



Presentación e instructivo de la herramienta fAlr Venture

1. Contexto general

La Inteligencia Artificial (IA) y los Sistemas de Decisión Automatizada (SDA) han adquirido un rol protagónico en nuestra sociedad y se encuentran inmersos en diversos aspectos de la vida cotidiana, la economía y el conocimiento. Según las previsiones de la consultora IDC, la inversión en sistemas basados en IA llegará a los 118.000 millones de dólares en 2022 y sobrepasará los 300.000 millones de dólares para 2026.

A pesar de su potencial transformador y sus posibles beneficios, las tecnologías de IA generan riesgos que pueden impactar de forma negativa a individuos, grupos, organizaciones, comunidades, la sociedad, el medio ambiente, y el planeta. (NIST, 2023).

Un caso emblemático de estos riesgos es el de la empresa de tecnología Amazon, en el cual se identificaron sesgos en el sistema de IA que había comenzado a implementar desde 2014 para automatizar tareas de selección de personal. Dados los sesgos de la base de datos de entrenamiento, el sistema de IA descartaba a aplicantes mujeres para los cargos. En 2015 la empresa abandonó el uso de este sistema de IA.

Organismos internacionales como la UNESCO y la OCDE, gobiernos nacionales y locales, y organizaciones del

sector privado han comenzado a formular marcos de referencia para orientar el desarrollo de estas tecnologías. Estudios recientes han identificado cierto grado de convergencia de estos marcos hacia algunos principios como privacidad, justicia y equidad, responsabilidad, no maleficencia y transparencia.

A nivel normativo, también hay esfuerzos para crear instrumentos que regulen el desarrollo y adopción de estos sistemas. Particular relevancia tiene el proyecto de ley de Inteligencia Artificial (2021) en la Unión Europea que se encuentra actualmente en trámite y la publicación de un Blueprint for an AI Bill of Rights (2022) por parte de la Oficina de Política de Ciencia y Tecnología de los Estados Unidos. Igualmente, en Chile se aprobó la Política Nacional de Inteligencia Artificial en el año 2021 que contiene los lineamientos estratégicos para el desarrollo de estas tecnologías en el país.

No obstante, la rápida evolución de estos sistemas de IA y sus potenciales oportunidades y riesgos, generan la necesidad de complementar estos marcos y regulaciones con herramientas prácticas para que distintos actores puedan contribuir al desarrollo, adopción y uso responsable de la IA.

"Un sistema computacional que puede, para un determinado conjunto de objetivos definidos por humanos, hacer predicciones y recomendaciones o tomar decisiones que influyen en entornos reales o virtuales. Los sistemas de IA están diseñados para operar con distintos niveles de autonomía." (OCDE, 2019)

<https://www.idc.com/getdoc.jsp?containerId=prUS49670322>

[https://www.nist.gov/itl/ai-risk-management-framework#:~:text=AI%20Risk%20Management%20Framework%20\(AI%20RMF%20.0](https://www.nist.gov/itl/ai-risk-management-framework#:~:text=AI%20Risk%20Management%20Framework%20(AI%20RMF%20.0)

Ethics of Artificial Intelligence. Case Studies and Options for Addressing Ethical Challenges. Springer Briefs. 2023. Page 10.

Jobin, A., Ienca, M., Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence.

2. Proyecto Algoritmos Éticos, Responsables y Transparentes (Chile)

El proyecto Algoritmos Éticos, Responsables y Transparentes es una iniciativa multi-actor que ejecuta la Universidad Adolfo Ibáñez (UAI), con apoyo de BID Lab, co-ejecutado por Magical Startups, y del cual son socios el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, ChileCompra, la División de Gobierno Digital, el Consejo para la Transparencia, el Fondo Nacional de Salud (FONASA), el Instituto de Previsión Social (IPS) y el Servicio Civil. Este proyecto busca instalar capacidades y estándares para incorporar consideraciones éticas en la compra y utilización de Inteligencia Artificial (IA) y algoritmos de decisión automatizada (ADA) en agencias estatales (vertical pública), y en la formulación y desarrollo de estas soluciones por parte de los proveedores tecnológicos e inversionistas (vertical privada). En el marco de esta vertical privada, iniciamos la co-creación de una herramienta de IA Ética para que los fondos de Venture Capital (VC) puedan identificar oportunamente riesgos éticos asociados al desarrollo e implementación de tecnologías de IA en las startups que financian, y gestionar proactivamente dichos riesgos.

