Floridi, L. (2023). *The Ethics of Artificial Intelligence*: *Principles, Challenges, and Opportunities*. Oxford University Press. 243 pp.

Estamos en un punto de inflexión histórica: las formas en que trabajamos, nos relacionamos y nos organizamos han cambiado de manera irreversible. Nuestras interacciones con los demás, con el mundo —y hasta con nosotros mismos— cada vez se ven más limitadas por las tecnologías digitales que no son simples mediadoras, sino que configuran las formas y el alcance de estas interacciones. Llevábamos años abandonando el mundo análogo de manera paulatina hasta que, en 2020, la pandemia arrojó a una buena parte de la sociedad a un mundo predominantemente digital (cfr. Floridi, 2023, p. 4). Ahora vivimos en una especie de mundo híbrido que bordea los límites del estar *en línea* (*online*) y en la realidad análoga. Luciano Floridi llama a esta nueva situación *onlife*.

Esta nueva realidad hasta ahora ha sido moldeada, en su mayoría, por las tecnologías y sus diseñadores. Lo que nos deja con la sensación —basada en la experiencia— de que el desarrollo tecnológico tiene su propio rumbo, aislado de los intereses de la humanidad. Se sigue un imperativo tecnológico —*sistematizar*, *optimizar*, *automatizar*— y nosotros nos atenemos a lo que resulte de ello. El poco espacio de acción que nos deja para actuar es moldeado por otras tecnologías, como el Estado y la ley, y desde estas tecnologías tenemos la posibilidad de ampliar nuestra área de injerencia. En este sentido, el objetivo de la filosofía —política, del derecho y moral— sería, de acuerdo con Floridi (p. xvi), "buscar, comprender, moldear, implementar y negociar lo moralmente bueno y correcto". Así ampliamos nuestro espacio de acción, tomamos el control sobre nuestro futuro.

Se trata de abrir brecha en la inercia de lo económico y lo político que estructura al mundo para encaminarnos al lugar deseable de la ética. Para ello, se vuelve necesario preguntarnos a qué futuro nos encaminamos y cuál deseamos. ¿Qué lugar debemos y queremos darles a las herramientas tecnológicas en nuestras vidas?, ¿qué vidas nos podemos imaginar gracias a ellas?, ¿qué limites tenemos que imponer ante la intrusión de nuevas tecnologías? ¿Cuál es el verdadero alcance de la inteligencia artificial (IA) y la robótica? Estas y muchas otras preguntas guían la discusión y preocupaciones de Floridi en *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities* (en adelante, *The Ethics of AI)*.

The Ethics of AI es el trabajo reciente más exhaustivo (en inglés o español) sobre la ética de la inteligencia artificial —un subcampo de la ética de la tecnología—. No solo identifica los daños morales y legales que la IA pueda causar, sino que sitúa el fenómeno en su dimensión histórica excepcional para trazar una ruta de futuro deseable, que le haga justicia a la humanidad y a la biosfera. El autor, el filósofo italiano Luciano Floridi, es uno de los principales exponentes de esta *nueva* disciplina, así como de la ética digital y la ética de la informática. Si bien estas tres áreas de estudio pueden intercambiarse en algunos casos, parte del valor de *The Ethics of AI* está en situar la especificidad que implica la inteligencia artificial para los retos de la ética y las sociedades actuales.

En este sentido, los capítulos del libro ofrecen: (1) coordenadas para leer la irrupción de la IA y lo digital en la historia, así como el desarrollo histórico de la noción de IA; (2) una crítica a la definición de "inteligencia artificial" y una propuesta para comprender *lo que sí es*; (3) una perspectiva sobre el desarrollo futuro de la IA; (4) un marco de principios éticos definido; (5 y 6) los retos del paso de la ética a su aplicación; (7 a 12) un panorama sobre riesgos y oportunidades de la IA; y (13) una propuesta de futuro para integrar las nuevas tecnologías con nuestros hábitats —naturales, urbanos y sociales—.

