

**MARK COECKELBERGH. *AI Ethics*. Cambridge, MA: The MIT Press, 2020.**

*Espejito, espejito en la pared*,<sup>1</sup> ¿quién podrá ayudarnos a superar el *hype*<sup>2</sup> que ha traído la Inteligencia Artificial (IA) y sus adeptos? Mark Coeckelbergh, sin duda alguna, es uno de los filósofos más capaces de lograr este cometido. Existimos junto a algoritmos que nos comparten información y nos convencen de tomar decisiones políticas, como también convivimos con vehículos autónomos que deben tomar decisiones —atropellar a alguien que se cruza por la calle o chocar y matar a sus pasajeros—, y, por supuesto, podemos ser juzgados por las técnicas analíticas de la policía predictiva (*Predictive policing*) y el reconocimiento facial. No obstante, paralizarse frente a estos eventos e intentar cancelar las investigaciones sobre IA no es una opción a la altura de las actuales urgencias planetarias. Así, el objetivo de este libro es generar una ética de la IA, capaz de reflexionar sobre el “technological change and its impact on individual lives, but also about transformations in society and in the economy”<sup>3</sup> (7), lo que se reflejará —en una primera instancia— en ofrecer una visión amplia de los problemas éticos relacionados con la IA y, —en una segunda instancia— en la presentación de consejos para la elaboración de políticas que nos permitan enfrentarnos a estos problemas antes de que sea demasiado tarde.

Ambos objetivos —que podríamos respectivamente denominar descriptivo y prescriptivo— son desarrollados a lo largo de los doce capítulos que componen el libro y responden a la idea con la que hemos terminado el párrafo anterior, ¿qué pasaría si es “demasiado tarde” para enfrentarnos a los problemas éticos que implica el desarrollo actual de la IA? (10). De esta manera, el segundo capítulo del libro *Superintelligence, Monsters, and the AI Apocalypse* se encarga de ilustrar las ideas de los ingenieros, futurólogos y transhumanistas que tanta expectación causan a la hora de pensar sobre la IA. Este *hype* se podría resumir en dos tópicos claves, la trascendencia y el apocalipsis, que en ambos casos nos enseñan los rasgos culturalmente específicos detrás de los adeptos a estas investigaciones. Bostrom, Kurzweil o Musk, siendo los portavoces de la *superinteligencia*, ilustran lo anterior

---

<sup>1</sup> El buen humor de Mark Coeckelbergh se hace presente desde el primer capítulo de su texto. Este se titula: “*Mirror, Mirror on the Wall*” y en un subtítulo se introduce la pregunta al espejo consejero: *Who is the Smartest of Us All?*

<sup>2</sup> La palabra *hype*, al menos en el contexto del libro, refiere a la sobrevaloración y las grandes expectativas que han generado los estudios sobre la IA, expectativas que muchas veces nos llevan a pensar en el fin de la humanidad o sobrevaloración que nos invita a imaginar cómo la IA podría solucionar muchos de los problemas de la humanidad. En ambos casos, Coeckelbergh se encarga de aterrizar la discusión tanto a lo realmente posible como a lo éticamente atingente.

<sup>3</sup> Traducción personal: “el cambio tecnológico y su impacto en las vidas individuales, pero también sobre las transformaciones en la sociedad y en la economía”.

tanto con sus miedos como con sus esperanzas: la *superinteligencia* podría decidir que lo mejor para nosotros sería nuestra destrucción o bien podría fusionarse con nosotros a fin de superar nuestra condición humana. Sin embargo, Coeckelbergh destaca que un examen crítico de los presupuestos sobre la relación entre la IA, la tecnología, el mundo y nosotros permitiría visualizar cómo es que la reflexión ética sobre la IA podría ser aplicada hoy. En primera instancia, este examen revelaría que quizás no todo tiene relación con la especie humana.

Por esto, el capítulo *All about the human* describe cómo es que las filosofías que han pensado la posibilidad de una inteligencia artificial gravitaron en torno al paradigma de la inteligencia humana —con el “ser-en-el-mundo” de Dreyfus como posición crítica y el “funcionalismo” de Churchland o Dennet a la defensa—. Sin embargo, determinar qué es la inteligencia solamente desde lo humano no está a la altura de nuestros tiempos. La comprensión por lo “humano” es una batalla que no ha terminado y gracias al avance de la discusión filosófica posthumanista de Donna Haraway y postfenomenológica de Don Ihde podemos pensar con mayor precisión sobre la relación entre nosotros y la tecnología. ¿No hemos sido siempre seres que usan la tecnología? (45) y si esto es así, entonces, nuestra relación con la IA ¿es una relación con *solamente máquinas*?<sup>4</sup> La pregunta es brillante, porque nos lleva a cuestionarnos cuál es el lugar de lo no-humano en la discusión sobre la IA. No es raro encontrar argumentos humanistas que degradan el estatus moral de entidades no-humanas debido a no cumplir con la característica de ser “racionales”, pero tampoco es seguro que la “racionalidad” sea una condición necesaria para que alguien tenga estatus moral. De esta manera, Coeckelbergh se pregunta si las IA pueden ser consideradas en términos de agentes (*agency*) y pacientes (*patiency*) morales. Este cuestionamiento nos permite reconocer que cuando discutimos sobre el estatus moral de la IA también discutimos sobre cómo deberíamos relacionarnos —y cómo lo hacemos hoy— con las entidades no-humanas con las que actualmente convivimos, y, a nuestro parecer, la clave de este cuestionamiento radica en lo *actual*, ya que hoy no nos relacionamos con *superinteligencias*, sino con tecnología que debe ser comprendida en su aplicación.

