

Las leyes de la robótica de Asimov, aunque pioneras, revelan ciertas fallas. Un pacto global digital podría aprender de estos errores para priorizar la sostenibilidad, la coexistencia y la anticipación de riesgos en un futuro justo y sostenible.

Las tres leyes de la robótica (Asimov, 1942: 94-110), aunque concebidas originalmente para un escenario ficticio en la década de 1940, representan un ideal visionario de la interacción humano-máquina, por lo que quizás puedan ofrecernos un punto de partida interesante para reflexionar sobre los principios básicos de un pacto global digital que no solo se enfoque en diagnosticar los problemas, sino que también intente diseñar un marco de acciones anticipatorias que, más que prevenir conflictos derivados de la interacción entre humanos y tecnología, fomente una simbiosis que promueva el bienestar colectivo y la sustentabilidad ética. Así que, antes de adentrarnos, conviene revisarlas:

1. Un robot no hará daño a un ser humano, ni por inacción permitirá que un ser humano sufra daño.
2. Un robot debe cumplir las órdenes dadas por los seres humanos, a excepción de aquellas que entren en conflicto con la primera ley.
3. Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la primera o con la segunda ley.

Como he mencionado, es importante resaltar que estas leyes no fueron diseñadas para su aplicación en el contexto real, así que puede que sea un poco injusto juzgarlas en comparación, no solo con el mundo real, sino con el actual; sin embargo, para fines reflexivos es conveniente que analicemos cuáles son sus fallas, ya que podría existir el riesgo de repetirlas al proponer caminos hacia una gobernanza efectiva de la tecnohumanidad¹.

Un pacto mundial digital podría incluir principios que regulen la convivencia entre seres humanos, máquinas y entidades no humanas

Falla 1. Lenguaje humano

Conceptos como daño, obedecer o ser humano son complejos y polisémicos, y dependen en gran medida del contexto cultural y social. ¿Qué se considera un ser humano? ¿Un feto? ¿Una persona en estado vegetativo? ¿Una copia digital de un cerebro? ¿Qué constituye un daño? ¿Dejar morir a alguien que sufre es causarle daño o ayudarlo? ¿Cuánto daño es aceptable en cada caso? Por ejemplo, en el uso de drones autónomos en operaciones militares, la definición de daño se torna crucial y controvertida, destacando las dificultades de programar estas nociones en algoritmos (Moor, 2006: 34).

Como se puede apreciar, mientras Asimov se centraba en la interacción entre humanos y robots autónomos, la actual interdependencia entre tecnología y biología sugiere un nuevo paradigma en el que la ética robótica debe evolucionar hacia una ética más amplia que abarque la tecnohumanidad. Sin embargo, antes de llegar a ese punto, valdría la pena aprovechar el momento y detenernos a preguntarnos si el desarrollo de un pacto mundial digital debería considerar seres no humanos u otras formas de vida (Floridi, 2013: 72).

Desde mediados del siglo XX, la deconstrucción del sujeto moderno y el antropocentrismo, es decir, la idea de que los humanos ocupan un lugar privilegiado como medida de todas las cosas, ha sido una constante. Enfoques teóricos como el posestructuralismo y el poshumanismo han descentralizado al hombre en una red de relaciones simétricas con diversas entidades como cosas, plantas y animales (Latour, 1993: 105). Este desplazamiento ha cuestionado la figura del hombre en todas estas tradiciones teóricas poscartesianas, deconstruyendo dualismos, esencialismos y antropocentrismo (Haraway, 2006: 117-158).

La crítica a la centralidad humana en la filosofía y las ciencias sociales ha promovido un cambio significativo en cómo percibimos la tecnología y su interacción con lo vivo. Este cuestionamiento ha permitido explorar las posibles simbiosis entre tecnología y biología, reconociendo a la tecnología no solo como herramienta sino como parte integral de los ecosistemas biológicos y sociales. Este enfoque plantea la necesidad de reconsiderar las éticas aplicadas, no solo en términos de cómo los humanos utilizan las máquinas, sino también cómo las máquinas pueden influir en nuestro entorno y coexistir con otras formas de vida (Bostrom, 2014: 112).

Por lo tanto, un pacto mundial digital que contemple estas interacciones complejas podría necesitar incluir principios que regulen la convivencia entre seres humanos, máquinas y entidades no humanas. Esto implica desarrollar una ética que no solo proteja la integridad humana, sino que también considere los derechos y bienestar de todas las formas de vida afectadas por la expansión tecnológica.