La irrupción de la IA y lo digital en la historia

El primer capítulo del libro plantea la transición del mundo enteramente análogo —que ya no existe para la mayoría de nosotros— a un mundo digital en donde hemos depositado en otros formatos los datos sobre nuestros entornos análogos, híbridos y digitales. Para comprender este cambio de entornos, Floridi emplea la analogía de la distinción entre *hardware* y *software*. No solo es que tengamos acceso a lo digital, sino que el *software* ya es una realidad integrada a la anterior, análoga —representada por el carácter material del *hardware*— y que, dentro de ella, se ha ido expandiendo.

Este cambio se logró gracias a la transferencia o traducción de la información como se presenta en la realidad al código legible por las máquinas. Dado que acumulamos datos sobre todo lo que se puede medir, etiquetar, categorizar y sistematizar, la disciplina de la IA pasó de ser una rama de la lógica a una rama de la estadística (p. 5). El punto clave, argumenta Floridi (p. 6), es que esta "revolución digital" transformó nuestra forma de comprender y conceptualizar la realidad.

Una vez que todo objeto adquiere un lugar en la red digital, este puede resituarse, expandirse, minimizarse, relativizarse y un sinfín de operaciones posibles gracias al "copia y pega" de lo digital (p. 6). Las implicaciones epistemológicas y ontológicas de estas operaciones van desde la disociación de la presencia con el lugar —el hecho de que podemos encontrarnos en espacios digitales— hasta las nuevas formas de leer nuestras identidades personales en términos de datos personales (p. 7). Nos hemos vuelto conjuntos de datos a los ojos de la máquina y a través de ella hemos interpretado otras realidades en estos términos. Floridi llama a este movimiento una re-ontologización (*re-ontologizing*) de la realidad y su re-epistemologización (*re-epistemologizing)* (pp. 9-10).

A través de esta *nueva realidad*, Floridi nos invita a interpretar las maneras en que relacionamos distintos objetos, conceptos y hechos: cómo ahora unimos o separamos elementos que, en el mundo análogo no se concebían bajo estas nuevas formas. La IA en particular plantea una nueva forma de comprensión de la inteligencia que la separa de la agencia (p. 24). Esto refiere al hecho de que ahora no se requiere inteligencia para hacer cosas que antes solo los humanos podían hacer.

¿Qué es, realmente, la inteligencia artificial?

De acuerdo con Floridi, la etiqueta "inteligencia artificial" es engañosa e incluso un oxímoron. En línea con el argumento ya tradicional de John Searle, Floridi sostiene que lo artificial nunca podrá ser inteligente porque carece de nuestras capacidades semánticas: se queda al nivel de la información, pero nunca podrá darle un sentido como lo hacemos nosotros, que la convertimos en conocimiento significativo (pp. xiii y 208). La clave sigue siendo la intencionalidad y parecería que es un umbral imposible de cruzar por vía de lo artificial.

Los humanos nos hacemos llamar inteligentes, en buena medida, por todas las cosas que podemos hacer y que los demás agentes a nuestro alrededor hasta hace muy poco no podían hacer. Nuestras formas particulares de conducirnos en el mundo tienen que ver con nuestras capacidades de comprensión semántica: los conceptos que nos mueven siempre se relacionan de alguna manera con objetos externos a ellos mismos. Esto es: en lo humano, inteligencia y agencia van de la mano.

Bajo esta comprensión de lo artificial y la inteligencia, Floridi propone una *nueva* manera de comprender a la IA como agencia artificial:

> La IA se vuelve posible gracias al desacoplamiento de la agencia, no de la inteligencia; de ahí que la IA sea mejor comprendida como una nueva forma de agencia, no de inteligencia. Entonces, la IA es una sorprendente revolución, pero en un sentido pragmático y no cognitivo; [en este sentido,] los retos y oportunidades concretos y urgentes derivados de la IA provienen de la separación entre la agencia y la inteligencia, que continuarán alejándose en la medida en que la IA se vuelva más exitosa (p. xiii).[1]

Esto es: la IA, en lugar de ser consciente o inteligente, es más bien la capacidad de realizar cosas y solucionar problemas que antes requerían de inteligencia, pero a la que esta última ya no le es necesaria. En este sentido, Floridi sugiere que sería más adecuado hablar de agencias artificiales (AA) que de inteligencias artificiales (p. 48).

