SOCIAL INNOVATIONS Journal

Vol. 23 (2024): Social Entrepreneurs Leveraging and Shaping Artificial Intelligence

Titulo: El qué, cuándo y cómo de la incorporación de la gobernanza IA en las empresas sociales.

Autora: Nidhi Sudhan, co-fundadora de Citizen Digital Foundation

Biografía Nidhi es una defensora de la tecnología y los medios de comunicación responsables. Lleva 20 años dirigiendo contenidos, editoriales, marcas y análisis en publicidad, televisión, radio y medios digitales en India, Emiratos Árabes Unidos y Reino Unido. Nidhi cofundó la Citizen Digital Foundation (CDF) en la India en 2021 para fomentar soluciones regenerativas que aborden las causas profundas de los complejos retos tecnosociales. Ha sido incluida entre las '100 Brilliant Women in AI Ethics™' 2024 por Women in AI Ethics (WAIE), Nueva York.

Nidhi también ha participado como voluntaria en la rehabilitación y el apoyo a víctimas de violencia doméstica, prostitución infantil y refugiados de guerra, y ha trabajado activamente en la coordinación de ayuda en caso de catástrofe. Apoya las iniciativas de Women in STEM.

Palabras clave: IA responsable, Inteligencia Artificial, IA generativa.

Abstract:

La Inteligencia Artificial Generativa (GenAI) y sus corolarios han empezado a impregnar el tejido social, incluido el ecosistema de impacto social. A medida que las organizaciones y los Estados-nación adoptan cada vez más formas incipientes de IA, los innovadores en impacto social se enfrentan al formidable reto de establecer bases sólidas para una IA Responsable (RAI).

En este artículo se describen los pasos a seguir para la implantación de marcos de RAI y se esboza cómo se pueden poner en práctica y cuál es el momento adecuado para hacerlo. A partir de la experiencia de Citizen Digital Foundation con empresas, instituciones educativas, medios de comunicación y administraciones públicas de todo el sur de la India, este artículo ofrece una guía práctica para la implantación de la RAI y arroja luz sobre las oportunidades potenciales y los retos reales a los que se enfrentan las organizaciones en las distintas fases de transformación de la IA.

Ecosistema de Tecnología Responsable. Responsible Tech Guide 2023, All Tech Is Human.

Articulo:

La aparición de la IA Generativa ha revigorizado el discurso en torno a la IA Responsable (RAI). Impulsadas por los incentivos económicos y la trampa multipolar, las organizaciones y los Estados-nación se han apresurado a desplegar y adoptar formas incipientes de IA, a menudo sin escrúpulos. Cada vez es más evidente que el miedo a abrirse al verdadero potencial de la IA para lograr el máximo impacto social podría conducir a un aumento de las disparidades, y la falta de coordinación en la carrera armamentística de la IA podría provocar tragedias sin precedentes y el colapso de nuestro tejido socioeconómico.

Las organizaciones de impacto social y los innovadores que decidan aprovechar la IA deben prepararse para ver cómo su intención, valor e impacto disminuyen o se vuelven en su contra si no refuerzan la adopción de la IA con prácticas de responsabilidad en todas las fases de diseño, desarrollo y despliegue de la IA, para mitigar los efectos adversos. Sin marcos de RAI, los esfuerzos destinados a abordar retos complejos mediante el uso de nuevas tecnologías acabarán resolviendo los problemas de forma limitada o, lo que es peor, redirigiéndolos a otra parte. Al mismo tiempo, la escasez de conocimientos y recursos, la falta de precedentes suficientes y el peso de la responsabilidad ética ante las presiones operativas pueden actuar justificadamente como elementos disuasorios. Nuestro objetivo es abordar algunos de los retos con los que nos hemos topado durante nuestras interacciones con las empresas, la educación, los medios de comunicación y las partes interesadas del gobierno en el sur de la India.

Haga de la gobernanza de la IA su centro de atención

A muchos responsables de la toma de decisiones les resulta abrumador y engorroso iniciar la RAI en las organizaciones y no saben por dónde empezar.