3. ¿Por qué una herramienta de IA Ética para VCs?

Los VC han desarrollado procesos sofisticados para analizar las startups que buscan financiar, particularmente en lo relativo a la visión estratégica, modelos de negocio, equipo ejecutor, unit economics, financiación previa, entre otras. Adicional a estos aspectos de negocio, la incorporación de criterios ambientales, sociales y de gobernanza (ASG o ESG en inglés) en la evaluación de startups es una tendencia creciente en los VC, pues permite analizar otras fuentes de riesgo (y oportunidad).

Esto cobra particular relevancia en el contexto actual, en el que existe mayor incertidumbre en el mercado debido a los efectos macroeconómicos post-pandemia en conjunción con situaciones de tensión política internacional, lo cual obliga a los VCs a seleccionar con mayor



cuidado sus inversiones. Además, en Chile este año comienza a regir la Norma 461 de la Comisión para el Mercado Financiero (CMF) que obliga a empresas financieras y listadas en la Bolsa a reportar indicadores ASG en sus memorias anuales.

Entre los temas ASG relevantes para la industria de software y tecnologías de la información (TI) están, por ejemplo: cambio climático (emisiones de gases de efecto invernadero), consumo de energía, diversidad (tanto en empleados como en los órganos directivos), bienestar de los empleados y buena gobernanza, entre otros.

No obstante, el uso y despliegue de tecnologías de inteligencia artificial (IA) genera riesgos sistémicos adicionales que ameritan un foco particular por parte de los VC. Adicionalmente, los fondos de VC en Chile y sus Limited Partners (LPs) han empezado a anticipar estas tendencias y riesgos emergentes, conscientes de que se convertirán en requisitos futuros para la expansión de sus startups hacia mercados más maduros como Europa y Estados Unidos.

En este contexto, BID Lab, a través de su iniciativa fAIR LAC ha desarrollado herramientas y guías que promueven un mejor entendimiento de los riesgos asociados a la implementación de tecnologías de IA en América Latina y el Caribe. Construyendo sobre estos avances, consideraremos que un entendimiento claro de los riesgos éticos de la IA por parte de los VC puede:

Contribuir a una comprensión más integral del valor de las inversiones que realizan.

Facilitar procesos de internacionalización de las startups en las que invierten.

Contribuir a mitigar riesgos sociales de los sistemas de IA y catalizar sus efectos positivos.

<https://goblab.uai.cl/algoritmos-eticos/>

United Nations. Principles for Responsible Investment (2022). Starting up. Responsible investment in venture capital.

World Economic Forum (2022). ESG Pulse Check: Getting the Basics Right for Start-ups and Venture Capital Firms.

<https://www.cmfchile.cl/portal/prensa/615/w3-article-49804.html>

[https://www.sasb.org/standards/materiality-finder/find/?industry\[0\]=TC-SI](https://www.sasb.org/standards/materiality-finder/find/?industry[0]=TC-SI)

<https://fairlac.iadb.org/>

<https://publications.iadb.org/es/autoevaluacion-etica-de-ia-para-actores-del-ecosistema-emprendedor-guia-de-aplicacion>

4. Proceso de co-creación

El proceso de desarrollo de la Herramienta de IA Ética para VC fue concebido desde el inicio como una co-creación con los usuarios objetivo, buscando entender sus prácticas actuales y expectativas, para maximizar su potencial adopción. En el proceso de co-creación participaron los siguientes actores clave:

- Fen Ventures.
- Dadneo (Fondo Vulcano).
- Magical (Administradora de fondos de inversión)
- The Yield Lab.
- Chile Global Ventures (Fondo Clin).

