu, 2024), ya que no fue algo deseado ni mucho menos planeado. Las plataformas de IAG son tan accesibles, incluso gratuitas, que los alumnos empleaban las IAG antes de que los docentes pudieran integrarlas en sus planeaciones y estrategias didácticas. Ante esta irrupción, en la mayoría de los centros educativos comenzaron a darse las prohibiciones debido a que la preocupación principal era que el aprendizaje se obstaculizaría, pues los alumnos recurrirían a las IAG para la elaboración de tareas, trabajos y actividades. Básicamente, la principal preocupación era la evaluación del aprendizaje (Chiu, 2024). Esta preocupación es más que válida, pues el estudio de Fleckenstein *et al*. (2024) demuestra que los docentes, aún hoy, siguen siendo incapaces de identificar correctamente si un texto ha sido redactado por una IAG o por un estudiante real. Por lo tanto, la pregunta de fondo en esta primera etapa fue: ¿cómo evaluar el aprendizaje de un estudiante si no se sabe si realizó él mismo la actividad o la realizó con ayuda de una IAG?

Posterior a la ola de prohibiciones de la IAG en las instituciones educativas, vino la necesaria reflexión por parte de los docentes y educadores donde se pusieron en la balanza los pros y contras de utilizar las IAG en la educación. Fruto de estas investigaciones se ha podido determinar que las IAG no solamente tienen un impacto en el proceso de enseñanza-aprendizaje, sino también en procesos de evaluación y en procesos administrativos (Chiu, 2024). Esto quiere decir que no solo los alumnos pueden beneficiarse de la utilización de las IAG, sino también los docentes e incluso los mismos centros educativos como conjunto.

Sin embargo, las dudas y los retos se siguen multiplicando en torno al empleo de las IAG en la educación. De entrada, la pregunta obligada es: ¿qué nuevos problemas genera la IAG tras resolver el problema original? (Schuurman, 2019). En el estudio cualitativo de Chiu (2024), los mismos estudiantes reconocieron que, ahora que utilizan de manera más o menos regular las IAG, requieren de unas habilidades de las que antes no eran conscientes: la habilidad de generar *prompts*, la alfabetización digital y, sobre todo, los fundamentos éticos. En el presente escrito, queremos poner el foco de atención en la alfabetización digital, pues su presencia o carencia determinará los fundamentos éticos que serán necesarios.

Se entiende por alfabetización digital “la habilidad para acceder, gestionar, entender, comunicar, evaluar e integrar información de manera segura y apropiada a través de tecnologías digitales” (Law, Woo, de la Torre, y Wong, 2018, p. 6). Esta habilidad es vista por Buchan, Bhawra y Katapally (2024) como crucial y esencial, ya que sin esta habilidad los jóvenes son más vulnerables a situaciones de desinformación, seguridad informática o problemas con la privacidad. Se trata de algo contraintuitivo, ya que los jóvenes interactúan todo el tiempo con dispositivos digitales conectados a la red, por lo que deberían ser más precavidos. Sin embargo, el hecho que las personas tengan habilidades digitales (sepan utilizar dispositivos digitales) no asegura que las sepan utilizar de manera eficaz en lo que respecta al aprendizaje, la seguridad y la ética (Getenet, Cantle, Redmond y Albion, 2024).

La alfabetización digital, de la que los estudiantes reconocieron estar carentes, está conformada por varios elementos. Siguiendo la investigación de Buchan

et al. (2024) consideraremos los siguientes cuatro: 1) fluidez digital; 2) seguridad digital y privacidad; 3) ética y empatía; y 4) sensibilización del usuario.

Se entiende por fluidez digital la consulta de información a través de la web, es decir, que las personas sepan buscar la información utilizando las categorías adecuadas de manera específica y sistemática, de manera que la búsqueda de información sea eficaz. De igual manera, la fluidez digital implica la habilidad de identificar, evaluar y autenticar tanto la confiabilidad de las fuentes como la veracidad de la información (Buchan *et al*., 2024). Prácticamente, todos los usuarios de internet saben buscar información de manera casual, pero cuando se trata de hacer investigación más seria o formal no todos los usuarios cuentan con esta fluidez.