Centrarse en la "gobernanza de la IA" en lugar de en la IA "responsable" o "ética" ayuda a quitarse un peso de encima. Ser ético -incluso desde el punto de vista performativo, como vemos en las iniciativas simbólicas de RSC o DE&I- exige un orden superior de comportamiento coherente, y es humano encontrarlo estresante, especialmente cuando va en contra del orden popular de las cosas.

Los directivos están más familiarizados con la "gobernanza": un viaje de colaboración que utiliza un conjunto de marcos normativos que ayudan a lograr un impacto a largo plazo y la creación de valor, pequeños hitos cada vez. Un marco así podría ayudar mucho a iniciarse en la RAI.

Evaluar detenidamente la oportunidad y la madurez de la IA en su organización

Las empresas sociales que trabajan con IA suelen tener la intención, pero rara vez están equipadas operativamente para pasar de la "intención" al "cumplimiento" en lo que respecta a la gobernanza de la IA, según la escala de madurez de Simon Zadek. Dependiendo del grado de madurez de una región en cuanto a la adopción de la IA, esto podría dificultar aún más los recursos de gobernanza de la IA.

Esbozar las oportunidades que las organizaciones pueden obtener mediante la adopción de la gobernanza de la IA puede ayudar a dar el impulso necesario para iniciar a tiempo los primeros pasos cruciales. Por ejemplo, puede ayudar a

1. Fortalecer su trabajo frente a las normativas sobre IA próximas y futuras.
2. Distinguir a la organización como investigadora y futurista; generar confianza entre las partes interesadas, y mejorar la reputación y el impacto.
3. Mitigar los riesgos económicos futuros: pérdida de datos, infraestructuras, mano de obra y externalidades negativas.
4. Atraer y retener nuevos talentos cualificados en tecnologías emergentes, sobre todo porque cuanto más jóvenes son los talentos, más buscan la alineación de valores en el trabajo.

Aproveche las redes afines para reducir costes y acelerar su curva de aprendizaje:

A pesar de las oportunidades, nos encontramos con obstáculos prácticos a la hora de aplicar la gobernanza de la IA.

Los elevados costes, las prioridades operativas, la falta de conocimientos o recursos para prever escenarios, la incapacidad para visualizar externalidades negativas y la priorización de objetivos a corto plazo se interponen en el camino de las empresas que consideran la gobernanza de la IA entre sus principales prioridades. Por otro lado, las promesas de eficiencia y rentabilidad de la IA resultan

más atractivas para las organizaciones del sector social y los innovadores que trabajan constantemente con recursos limitados y el objetivo de lograr un impacto social a gran escala.

Colaborar con organizaciones de la sociedad civil de "buena tecnología", comunidades y foros RAI, explorar la IA de código abierto y llevar a cabo evaluaciones de madurez/riesgo de la IA utilizando la infraestructura de apoyo de organismos comerciales e industriales podría ayudar a reducir los costes iniciales, asignar recursos y establecer objetivos mientras se tantean las aguas.

La visualización de las externalidades negativas se ve facilitada por varios marcos de gobernanza de la IA (que se comentan más adelante) que ayudan a establecer la higiene de los equipos y los datos, planificar ejercicios de redistribución de tareas, tener en cuenta los sistemas de reparación y las prácticas de explicabilidad como la procedencia y las tarjetas de modelos de IA, estructurar revisiones y auditorías externas con expertos en la materia y las comunidades afectadas, etc. Los informes precedentes y los enfoques recomendados por los gobiernos ayudan a reforzar los argumentos que convencen a inversores y donantes de por qué estas medidas son indispensables para la escala y la eficacia del impacto, así como para su sostenibilidad.

Se trata de conversaciones complejas, cuya gestión exige manejar polaridades, en las que podría utilizarse este útil marco.

Para lograr un futuro digital justo, de principio a fin, que recompense equitativamente a todas las partes interesadas, necesitamos un cambio de paradigma que maximice el impacto social. Esto puede consistir en unirse a los esfuerzos locales y representativos para crear cooperativas de plataforma en lugar de nuevos modelos impulsados principalmente por el crecimiento rápido y los beneficios a corto plazo. En la misma línea, las empresas que aspiran a aportar valor en lugar de conseguir valoraciones irreales se están agrupando como "cebras" en una narrativa contraria a los "unicornios".