Esta definición propuesta por Floridi para comprender a las IA es una de las tesis centrales del libro y es la clave para entender estas herramientas: cuáles son sus limitaciones y hacia dónde es razonable esperar que avancen. Además de ser una de las tesis centrales del libro, esta nueva aproximación conceptual tiene el potencial de cambiar las comprensiones convencionales de la IA. Entre otras razones, porque aporta un buen referente para superar las narrativas que exageran o minimizan las capacidades de la IA. Brinda un punto de partida para la comprensión del impacto que tienen las AA y tendrán en un futuro en nuestras sociedades.

Un segundo argumento central del libro, que se sigue de la definición y que comparte este profundo impacto para la propuesta y la disciplina, explica la idea de que las condiciones de éxito de la IA tienen que ver con los entornos controlados que construimos a su alrededor. Esto es, que las máquinas son más exitosas cuando les creamos una especie de envoltura (*envelope*) que simplifica el entorno para facilitar sus tareas (p. 25). Floridi pone un lavavajillas como ejemplo para comprender tanto la noción de "agencia artificial" como este envolvimiento de las máquinas. No consideramos inteligentes a estas lavadoras por hacer algo que antes solo hacíamos los humanos, y estas máquinas son tan efectivas porque vienen con su propia *envoltura* que permite simplificar la tarea a

[1] Las traducciones son mías.

las capacidades del sistema. En la misma línea, los robots más exitosos son aquellos que operan en un entorno simple —a menudo creado para ellos—, como en las fábricas.

El futuro de la IA

Para comprender hacia dónde pueden llegar las IA, Floridi presenta el estado del desarrollo actual y sus retos técnicos principales. Argumenta que para un desarrollo exitoso de estas herramientas hay que reestructurar los problemas que les planteamos. Las IA son muy buenas para resolver problemas complejos, pero les cuestan más trabajo las tareas difíciles que requieren muchas destrezas (p. 42). De ahí, por ejemplo, que los coches automáticos no requieran de un androide que aprenda todas las destrezas humano-somáticas, sino una computadora que reestructure el problema de manejar y un entorno que se lo simplifique (pp. 39-43).

La irrupción de los modelos generativos también ofrece claves importantes para comprender qué aspectos de nuestras sociedades serán afectados por la IA en el futuro próximo y de qué maneras. Al respecto, Floridi argumenta que debemos ser conscientes tanto de los mejores usos como de los usos cuestionables de los modelos generativos, así como de las maneras en que trabajamos con ellos. Ahora que estas herramientas se usan especialmente para escribir cualquier tipo de texto, habría que definir el lugar de los humanos en la redacción de textos relevantes, como la legislación, contenidos educativos, la programación y la investigación científica (p. 48). Además de estos cambios, al impacto moralmente relevante de modelos como el ChatGPT se suma el costo ambiental y el uso de mano de obra barata en el sur global (p. 47).

Principios para la ética de la IA

A partir de los dos argumentos medulares —que la IA es más bien AA (agencia artificial) y que su éxito depende de la posibilidad de crearle entornos propicios—, Floridi presenta un marco de cinco principios centrales para una ética de la IA (p. 55): (1) beneficencia, (2) no maleficencia, (3) autonomía, (4) justicia y (5) explicabilidad. El conjunto de estos principios integra, de acuerdo con un análisis comparativo de Floridi, las principales preocupaciones en las listas más relevantes de principios éticos para la IA que se han presentado recientemente en Europa y Norteamérica (pp. 58- 60). Los primeros cuatro principios

provienen de la bioética y el quinto refiere a un problema propio de la IA.