Cuando se habla de seguridad digital y privacidad se está tocando uno de los puntos álgidos de las actividades que se realizan en la web. Lo principal en esta habilidad es que el usuario comprenda el concepto de privacidad, pues no es tan obvio como parece (Buchan *et a*l., 2024). De entrada, el usuario siempre debe ser consciente de la información que se le solicita en los muchos sitios web que consulta y apps que utiliza. En algunas ocasiones se trata de su dirección IP, en otras se trata de su nombre, y a veces es su ubicación. No todos los sitios web y todas las apps solicitan la misma información o mismos permisos, por lo que el usuario debería ser capaz de identificar qué información es la que se le está solicitando y evaluar si es pertinente brindar esa información. Sumado a esto, la protección de la privacidad debe llevar al usuario a ser consciente de la necesidad de tener contraseñas seguras y diversificadas limitando el acceso a la información que comparte, para evitar, entre otras cosas, un posible robo de identidad (Buchan *et al*., 2024).

En tercer lugar, Buchan *et al*. (2024) identifican la ética y la empatía como componentes esenciales de la alfabetización digital. Estos componentes tienen dos orientaciones. La primera de ellas se orienta hacia el usuario mismo, es decir, se trata de la responsabilidad que tiene el usuario de su propio comportamiento para evitar situaciones indeseables como lo puede ser el ciberacoso; pero también, implica la responsabilidad ética al saber que toda la información que suben o comparten y toda la actividad que realizan en la red deja una huella digital que puede ser rastreada por personas que realizan actividades delictivas como pueden ser chantajes y amenazas. La segunda orientación del componente ético se dirige hacia los demás y se manifiesta principalmente en lo relacionado con los derechos de autor y el plagio. Esta segunda orientación busca que el usuario comprenda la importancia de citar todas las fuentes que utiliza en sus búsquedas y por qué el no hacerlo es considerado plagio. Esta necesidad es tan urgente que ha habido algunas propuestas como la de Dutceac Segesten, Larsson, Åström y Aits (2023) de elaborar un programa de doctorado en IA donde se estudien, entre otras cosas, los fundamentos éticos del uso de la IA con la intención de de asegurar sus beneficios y reducir sus efectos negativos.

Aunque Buchan *et al*. (2024) se quedan en un nivel general, este aspecto ético del plagio es particularmente relevante cuando lo aplicamos a las IAG, pues una práctica común por parte de los alumnos consiste en solicitar a las IAG que les elaboren productos que luego entregarán en la escuela como si fueran hechos por ellos mismos (Abbas, Jam, y Khan, 2024). Esta situación tiene una particularidad, pues técnicamente no se está afectando la propiedad intelectual ni se está plagiando a nadie, pero es cuestionable cuando el alumno lo entrega como si fuera de su autoría.

Finalmente, el último elemento de la alfabetización digital es la sensibilización al usuario. Este componente se relaciona, principalmente, con la aceptación de los términos de servicio y la recolección de datos (Buchan *et al*., 2024). Es muy raro el usuario que dedica tiempo a leer los términos y condiciones que las *apps*, redes sociales y otros servicios le piden aceptar. En dichos contratos se especifica quién será el dueño de la información que se suba a la red, quién puede hacer uso de dicha información y a quién se puede compartir o vender. Si el usuario no leyó ese contrato, no sabrá qué es lo que está autorizando. Por otro lado, el usuario debería sensibilizarse en lo relativo a la recolección de datos y el para qué se recolectan. Principalmente, la recolección de datos tiene como finalidad ofrecer mejores experiencias durante la navegación del usuario por la red, sugiriendo información o productos afines a sus gustos y pensamiento (Buchan *et al*., 2024). Este último punto es particularmente relevante al hablar de las IAG, ya que éstas han sido entrenadas con cantidades masivas de información disponible en la web, es decir, las IAG han sido entrenadas también con los datos que se han recolectado de cada usuario de la red (*Research*, s/f). Más aún, ChatGPT está continuamente "aprendiendo" del usuario preferencias, modos de redactar, vocabulario típico, etc., a partir de los *prompts* que le son dados.