Abstenerse de antropomorfizar la IA

Una de las cuestiones más problemáticas es la capacidad de la IA emergente para producir textos, imágenes, vídeos y sonidos realistas y convincentes, incluso cuando distan mucho de ser exactos. Esto, unido a la inclinación humana a antropomorfizar las tecnologías, hace que los productos finales, incluso con la mejor intención, adopten a menudo un rostro, un nombre, un tono o una personalidad humanos. Estos productos están diseñados para obtener una confianza implícita en su

eficacia, pero dificultan que los usuarios finales distingan dónde acaba el acto humano y dónde empiezan las limitaciones de la tecnología. Este enmascaramiento de las limitaciones de la tecnología mediante rasgos humanos podría, por ejemplo, tener consecuencias nefastas en las líneas de ayuda para niños, mujeres u otras comunidades vulnerables, o en las aplicaciones basadas en IA que la gente utiliza para obtener apoyo en salud mental, compañía o asesoramiento astrológico.

Nuestra confianza innata en la precisión empírica de los algoritmos y en su eficacia para reducir los errores humanos conduce, sin saberlo, a la propagación de patrones discriminatorios incluso sin intención, como detalla Cathy O'Neil en su libro "Weapons of Math Destruction" (Armas de destrucción matemática). Esto conduce a la exacerbación de las desigualdades sociales existentes en las soluciones basadas en la IA para la aplicación de la ley, el empleo, los servicios financieros, el bienestar social, la difusión de la información, el bienestar laboral, las injusticias de género, etc.

Abstenerse de antropomorfizar los productos basados en IA, por muy tentador que resulte, es fundamental. Mantener la autodivulgación de los personajes de la IA es crucial para generar confianza con las partes interesadas. Crear conscientemente equipos de diseño y bases de datos representativos puede evitar que los modelos de IA se basen en conjuntos de datos discriminatorios y produzcan falsos positivos y negativos.

Preste mucha atención a la procedencia de sus datos y a los derechos de propiedad

Las empresas que adoptan la IA a menudo tienden a compensar los problemas de derechos de autor a los que se enfrentan los modelos fundacionales con las empresas que los construyen, sin comprender cómo estos pueden tener efectos dominó en nuestros productos o, en la lucha de poder que acompaña al auge de la IA, incluso repercutir en los usuarios. Dado que gran parte de la cuestión está aún en el aire, lo que podemos hacer en este momento es desarrollar buenas prácticas que absorban las ondas expansivas cuando se produzcan (no si se producen).

Algunas de estas mejores prácticas incluyen la creación de registros de procedencia en todo lo que se crea. Los trabajos de la Iniciativa para la Autenticidad de los Contenidos (CAI) y el Centro para la Comunicación Constructiva del MIT muestran cómo conseguirlo mediante herramientas y bases de datos como *Content Credentials* y *Data Provenance Explorer*. Además, las organizaciones deben esforzarse por respetar la legislación vigente sobre derechos de autor y registrar y acreditar las fuentes siempre que sea posible. Es imperativo establecer mecanismos de consentimiento, crédito, control y compensación cuando se obtienen contenidos de artistas.

¿Cómo conseguirlo?

Puede que no sea posible conseguir cien puntos en la RAI desde el primer día. No es un objetivo que alcanzar y seguir adelante, sino más bien un viaje. Abogamos por algunos de los recursos existentes y en continua evolución que guían a las organizaciones en su viaje hacia la gobernanza de la IA.

El marco *AI Blindspot Discovery* desarrollado por el Assembly Program del Berkman Klein Center es un marco completo que propone varios pasos para realizar un análisis exhaustivo del statu quo e iniciar sistemas y procesos que ayuden a establecer la gobernanza de la IA en las fases de planificación, construcción, despliegue y supervisión de un modelo o sistema de IA.