(1) El primer principio, de *beneficencia*, busca asegurar que el desarrollo de la IA sea para el bien común de la humanidad y sus entornos naturales. (2) Como complemento, se introduce el principio de *no maleficencia*: no solo se trata de que las IAs hagan el bien, sino también de que no produzcan daños (p. 61). (3) Para responder a la pregunta sobre quién debe encargarse de que la IA cumpla con estos principios —quién debe tomar las decisiones—, se introduce el principio de *autonomía*. A medida que la IA va adquiriendo más agencia —o autonomía—, debemos asegurarnos de que esto no minimice la autonomía humana. Sobre este principio, Floridi argumenta que los humanos deben decidir qué delegarle a las máquinas y por qué, al mismo tiempo, siempre deben tener la capacidad de restringir y recuperar la agencia delegada (p. 62).

Una de las razones por las que los humanos deben tener siempre el control último es que nosotros comprendemos en términos morales las condiciones del actuar, así como sus consecuencias. (4) De ahí que se requiera un principio para asegurarnos de que estas decisiones humanas beneficien a todas las personas: el principio de *justicia*. La importancia de este principio, considero, reside en que le da un sentido a los valores humanos que debe perseguir la IA: no se puede optimizar hacia el crecimiento económico si este no es justo —compartido, sostenible—, por ejemplo. Solo se podrían hacer efectivos estos principios una vez que se puedan identificar adecuadamente las responsabilidades. (5) Para ello, el principio de *explicabilidad* integra tanto su sentido epistemológico como el ético: busca hacer los procesos de la IA inteligibles para poder responder a la pregunta "quién es responsable de la manera en que funcionan" (p. 60).

De la ética a su aplicación

Para hablar de la traducción de los principios éticos en buenas prácticas, Floridi presenta primero las malas prácticas o riesgos en el mal uso de la ética y, después, introduce una distinción entre *ética suave* y *ética fuerte* para profundizar en el alcance de la ética. Cabe destacar que si bien los principios que presenta fueron concebidos en y para el norte global (principalmente en el contexto europeo), el problema de su aplicación ilustra dinámicas moralmente relevantes sobre la aplicación de estos principios en una escala global. Floridi señala, por ejemplo, el riesgo de que las compañías ejecuten sus malas prácticas en países con

regulaciones laxas (o inexistentes) para usar los resultados de dichas malas prácticas en contextos bien regulados (pp. 72-74).

También observa la posibilidad de que se los principios éticos se usen como estrategia de *marketing* engañosa: para publicitar la implementación de prácticas éticas superficiales o insuficientes y así limpiar su imagen sobre otras malas prácticas (p. 70). Igualmente, se corre el riesgo de que las empresas elijan el conjunto de principios éticos que les convengan y omitan los que no les beneficien o que dirijan la atención hacia la ética para distraer sobre la necesidad de legislar y aplicar la ley (p. 71). En conjunto, todas estas prácticas viciadas señalan un riesgo más grande: la evasión de las responsabilidades éticas (p. 74). Si la ética se vuelve optativa o maleable a los intereses particulares, entonces se podrá evitar.

En buena medida, el problema de la aplicación de los principios éticos tiene que ver con las diferencias entre lo que Floridi llama *ética suave* y *ética fuerte*, que serían, respectivamente, la brecha entre gobernanza (basada en la ética) y moral. La distinción refiere a la *ética fuerte* como aquella serie de principios éticos de los que se deriva la regulación y a la *ética suave* como los principios morales que quedarían fuera de la ley —pero constituyen la distinción entre lo que se debe hacer y lo que se debe evitar— (p. 82). Esta distinción permite señalar que la ética va más allá de la ley y este espacio externo a la ley tiene un valor en tanto que le da contexto y sentido a las normas a la vez que motiva (o debería motivar) mejores prácticas.