El estudio de Buchan *et al*. (2024) muestra la generalidad de las carencias de alfabetización digital de los usuarios, principalmente los más jóvenes. Sin embargo, cuando estas carencias se observan en el uso de las IAG, se vuelven más preocupantes, ya que la mayoría de los estudiantes utilizan las IAG para fines escolares y las usan, en general, de manera eficaz. No obstante, el estudio de Chiu (2024) dejó en evidencia que los estudiantes utilizan las IAG sin saber cómo funcionan, con qué información fueron entrenadas y cómo procesan la información que reciben. Este déficit conduce a dos situaciones concretas: la primera de ellas tiene que ver, nuevamente, con la vulnerabilidad a la desinformación, pues los estudiantes asumen que la información que reciben de las IAG es siempre verdadera; incluso, en el caso que duden de la veracidad de la información arrojada por las IAG, su carencia de alfabetización digital les dificultará corroborar la información al no poder diferenciar las fuentes confiables de las que no lo son. Pero, la segunda situación es la más relevante para este estudio, pues el no entender cómo funciona la IAG y cómo se procesa la información conducen directamente a riesgos éticos que deben ser atendidos.

4. Perspectiva moral en la Inteligencia Artificial

Los problemas morales que vivimos en este siglo XXI tienen sus inicios con René Descartes, quien es considerado el padre de la Modernidad. En la primera de las *Meditaciones Metafísicas*, Descartes (2011) desarrolla un proyecto que había estado planeando desde mucho tiempo atrás: destruir en general todas las opiniones antiguas, tanto en el ámbito del conocimiento, como de la moralidad. Como es sabido, Descartes propone la duda como método de conocimiento y, con éste, llega a un punto donde no existe ninguna duda: el *cogito*. Este inicio, además de llevar a un subjetivismo (solipsismo) y a la autotrascendencia en sí mismo, también llevó a la despersonalización a través de eliminar el rostro de uno mismo y del otro (Medina-Delgadillo, 2010).

El rostro, dentro de la filosofía, será recuperado por el filósofo lituano Emmanuel Lévinas. Específicamente, Lévinas buscará desdecir lo que el idealismo y la fenomenología habían desarrollado y pretenderá decir en griego la novedad de lo hebreo (Medina-Delgadillo, 2017). Es decir, Lévinas volverá a centrar lo que la filosofía moderna des-ordenó: el orden de la relación del Yo-Tú. En efecto, para Lévinas la forma de conocimiento se puede dar en lo sensitivo, cognitivo y el recibimiento o la experiencia sensible del rostro; esta recepción del rostro permite la relación ética, puesto que responsabiliza al yo a través de la palabra, del discurso y de la dinámica de la mirada (Navarro, 2008).

Con esto, queremos decir la parte fundamental de todo nuestro escrito: el rostro es condición de moralidad. Retomando nuestra metodología, podemos decir que la propiedad natural por la cual emerge la posibilidad metodológica de hacer ética es el rostro del otro. En efecto, cuando nos encontramos rostro-a-rostro (cara-a-cara) se apunta a un acto de interpelación del otro, así como de que el "yo" sea interpelado. Gracias a este encuentro, también el "yo" es cuestionado, además la presencia del otro colma y goza al amor.

La relación Yo-Tú, estar en la presencia rostro-a-rostro, permite que el Yo no se encierre en sí mismo, sino que trascienda de sí y busque la alteridad. Una forma de trascender se da en la búsqueda, no del bien propio, sino del bien del otro. Así pues, la búsqueda y recognición del rostro del otro me posibilita a vivir conforme al bien. Esta propuesta Levinasiana rompe con la perspectiva de primera persona (PPP) y la perspectiva de tercera persona (PTP), y abre a la perspectiva de segunda persona (PSP). Estas perspectivas se pueden usar en diferentes disciplinas, la ética también tiene estas perspectivas: la PPP es una en la cual sólo se busca el bien individual, la PSP es la búsqueda del bien común (BC), y la PTP es la búsqueda del bien en un sentido general/universal o de manera teórica. ¿Desde qué perspectiva se están generando los códigos morales en la actualidad? Para el caso específico de nuestro tema, ¿cómo se están configurando los códigos morales con respecto a la IA?