El Responsible AI Resource Kit de NASSCOM es un recurso independiente del sector con aportaciones de múltiples partes interesadas que consiste en principios de RAI, evaluación de madurez de RAI, marco de gobernanza y una guía expansiva para arquitectos de diseño.

Llevar a cabo evaluaciones anuales de la madurez de la RAI utilizando cualquiera de estos marcos ayudaría a evaluar el progreso o el retroceso en cada proceso, lo que permitiría tomar las medidas necesarias para corregir el rumbo. Un proceso Planificar-Hacer-Estudiar-Actuar (ciclo de Deming) en cada fase de prueba mejoraría además la calidad del producto y los procesos, permitiendo una aplicación dinámica de lo aprendido.

¿Cuándo es el momento adecuado?

Un punto muy debatido, el momento adecuado para que cada organización inicie la gobernanza de la IA puede evaluarse utilizando el Dilema de Collingridge. En palabras del propio Collingridge en su libro El control social de la tecnología: "Cuando el cambio es fácil, no se puede prever su necesidad; cuando la necesidad de cambio es evidente, el cambio se ha vuelto caro, difícil y lento". Creemos que una evaluación de los riesgos de la IA o de la madurez de la RAI desde el principio podría servir de brújula para iniciar un sistema de gobernanza de la IA y a la larga merecería la pena. Del mismo modo que las vulnerabilidades en caso de incendio se evalúan mejor como parte del diseño arquitectónico para permitir la planificación de las vías de salida y la instalación de mecanismos de prevención durante la construcción.

Los conocimientos y la experiencia de las personas impulsan su trabajo en sus respectivos campos, y hay un compromiso añadido en el sector social. Es natural que a veces las organizaciones de la sociedad civil, los organismos gubernamentales, las startups, las comunidades tecnológicas, médicas y periodísticas se vuelvan ferozmente territoriales en el trabajo que realizan. El Ecosistema de la Tecnología Responsable (pág. 14, Guía de la Tecnología Responsable, ATIH 2023) subraya la necesidad de abrirnos a conversaciones paralelas en campos afines, pues de lo contrario corremos el riesgo de aislarnos en nuestros planteamientos para la resolución de problemas. Los esfuerzos aislados sólo pueden ser graduales y el verdadero cambio de sistemas exige la colaboración interdisciplinar.

Las consecuencias de un accidente de bicicleta son distintas de las de un accidente de tren. La responsabilidad y la obligación de rendir cuentas son proporcionales a las capacidades de los vehículos.

Del mismo modo, por mucho que haya una carrera en todos los ámbitos para aprovechar la IA, no se trata de una carrera básica, sino más bien de una carrera del limón y la cuchara. De nada sirve llegar el primero si al final se cae el limón a mitad de camino.

Referencias:

Gurteen, David."Trampas multipolares: Actuar contra nuestros intereses colectivos" En Multipolar Traps. https://conversational-leadership.net/multipolar-trap/

All Tech is Human. 2023. “Responsible Tech Guide” https://www.scribd.com/document/476272088/Responsible-Tech-Guide-by-All-Tech-Is-Human

Assembly Program -Berkman Klein Center: https://www.berkmankleinassembly.org/

BCG Gamma. “Are You Overestimating Your Responsible AI Maturity?” March 2021.https://web-assets.bcg.com/b5/4b/8386b5cf409e835bba50306c39d2/slideshow-final-website-version-2021-rev.pdf

Braungart, Michael, and McDonough, William. 2002. *Cradle To Cradle: Remaking the Way We Make Things*. North Point Press.

Calderon, Ania, Taber, Dan, Qu, Hong and Wen, Jeff. "AI Blindspots". https://aiblindspot.media.mit.edu/

Collingridge, David. 1980. *The Social Control of Technology*. Cambridge University Press.