Más allá de moldear la regulación, la ética presenta oportunidades para el desarrollo mismo de la IA, como bien expone Floridi (pp. 89-91). En primer lugar, permite que las personas involucradas aprovechen los usos socialmente valiosos de las tecnologías digitales. En segundo lugar, brinda una perspectiva amplia para la pronta identificación de posibles riesgos en la implementación que podrían ser dañinos o costosos. En este sentido, lejos de presentarse como una limitante para el progreso de la IA, la ética debe verse como una herramienta para un mejor desarrollo de la tecnología: una herramienta que debe guiarla, no acotarla.

Riesgos y oportunidades de la IA

Floridi dedica seis capítulos del libro (7-12) a exponer un panorama robusto y necesario sobre los riesgos y oportunidades concretos que abre la IA en nuestras sociedades. De manera general, considero que sobre-enfatiza el potencial positivo de la IA para la solución de problemas

complejos —que requieren del análisis de muchos datos y predicciones efectivas— frente a los riesgos (y su gravedad). Entre los riesgos, destaca los malos usos de los algoritmos que dan con resultados injustos, opacos o inadecuados, y el uso deliberadamente criminal de la IA —para el fraude, robo de datos, especulación financiera y tortura, en particular—. Entre las oportunidades, se muestra optimista sobre el uso de la IA para el bien social (AI4SG: *AI for Social Good*). Señala que uno de los riesgos de la IA (o de su ética) está en desaprovechar su potencial para resolver algunos de los problemas que afectan a toda la humanidad, como la injusta distribución de los recursos y la crisis climática.

Más allá de la valiosa exposición de un catálogo de riesgos y oportunidades, considero que esta lista podría señalar un posible problema en el uso de un marco de principios de la ética práctica —que se presenta en relación con los casos y no a partir de una teoría de base determinada, como lo hace la ética aplicada—. Dado que Floridi usa el modelo de la bioética actual —donde predomina la ética práctica—, no parece claro bajo qué marco subyacente se resolverían aquellos casos controversiales en donde se tendrían que ponderar dos principios distintos. Este problema se expresa en que encontramos casos donde un riesgo y una oportunidad podrían ser contradictorios —como sugerir un mayor desarrollo de la IA y señalar su gran impacto ambiental, o como dar por hecho la inviabilidad de prohibir las armas autónomas y el principio de no maleficencia—.

A dónde nos lleva *The Ethics of AI*

Los problemas que podemos encontrar en el marco presentado por Floridi para la ética de la IA se dan en buena medida porque esta ya nos da una guía de acción con bases conceptuales sólidas. En este sentido, incluso ante sus propios retos es una propuesta valiosa en la medida en que se unifica y justifica una serie de pautas para moldear el desarrollo y la regulación de la IA. Funciona, así, como un excelente punto de partida para la evaluación ética de la inteligencia artificial, de las prácticas de sus creadores y usuarios. Pero la propuesta de Floridi, como mencioné al inicio de esta reseña, va más allá de los principios éticos, riesgos y oportunidades.

Un elemento que vuelve valiosa la obra es que se sitúa en su dimensión histórica: en la búsqueda de un objetivo conjunto para la especie. El desarrollo tecnológico debe estar en manos de los humanos y nosotros tenemos la obligación de conciliar la armonía entre *lo verde*

de nuestros entornos naturales y sociales y *lo azul* de las tecnologías digitales (p. 202). Las formas de hacerlo posible involucran a la filosofía, de manera general, para la reconceptualización de la nueva realidad y del horizonte de futuro que deseamos, y a la filosofía política en particular: tenemos el reto de *volver a lo humano*, de poner en el centro el bienestar de la especie y de sus entornos sociales. De esta manera, Floridi concluye su *capolavoro* con una invitación a hacernos cargo de nuestro futuro y una invitación a construir una nueva narrativa donde la humanidad integre lo sintético con lo natural de maneras que sean beneficiosas para todo lo vivo. Para efectuar su propuesta ética, mantener un nivel de esperanza se vuelve un imperativo: las pautas para organizar ese proyecto deseable se presentarían en su próximo libro, *The Politics of Information*.

Tatiana Lozano Ortega
Universidad de Birmingham
t.lozano.ortega@gmail.com