Antes de responder a esta pregunta, queremos hacer una observación de la PSP con respecto a la empatía y el argumento de la analogía. Si nos quedamos en una PPP decimos que sólo tenemos acceso fenoménico a nuestra propia interioridad, es decir, a nuestra propia mente (pensamientos, intenciones, voliciones, etc.). ¿Cómo tener acceso a la mente del otro? Ya Lévinas nos ha dicho que es gracias al rostro, pero ¿cómo la percepción del rostro del otro nos brinda información sobre su mente? Aquí es donde entra el argumento de la analogía: el "yo" se puede dar cuenta de sí mismo, se puede percibir en su corporalidad y hay ciertas experiencias que tiene y que siente en sí mismo a través de su cuerpo, las cuales le causan ciertas acciones. Por ejemplo, puede tener la experiencia de caminar descalzo y golpearse el dedo chiquito del pie derecho con la pata de una mesa; esta experiencia causa en sí mismo un dolor en todo su ser. Ahora bien, dado que el "yo" experimenta esto en sí mismo, puede inferir que sucede lo mismo en los cuerpos de los otros, pues también están influidos y actúan de maneras similares. Así, por analogía, se infiere que cuando alguien tiene una experiencia similar a la del "yo", el cuerpo del otro va a reaccionar de la misma manera que el "yo". Si bien, Scheler y Sartre estarían en desacuerdo con esta explicación, porque no podemos dividir el fenómeno en un aspecto psicológico y un aspecto de comportamiento (el cuerpo y la *psiqué* son una unidad expresiva donde se nos da en un contexto significativo), esta explicación manifiesta, de alguna manera, un modo de la conciencia que se denomina empatía: un acto intencional que nos descubre la experiencia vivida del otro (Gallagher y Zahavi, 2014, 268-273).

La propuesta de conformar una ética desde la PSP parece plausible; sin embargo, existen algunos códigos éticos que han tratado de conciliar el tema de la IA con el BC. Berendt (2019) menciona que el BC es un objetivo de la IA, el problema de esta propuesta es que pretende supeditar el BC a la IA; cuando lo que debe suceder, más bien, es que el BC gobierne a la IA. Berendt (2019) muestra tres códigos de ética, pero en este escrito sólo mencionaremos dos por sus referencias al BC: 1) el código de ética de la *Association for Computing Machinery* (ACM) que sólo busca la contribución al bienestar de las personas; y, 2) dentro de los Principios de Asilomar, encontramos el principio de BC, el cual no es definido, pero dice que la superinteligencia se debe desarrollar "al servicio de ideales éticos ampliamente compartidos" (Berendt, 2019, 45) y que generen un beneficio para la sociedad o institución. Lo que se quiere mostrar con estos dos ejemplos es que, en realidad, estos códigos de ética para la IA no tienen como objetivo el BC, pues en el código de ACM sólo habla de bienestar (el cual es un elemento del BC, mas no el BC) y el segundo menciona que está al servicio de ideales éticos.

Por otro lado, Berendt (2019) menciona que existen cuatro características de la IA que desafían al BC: 1) la mentalidad de resolución de problemas por parte de los ingenieros; con esto, Berendt (2019) quiere decir que es importante la cuestión de quién define el problema; 2) la integración de las partes interesadas, pues la unión de muchos tomadores de decisiones ha conseguido que la obtención de requisitos sea un estatus de obligación legal; 3) el papel del conocimiento; y, 4) efectos secundarios (Berendt, 2019, 50). Para superar estos cuatro desafíos al BC, propone nueve recomendaciones: estudiar las distintas caras del conocimiento, identificar e implicar a las partes interesadas, buscar efectos individuales y sociales, incluir el pensamiento de proporcionalidad, centrarse en sistemas completos, considerar bucles de retroalimentación, aprovechar el estado de la técnica, hacer preguntas y estar preparado para equivocarse (Berendt, 2019, 60-61).

Las propuestas antes mencionadas tienen un problema de raíz: la concepción que tienen sobre el BC. Berendt (2019) define el BC como "aquello que beneficia a una sociedad en su totalidad", siguiendo a Hussain (citado en Berendt, 2019) entiende por BC aquello como recursos naturales e instituciones creadas por el hombre, teniendo como más importantes las instituciones y las prácticas sociales (Berendt, 2019, 46). Además, entiende el beneficio de la utilidad individual o grupal en un sentido consecuencionalista. Si seguimos esta propuesta, podemos decir que un grupo de malhechores puede robar dinero usando de la aplicación (IA, entre otras tecnologías) una institución creada por el ser humano (un banco) para obtener un beneficio grupal (tener dinero de cuentahabientes), y que éstos trabajaron por el BC.