Content Authenticity Initiatives https://contentauthenticity.org/

Content Credentials https://contentcredentials.org/

Data Provenance Explorer https://www.dataprovenance-explorer.org/

Deloitte. 2023 Gen Z and Millennial Survey. https://www.deloitte.com/global/en/issues/work/content/genzmillennialsurvey.html

Karkera, Kiran. "Why is Provenance Important for AI?" July 10, 2020. https://kaal-daari.medium.com/an-example-of-art-provenance-records-for-the-curious-d3a5e4a1dd77

Genus, Audley, and Stirling, Andrew. 2018. "Collingridge and the Dilemma of Control." Research Policy. February 2018. https://www.sciencedirect.com/science/article/pii/S0048733317301622?via%3Dihub

Google. "Model Cards" https://modelcards.withgoogle.com/about

Gurteen, David. "Multipolar Traps: Acting against our collective interests" In Multipolar Traps. https://conversational-leadership.net/multipolar-trap/

Hays, Kali. 2023. "Google, OpenAI, and Microsoft are Blaming Users When Generative-AI Models Show Copyrighted Material". *Business Insider*. Nov 7, 2023. https://www.businessinsider.com/google-openai-microsoft-users-responsible-ai-copyrighted-material-2023-11?IR=T

Hendrycks, Dan, Mazeika, Mantas, and Woodside, Thomas. "An Overview of Catastrophic AI Risks". Center for AI Safety. 9 Oct. 2023. https://arxiv.org/pdf/2306.12001.pdf

IBM. Global AI Adoption Index 2022. May 2022. https://www.ibm.com/downloads/cas/GVAGA3JP

Lu, Yingying. 2023. "AI Will Increase Inequality and Raise Tough Questions About Humanity, Economists Warn". *The Conversation*. April 27, 2023. https://theconversation.com/ai-will-increase-inequality-and-raise-tough-questions-about-humanity-economists-warn-203056

MIT Center for Constructive Communication. "Data Provenance for AI" https://www.ccc.mit.edu/project/data-provenance-for-ai/

NASSCOM. "Responsible AI Resource Kit". https://indiaai.gov.in/responsible-ai/homepage

NITI Ayog. "RESPONSIBLE AI. Approach Document for India: Part 1 –Principles for Responsible AI" February 2021. https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf

NITI Ayog. "RESPONSIBLE AI. Approach Document for India: Part 2 -Operationalizing Principles for Responsible AI." August 2021. https://www.niti.gov.in/sites/default/files/2021-08/Part2-Responsible-AI-12082021.pdf

Center for Creative Leadership. “Are You Facing a Problem or a Polarity?” November 18, 2022. https://www.ccl.org/articles/leading-effectively-articles/are-you-facing-a-problem-or-a-polarity/

Olay. “AI-shu Chatbot” https://ai-shu.in/

O’Neill, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.

Platform Cooperativism Consortium: https://platform.coop/

Saujani, Reshma. 2023. “We Don’t Have to Choose Between Ethical AI and Innovative AI” *Time*. Dec. 5, 2023. https://time.com/6342280/ai-paid-leave-social-good/

Center for Humane Technology. “The A.I. Dilemma”. March 9, 2023. https://www.youtube.com/watch?v=xoVJKj8lcNQ

Tech Stewardship. “How Can We Ensure Tech is Beneficial for All?” 2023. https://techstewardship.com/the-change/

The Authors Guild. 2023. “More than 15,000 Authors Sign Authors Guild Letter Calling on AI Industry Leaders to Protect Writers.” July 18, 2023. https://authorsguild.org/news/thousands-sign-authors-guild-letter-calling-on-ai-industry-leaders-to-protect-writers/

The Consilience Project. “Challenges to Making Sense of the 21st Century”. March 30, 2021.https://consilienceproject.org/challenges-to-making-sense-of-the-21st-century/

The W. Edwards Deming Institute. “Deming Cycle” https://deming.org/explore/pdsa/

Tveit, Alex, Abbott, Mark, and Lajoie, Jason. “Tech Stewardship as a foundation for Multi-Stakeholder Collaboration (MSC) to enable STI4SDGs” https://sdgs.un.org/sites/default/files/2023-05/B47%20-%20Tveit%20-%20Tech%20Stewardship.pdf

Zebras Unite Cooperative: https://zebrasunite.coop/

Zadek, Simon. “The Path to Corporate Responsibility” Harvard Business Review. Dec. 2004. https://hbr.org/2004/12/the-path-to-corporate-responsibility