Falla 2. Ética y moral

Las perspectivas sobre lo que es éticamente correcto pueden variar significativamente entre diferentes culturas. El experimento *Moral Machine*² ilustra cómo las decisiones éticas pueden diferir globalmente, lo que plantea desafíos significativos en la posible interacción de seres humanos y máquinas con programación de decisiones éticas.

Con lo anterior surgen nuevas preguntas: ¿Qué valores éticos se deberían priorizar? ¿Cómo se pueden conciliar las diferentes perspectivas morales? ¿La tecnología que diseñamos debería parecerse a nosotros o ser radicalmente diferente? ¿Debería reflejar nuestra humanidad o ser completamente distinta? ¿Qué papel queremos que desempeñen estos dispositivos en nuestra sociedad? ¿Deben ser herramientas que nos sirvan o entidades con las que interactuemos como iguales?

El principal desafío de codificar la ética en algoritmos radica en la variabilidad y complejidad de los juicios éticos humanos. Los dilemas éticos son dinámicos y a menudo requieren una comprensión profunda del contexto, algo que las máquinas aún no pueden replicar completamente. Esto plantea preguntas sobre cómo y hasta qué punto las decisiones automáticas deben ser supervisadas o modificadas por intervención humana.

Al final del día, la ética es un campo dinámico que no se presta fácilmente a la codificación fija. Proyectos como *Ethics and Algorithms Toolkit*³ trabajan para mitigar los riesgos algorítmicos, pero destacan la complejidad de integrar la ética de manera efectiva en sistemas autónomos.

De ahí que entonces, más allá de la ética, el pacto global digital podría abordar los riesgos que amenazan la seguridad, la privacidad y el bienestar de las sociedades en todo el mundo. Es decir, priorizar la gestión de los siguientes riesgos sobre los aspectos éticos, asegurando así un entorno digital más seguro y confiable para

todos.

1.-Veracidad y autenticidad de la información. El pacto internacional debe exigir a las plataformas digitales la implementación de tecnologías de verificación de hechos y moderación de contenido. Esto incluye algoritmos que detecten y desacrediten automáticamente la información falsa, así como etiquetas de verificación para fuentes confiables.

2.- Prevención de la adicción digital. Es esencial abordar la creación de mecanismos adictivos en aplicaciones y plataformas digitales. El pacto debe promover el diseño de tecnologías equilibradas y saludables, limitando prácticas como las notificaciones constantes y los bucles de retroalimentación.

3.- Equidad económica y tecnológica. La tecnología debe ser un puente hacia la igualdad de oportunidades. El pacto debe asegurar la distribución equitativa de los beneficios tecnológicos, apoyando iniciativas de inclusión digital y estableciendo marcos regulatorios para incentivar la inversión en tecnología educativa e infraestructuras.



4.- Protección de la privacidad y los datos. Se requiere una regulación estricta sobre la vigilancia y la

recolección de datos. El pacto debe garantizar que el uso de datos personales sea transparente, consensuado y regulado para proteger la privacidad de los individuos.

5- Entorno digital seguro y respetuoso. Es necesario actuar contra los actores criminales y de odio en plataformas digitales. Esto incluye la creación de un cuerpo de supervisión digital internacional y la promoción de la educación digital y la conciencia sobre el impacto social de las actividades en línea.

Este enfoque prioriza medidas concretas para abordar los riesgos emergentes en la esfera digital, fortaleciendo nuestra capacidad colectiva para enfrentar los desafíos del futuro digital.

Falla 3. Líneas borrosas

La tecnología no es solo una extensión de nuestras capacidades físicas o mentales, sino que también contribuye a la formación de nuevas estructuras sociales, económicas y culturales. Esta interacción constante ha llevado a lo que algunos teóricos describen como una *coevolución*, donde la evolución de la tecnología y la evolución humana se influyen mutuamente de maneras complejas y a menudo impredecibles.

Hessler (2019), por ejemplo, argumenta que la historia de la tecnología debería incluir una antropología histórica de la tecnología, lo que implica una reevaluación de la interacción entre seres humanos y máquinas. Este enfoque resalta la necesidad de considerar cómo las imágenes y conceptos de relaciones humano-máquina juegan un papel crucial en la construcción, desarrollo y aceptación de la tecnología.