Por otra parte, Schuurman (2019) sugiere seis preguntas que propone Neil Postam para entender el impacto de la IA en nuestras vidas. Estas seis preguntas que plantea no se basan en cuestiones técnicas, sino para cubrir algunos sesgos introducidos por la tecnología: 1) ¿qué problema soluciona la IA?, 2) ¿de quién es el problema que está resolviendo la IA?, 3) ¿qué problemas creará la IA incluso cuando resuelva un problema?, 4) ¿qué personas o instituciones se verán perjudicadas por el IA?, 5) ¿qué cambios de lenguaje está forzando la IA?, 6) ¿qué tipo de personas e instituciones obtienen un poder económico y político especial a través de la IA? (Schuurman, 2019, 78). También propone siete preguntas para comprender los problemas éticos que surgen del uso de la IA (Schuurman, 2019, 79). El problema es que en su artículo busca acercar el pensamiento filosófico y teológico y, de esta forma, contribuir al BC, pero en

ningún momento lo hace, pues la única acción que propone es que cristianos se unan a la conversación para así contribuir al BC.

Por último, Oliver (2018) ha mencionado que es necesaria la regulación y desarrollo de la IA con una perspectiva de BC. Para poder realizar esta regulación y desarrollo, propone cinco pilares: 1) autonomía (y dignidad), 2) justicia (y solidaridad; no discriminación; cooperación), 3) beneficiencia (sostenibilidad; veracidad; diversidad), 4) explicabilidad (transparencia; responsabilidad y el papel del humano) y 5) no maleficencia (fiabilidad y seguridad; reproducibilidad; prudencia; protección de datos y privacidad) (Oliver 2018, 42-49). Estos cinco pilares pueden ayudar a comprender la humanidad que se busca alcanzar con la IA, pero no cómo toda la comunidad política debe gobernar, distribuir (justicia) y mantener (estabilidad) los bienes comunes que genere la IA o a la misma IA.

Las propuestas de Berendt (2019), Schuurman (2019) y Oliver (2018) no llegan al problema de raíz: la ontologización, subjetivación o sustanciación de la IA. Esto quiere decir que cada vez que hablamos de la IA es como si fuera un sujeto con una mente fenomenológica que goza de inteligencia, pensamientos, voliciones, intencionalidad y, una característica más a destacar, un rostro. La ausencia del rostro del otro genera un infierno en la tierra y lleva a destruir el BC. Sedmak (2020) nota que el concepto de infierno significa lo que no podemos desear, lo que debemos evitar y de lo que debemos protegernos. Siguiendo a Macario, el Grande, Sedmak (2020) describe el infierno como "un lugar en el que las personas están atadas por la espalda y no pueden verse a los ojos. [...] No es posible conocer el rostro de los demás" (p. 51). Otras dos características que retoma Sedmak (2020), de Evagrio Póntico, es que el infierno en la tierra es «un lugar de quietud dolorosa, silencio doloroso, y un lugar de espera atormentada» (p. 52). Estas tres características nos llevan a la destrucción de la persona humana, pues quedan excluidas de la comunidad de fraternidad.

Dicho todo lo anterior, podemos comprender por qué uno de los dilemas éticos que está generando las IAG es la imposibilidad de empatía en la toma de decisiones (expuesto en el primer apartado de este escrito). Muchas personas que usan las IAG no pueden captar la moralidad de su uso, porque no hay un encuentro con un rostro en el cual los interpele a la responsabilidad. Esto se da de dos maneras: por un lado, aquellos que usan todos los datos obtenidos de las IAG no sienten el compromiso con el otro como un imperativo porque no lo conocen cara-a-cara; por el otro lado, los usuarios de estas tecnologías no conocen los límites morales a los que se están enfrentando, porque no pueden reconocer la intencionalidad del que se encuentra del otro lado a punto de beneficiarse por la explotación o venta sus datos informáticos. Es por ello que, cuidar de nuestra propia identidad personal es un deber moral, pues la persona humana no debe estar a disposición de proyectos tecnológicos, los cuales llevan a perder el significado de sí mismo (Sanguineti, 2021).