Para iniciar el proceso, elaboramos una versión inicial de la herramienta partiendo de un referenciamiento de marcos, metodologías y herramientas existentes para analizar aspectos de IA Ética en VCs y startups. Ejemplos de estos marcos son:

- BID Lab (2021). Auto-evaluación ética para el ecosistema emprendedor.
- OCDE (2022). Framework for the Classification of AI Systems.
- NIST (2023). Artificial Intelligence Risk Management Framework.
- Dotan, R. (2022). Responsible Investment in AI. A Guidebook for VCs.
- Harvard Kennedy School of Government.
- Belfer Center. Technology and Public Purpose Project

Una vez desarrollada la versión inicial, surtimos el siguiente proceso con los fondos participantes:

- Entrevistas iniciales de entendimiento de los procesos de cada VC.
- Sesión de trabajo conjunta para realizar una nivelación conceptual.
- Sesiones de trabajo para revisar la primera versión de la herramienta y recibir feedback.
- Pilotaje en un startup del portafolio de tres VCs: Fen Ventures, Dadneo y Magical.

Este proceso nos mostró la acogida positiva de este tipo de herramientas aplicadas y la importancia de abrir una conversación alrededor de temas de IA Ética entre el VC y las startups. También nos ayudó a simplificar su contenido y nos condujo a aclarar conceptos para facilitar su adopción.

5. ¿Cómo funciona la herramienta?

5.1. Dimensiones de la herramienta: Sociedad, Solución y Sistema

La herramienta se alinea con las dimensiones del marco S3 (Sociedad, Sistema y Solución) de BID Lab, aquellas detallamos a continuación:

a) Dimensión Sociedad

En esta dimensión, las preguntas buscan comprender a los usuarios del sistema de IA y los grupos impactados por la tecnología. Se evalúa el grado de notificación que genera el sistema sobre la interacción con los usuarios y su potencial impacto negativo en áreas relevantes para el bienestar humano, entre otros aspectos.

b) Dimensión Solución

Las preguntas en esta dimensión buscan entender las áreas de aplicación y los usos del sistema de IA, así como la criticidad de las actividades relacionadas y el alcance de su despliegue, entre otros factores.

b) Dimensión Sistema

En esta dimensión, las preguntas se centran en aspectos técnicos de la tecnología, como las características de los datos de entrada del modelo y la base de datos de entrenamiento, la adopción de prácticas de evaluación de sesgos y de transparencia y explicabilidad, y el grado de autonomía del modelo, entre otros aspectos técnicos.



<https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm>

[https://www.nist.gov/itl/ai-risk-management-framework#:~:text=AI%20Risk%20Management%20Framework%20\(AI%20RMF%201.0\)](https://www.nist.gov/itl/ai-risk-management-framework#:~:text=AI%20Risk%20Management%20Framework%20(AI%20RMF%201.0))

<https://www.ravidotan.com/ai-due-diligence>

Harvard Kennedy School (2020). Responsible Investing in Tech and Venture Capital. Advancing Public Purpose in Frontier Technology Companies. Belfer Center for Science and International Affairs. Technology and Public Purpose Project (TPPP). August 2020

Gráfica - Dimensiones y temas de la Herramienta (Versión 1.0)

Fuente: Elaboración propia



5.2. Preguntas y opciones de respuesta

La herramienta contiene **21 preguntas** cortas con opciones de respuesta predefinidas, en su mayoría adaptadas y traducidas del OCDE (2022) Framework for the Classification of AI Systems. Estas preguntas están asociadas a las características de sistemas de IA confiables propuestas por el National Institute of Science and Technology - NIST (2023) en su Artificial Intelligence Risk Management Framework.