Desde una perspectiva histórica, la tecnología ha influido decisivamente en cómo los humanos se perciben a sí mismos y en cómo entienden su lugar en el mundo. Cada nueva herramienta, cada nueva máquina, y cada nuevo sistema informático no solo cambian lo que hacemos, sino fundamentalmente quiénes somos. Esto se ilustra, por ejemplo, en la industria cinematográfica, donde la adopción de nueva tecnología no solo ha transformado la forma en que se produce y consume el contenido, sino que también ha reconfigurado la experiencia del espectador al desafiar las convenciones tradicionales tanto de la narrativa como de la interacción audiovisual.

Así como los humanos antropomorfizamos a la tecnología, de una forma u otra, la tecnología, sobre todo las tecnologías actuales relacionadas con la IA, están cambiando nuestras formas de pensamiento, provocando en nosotros un mecanomorfismo (Waters, 1948: 139-142). Este concepto nos ayuda a describir el proceso del cual hemos sido parte, de manera consciente o inconsciente, durante los últimos años. Hoy, queramos o no, atribuimos características mecánicas al individuo humano e interpretamos el comportamiento humano en términos de conceptos y procesos característicos de las máquinas.

La influencia de la tecnología en la sociedad ha sido un tema de debate durante mucho tiempo. McLuhan (1964) afirmó que damos forma a nuestras herramientas, y luego nuestras herramientas nos dan forma, mientras que Postman (1982) argumentó que la tecnología puede adormecer nuestras mentes.

Los algoritmos son una gran muestra de este fenómeno. Hoy en día, las personas que crean contenido buscan satisfacer al algoritmo para ganar visibilidad, y al hacerlo, se están mecanomorfizando. Este fenómeno se observa especialmente en las plataformas de redes sociales, donde la relevancia del contenido está determinada en gran medida por algoritmos que priorizan la participación y el compromiso del usuario. En lugar de enfocarse en la calidad o la originalidad del contenido, los creadores a menudo se ven impulsados a producir contenido que se alinee con las preferencias algorítmicas, lo que puede llevar a una homogeneización y simplificación del contenido generado en línea.

Esta dinámica plantea interrogantes sobre la diversidad y autenticidad del contenido en el panorama digital, así como sobre el impacto en la percepción de la realidad y la formación de opiniones de los usuarios.

Un pacto global digital, por tanto, necesita contemplar estos cambios y desafíos fundamentales. Debe abordar cómo la tecnología redefine la autonomía, la autoridad y la identidad humana en un mundo interconectado. Esto incluye consideraciones sobre cómo las tecnologías emergentes, como la edición genética y la neurotecnología, podrían cambiar nuestra comprensión de lo que significa ser humano.

De modo que, después de revisar las fallas y a modo de conclusión propongo tres principios, derivados de la reflexión de las tres leyes de Asimov, que considero imprescindibles para un posible pacto global digital:

1. Principio de inclusión total. Todas las tecnologías, incluyendo la inteligencia artificial y los sistemas robóticos, deben diseñarse y operarse de manera que minimicen el impacto en el medio ambiente, protejan la biodiversidad, conserven los recursos naturales y contribuyan activamente a la salud y estabilidad de los ecosistemas, incluyendo elementos no vivos como el aire y el suelo. Además, las políticas tecnológicas deben ser formuladas y ejecutadas de manera participativa, incluyendo a representantes de todos los sectores de la sociedad, asegurando que las voces de comunidades marginadas sean escuchadas en la toma de decisiones.

La tecnología redefine la autonomía, la autoridad y la identidad humana en un mundo interconectado

2. Principio de reducción y anticipación de riesgos. Debe existir una completa transparencia en el desarrollo, implementación y operación de tecnologías avanzadas. Las entidades que diseñan, operan o gestionan sistemas tecnológicos deben demostrar su cumplimiento ante los organismos reguladores y la sociedad. La innovación tecnológica debe guiarse no solo por lo técnicamente posible o económico, sino también por lo deseable y sostenible, evaluando los impactos a largo plazo en la sociedad y en todas las formas de vida.

3. Principio de coexistencia y coevolución. La interacción entre la tecnología y la humanidad debe ser vista como un proceso de coexistencia dinámica y coevolución, eso implica que la tecnología debe promover un beneficio mutuo y equitativo para todos los seres, apoyando no solo el progreso humano, sino también mejorando la calidad de vida de todas las especies y la salud del planeta, fomentando la resiliencia y la adaptación ecológica frente a los desafíos globales como el cambio climático.

Estos principios reconocen la interdependencia entre tecnología, humanidad y el medio ambiente natural. Su inclusión fomentaría una mayor coherencia y colaboración internacional en el manejo ético de la tecnología, asegurando que la tecnohumanidad avance hacia un futuro más justo y sostenible para todos los habitantes del planeta.