4. Conclusiones

En este estudio, se han identificado las nuevas necesidades que surgen de la utilización de las IAG en el ámbito educativo, siendo la más urgente la alfabetización digital, no en un sentido reducido a la eficacia sino en uno más amplio que considere el desarrollo integral del alumno. Una de las manifestaciones de la desalfabetización digital es el desconocimiento por parte

del alumno de cómo funcionan las IAG, quién está detrás del entrenamiento de las IAG, para qué se recolectan los datos, etc., es decir, el alumno no puede colocar un rostro a la IAG, un rostro que, como ya se dijo en el último apartado, es una condición que facilita la vivencia de lo moral.

Normalmente, las regulaciones éticas tienen como objetivo principal la normatividad; sin embargo, nuestra postura busca ampliar los alcances que se pueden obtener, partiendo de una PSP y, como consecuencia, del concepto de BC. El BC tiene un elemento teleológico que es el desarrollo humano integral o, dicho en términos aristotélicos, la *eudaimonia* (felicidad o florecimiento) a través de las virtudes. Baehr (2011) propone nuevas virtudes intelectuales, entre las que se pueden destacar: cuidado y minuciosidad en la investigación, capacidad de hacer buenas preguntas, imparcialidad, rigor, prudencia, pensamiento crítico y veracidad.

Si no hay rostro en las IAG por desconocimiento de su funcionamiento, se dificulta, por tanto, la moralidad. Asimismo, se dificulta la vivencia de las virtudes que acabamos de enlistar, impidiendo el pleno desarrollo de la *eudaimonia* en el alumno, es decir, su florecimiento integral como persona. Por las razones presentadas, consideramos que es necesaria la alfabetización digital, para lograr el desarrollo pleno del alumno.

Para lograr este objetivo, se pueden reconocer dos acciones necesarias. La primera de ellas es la necesidad de programas de capacitación en alfabetización digital, para que el estudiante sea consciente del funcionamiento de las IAG, con el objetivo de ponerles un rostro y facilitar, así, el proceso moral. La segunda acción es seguir ofreciendo estrategias para el desarrollo del pensamiento crítico de los estudiantes, como virtud intelectual. Siendo conscientes de que, alcanzando esta virtud, se puede vivir en la *eudaimonia*. La alfabetización digital, por tanto, tiene un alcance superior a una visión meramente tecnicista, ya que apunta a la plenitud y felicidad del estudiante.

Aunque no se desarrolló en este escrito, encontramos varias vetas para futuras investigaciones, por ejemplo, la necesidad de un modelo epistemológico sólido para que se puedan evaluar las fuentes de información de la web para diferenciar correctamente entre hechos y opiniones. Otro tema de interés, es la propuesta de la regulación moral del uso de este tipo de tecnologías con una PSP en el proceso de enseñanza-aprendizaje.

Referencias

Abbas, M., Jam, F. A., y Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, *21*, 10. https://doi.org/10.1186/s41239-024-00444-7

Aligning language models to follow instructions. (s/f). Recuperado el 29 de febrero de 2024, de https://openai.com/research/instruction-following

Baehr, Jason. *The inquiring mind: On intellectual virtues and virtue epistemology*. Oxford: Oxford University Press, 2011. https://doi.org/10.1093/acprof:oso/9780199604074.001.0001

Berendt, B. (2019). AI for the common good?! Pitfalls, challenges, and ethics Pen-Testing. *Paladyn. Journal of behavioral robotics, 10*(1), 44-65. https://doi.org/10.1515/pjbr-2019-0004

Buchan, M. C., Bhawra, J., y Katapally, T. R. (2024). Navigating the digital world: development of an evidence-based digital literacy program and assessment tool for youth. *Smart Learning Environments*, *11*, 8. https://doi.org/10.1186/s40561-024-00293-x

Chiu, T. K. F. (2024). Future research recommendations for transforming higher education with generative AI. *Computers and Education: Artificial Intelligence*, *6*, 100197. https://doi.org/10.1016/j.caeai.2023.100197

David, M. (Summer 2022 Edition). "The correspondence theory of truth", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2022/entries/truth-correspondence/>.

Descartes, R. (2011). *Meditaciones metafísicas*. Gredos.

Dehaene, S. (2018). *En busca de la mente. El largo camino de la ciencia para comprender la vida mental (y lo que aún queda por descubrir)*. Siglo Veintiuno Editores.