Superar el *hype* de la IA se lograría, entonces, teniendo una idea de la *tecnología*<sup>5</sup> y la *ciencia*<sup>6</sup> que conforman actualmente las investigaciones sobre IA. La urgencia de la reflexión ética nos invita a reflexionar más sobre el hoy antes que sobre los futuros utópicos, objetivo que el autor logra gracias a la exposición de una breve pero consistente historia de los aspectos que cada disciplina ha conjugado para que la IA se nos presente tal y como la conocemos hoy: como un

---

<sup>4</sup> El título del cuarto capítulo es: “*Just Machines?*”

<sup>5</sup> Este es el nombre del capítulo quinto, en inglés: *The Technology*.

<sup>6</sup> Este tópico es desarrollado en el sexto capítulo del libro, que se titula: *Don't Forget the Data (Science)*

dispositivo tecnológico que, en su aspecto más pragmático, tiene un carácter ético y social (68). Con esto en mente, es fácil entender por qué no hay una IA en sí, ya que su posibilidad de ser radica en apoyarse en muchas tecnologías y procedimientos científicos de los que también hereda sus problemas éticos específicos. De hecho, si consideramos la descripción del *Machine Learning* presentada en el sexto capítulo, entonces, se ilustra muy bien que las IA pueden obrar autónomamente con datos que deben ser explicados e interpretados por algún humano. Los datos siempre son sesgados<sup>7</sup> y, por lo tanto, los patrones encontrados en este proceso también heredan estos sesgos. No obstante, esto no ha sido una razón suficiente para no aplicar el *Machine Learning* en todos los ámbitos sociales, lo que innegablemente ha afectado a nuestra privacidad a la vez que a la economía mundial (100)<sup>8</sup>. Hay que advertir que, si bien el procesamiento de datos es autónomo, seguimos siendo nosotros los que decidimos qué aplicación tendrán en nuestro entorno social.

Es inevitable preguntarnos qué tan dependientes queremos hacernos de esta tecnología (114) dada su cualidad de ser “arresponsables” y su incapacidad de explicar muchas de sus decisiones<sup>9</sup>. Si bien no podemos exigir a las IAs que tomen responsabilidad por sus acciones —porque no sabemos si saben lo que están haciendo—, no podemos obviar que ellas están configuradas por personas reales, con historia, que podrían transparentar y explicar por qué estas tecnologías han tomado ciertas decisiones. Porque pueden... ¿no?

Este movimiento, escéptico si se quiere, le permite al autor del libro acabar con las expectativas utópicas en torno a las capacidades de la IA en la medida en que tanto ellas como nosotros nos movemos otorgándole sentido al mundo desde nuestros prejuicios<sup>10</sup>. Es cierto, los sesgos no siempre son intencionales, pero el *Machine Learning* tampoco diferencia entre conjuntos de datos que discriminan “positiva” o “negativamente” a ciertos grupos de personas. Nos enfrentamos a una caja negra que impacta en nuestra comprensión de la realidad, y si esto es así: ¿qué se debe hacer frente a los problemas éticos que nos trae esta tecnología? La respuesta de Coeckelbergh es tajante: legislar<sup>11</sup>. O, al menos, aconsejar a los que

---

<sup>7</sup> ¿Cómo podríamos definir la objetividad de nuestra actividad en la web si estas acciones no pueden ser más que personales y en extremo subjetivas?

<sup>8</sup> El caso de Cambridge Analytica y la minería de datos es uno de los eventos más ilustrativos de este fenómeno. Estas ideas están analizadas en el séptimo capítulo, titulado: *Privacy and the Other Usual Suspects*.

<sup>9</sup> El octavo capítulo, que discute cómo podríamos atribuir responsabilidad moral a una IA, se titula: *A-responsible Machines and Unexplainable Decisions*.

<sup>10</sup> Así, el capítulo 9 del libro se titula: *Bias and Meaning of Life*.

<sup>11</sup> Un ejemplo de cómo se legisla, tanto en Occidente como en Oriente, es presentado en el capítulo *Policy Proposals* y en el capítulo 10, *Challenges for Policymakers*. En ambos el autor detalla a qué desafíos se enfrentan las personas que quieren legislar sobre tecnología.

legislan, y para lograr esto se requeriría una metodología capaz de seguir el ritmo del desarrollo de la tecnología. No basta con delimitar lo que no se puede hacer, sino también de generar propuestas que busquen asegurar la prosperidad de todas las realidades que son afectadas por la tecnología. A nuestro parecer, uno de los mejores consejos que se encuentran en esta obra es intentar evitar un ‘Invierno IA’ —el uso indiscriminado de esta tecnología— por medio del fomento de la educación desde perspectivas inter y multidisciplinarias. De esta manera, la prioridad<sup>12</sup> de este legislar apuntaría a alcanzar maneras menos antropocéntricas de relacionarnos con el mundo, porque el desarrollo de esta tecnología podría convencernos de que antes de colonizar la galaxia con vehículos Tesla quizás deberíamos volver a pensar nuestra relación con nuestra tierra.

*Nicolás Rojas Cortés*<sup>13</sup>  
Universidad de Chile, Santiago, Chile  
nicolas.rojas.c@ug.uchile.cl

---

<sup>12</sup> El filósofo concluye su libro ofreciendo una crítica al antropocentrismo de las investigaciones actuales sobre IA y sus propósitos —la colonización de otros planetas— en el capítulo: *It's the Climate, Stupid! On*

*Priorities, the Anthropocene, and Elon Musk's Car in Space.*

<sup>13</sup> Agradecimientos: Agencia Nacional de Investigación y Desarrollo / Subdirección de Capital Humano / Beca de Doctorado Nacional 21210804.