Dutceac Segesten, A., Larsson, S., Åström, K., y Aits, S. (2023). *A concept for interdisciplinary PhD education in artificial intelligence*. https://doi.org/10.5281/ZENODO.7576143

European Commission, Directorate-General for Research and Innovation, Nordmann, A. (2004). *Converging technologies: shaping the future of European societies*. Publications Office.

European Commission, Directorate-General for Research and Innovation, Nordmann, A. (2016). *Digital futures final report- A journey into 2050 visions and policy challenges*. Publications Office.

Fleckenstein, J., Meyer, J., Jansen, T., Keller, S. D., Köller, O., y Möller, J. (2024). Do teachers spot AI? Evaluating the detectability of AI-generated texts among student essays. *Computers and Education: Artificial Intelligence*, *6*. https://www.sciencedirect.com/science/article/pii/S2666920X24000109

Gallagher, S., y Zahavi, D. (2014). *La mente fenomenológica*. Alianza Editorial.

Getenet, S., Cantle, R., Redmond, P., y Albion, P. (2024). Students' digital technology attitude, literacy and self-efficacy and their effect on online learning engagement. *International Journal of Educational Technology in Higher Education*, *21*, 3. https://doi.org/10.1186/s41239-023-00437-y

González-Arencibia, M., y Martínez-Cardero, D. (2020). Dilemas éticos en el escenario de la inteligencia artificial. *Economía y Sociedad, 25*(57), 1-17. https://doi.org/10.15359/eys.25-57.5

Gutiérrez, K. (2023). Inteligencia artificial generativa: Irrupción y desafíos. *Revista Enfoques*, *4*(2), 57–82.

Landa-Arroyo, C. (2021). Constitución, derechos fundamentales, inteligencia artificial y algoritmos. *Themis. Revista de Derecho,* 79, 37-50. https://doi.org/10.18800/themis.202101.002

Lariguet, G., Yuan, M. S., y Alles, N. (2023). *La metaética puesta a punto*. Ediciones Universidad Nacional del Litoral.

Law, N., Woo, D., de la Torre, J., y Wong, G. (2018). *A global framework of reference on digital literacy skills for indicator 4.4.2*. UNESCO. https://uis.unesco.

org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf

Llamas-Covarrubias, J. Z. (2020). "Derechos humanos, transhumanismo y posthumanismo: una mejora tecnológica humana". *Derechos fundamentales a debate/ Comisión Estatal de Derechos Humanos Jalisco, 12*(1), 85-104.

Lucas-Lucas, R. (2016). *Bioética para todos*. Trillas.

Massini-Correas, C. (2019). *Alternativas a la ética contemporánea. Constructivismo y realismo ético*. RIALP.

Medina-Delgadillo, J. (2010). *¿El mesías soy Yo? Introducción al pensamiento de Emmanuel Levinas*. Conspiratio.

Medina-Delgadillo, J. (2017). *Decir en griego la novedad del hebreo*. Libros Certeza.

Miller, J. (2022). Catholic health care and AI ethics: Algorithms for human flourishing. *The Linacre Quartely, 89*(2), 152-164.

Navarro, O. (2008). El «rostro» del otro: Una lectura de la ética de la alteridad de Emmanuel Lévinas. *Revista Internacional de Filosofía,* 13, 177-194.

Oliver, N. (2018). *Inteligencia artificial: ficción, realidad y... sueños*. Real Academia de Ingeniería.

Research. (s/f). Recuperado el 1 de marzo de 2024, de https://openai.com/research/overview

Sanguinetti, J. J. (2021). Ciencia, tecnología y mundo humano. Ediciones Logos.

Sedmak, C. (2020). El bien común, bajo cero. En M. Nebel (Ed.), *Generar un porvenir compartido. Cómo crear dinámicas de bien común en México*, pp. 45-64. Tirant lo Blanch

Schuurman, D. (2019). "Artificial intelligence: Discerning a christian response". *Perspectives on Science and Christian Faith*, 71: 75-82.

Teixidó Durán, O. F. (2023). La ética de la automatización en vehículos y enfoques alternos para problemas morales actuales. *Revista de Bioética y Derecho,* 57, 153-180. https://doi.org/10.1344/rbd2023.57.3802