5.3. Niveles de riesgo

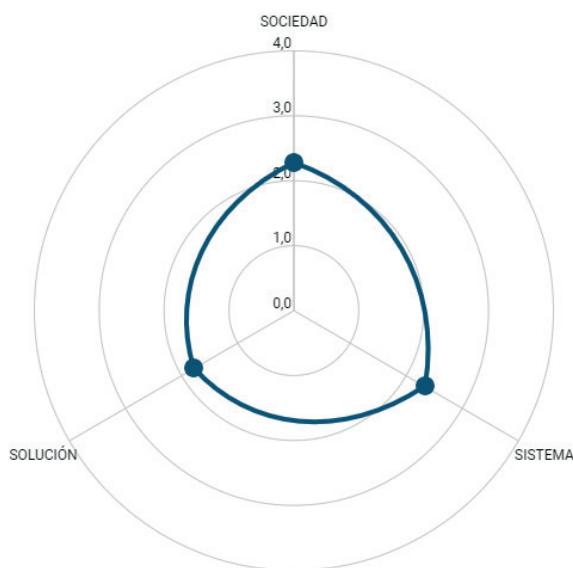
Cada opción de respuesta tiene un nivel de riesgo correspondiente, desde bajo hasta muy alto. Estos niveles de riesgo se utilizan para otorgar una calificación general del nivel de riesgo de la startup.

LEAME (READ ME)		fAIr Venture				
Dimensión	Tema	#	Descripción	Opciones de respuesta	Característica NIST Asociada	Riesgo
SOCIEDAD	Usuarios del sistema de IA	1	¿Cuál es el nivel de competencia de los usuarios directos que interactúan con el sistema de IA?	Experto en IA	Responsabilidad y Transparencia	Medio
	Grupo impactado	2	¿Qué grupo se verá principalmente impactado por el sistema?	Amateur	Responsabilidad y Transparencia	Alto
	Información	3	¿El sistema de IA informa a los usuarios o grupos impactados sobre su interacción con dicho sistema?	Profesional entrenado pero no experto en IA	Explicabilidad e Interpretabilidad	Alto
	Opcionalidad	4	¿Pueden los grupos impactados por el sistema de IA escoger no participar, cuestionar o revertir sus resultados?	Experto en IA	Responsabilidad y Transparencia	Bajo
	Derechos Humanos	5	¿Pueden los resultados del sistema de IA impactar Derechos Humanos fundamentales como los listados a continuación?	Otro	Beneficencia (Safety)	Alto
	Bienestar, sociedad y medio ambiente	6	¿Cuál es el principal ámbito de la vida diaria donde el sistema de IA produce efectos?	Salud física o mental	Beneficencia (Safety)	Alto
	Potencial de desplazamiento	7	¿Podría el sistema automatizar tareas que actualmente ejecutan humanos?	Nulo potencial de automatización	Justicia y Equidad	Bajo
SOLUCIÓN	Áreas específicas de aplicación	8	¿El sistema de IA se aplica(rá) en alguna de las siguientes áreas?	Identificación biométrica y categorización de personas físicas	Valididad y Confianza	Alto
	Tipo de uso	9	¿El sistema será utilizado para alguno de los siguientes propósitos?	Utiliza(rá) o habilita(rá) sistemas de identificación biométrica remota en tiempo real	Valididad y Confianza	Muy Alto
	Función del negocio	10	¿En qué función del negocio se va a implementar el sistema (ventas, servicio al cliente, etc.)?	Ventas	Beneficencia (Safety)	Alto
	Criticidad de las actividades	11	¿Una disrupción en la funcionalidad o actividad del sistema podría afectar una actividad crítica, según la normatividad aplicable?	No se despliega en actividades críticas ni presta servicios a ellas	Seguridad y Resiliencia	Bajo
	Tipo de modelo	12	¿Qué tipo de modelo utiliza el sistema de IA?	Discriminativo	Explicabilidad e Interpretabilidad	Bajo
	Alcance del despliegue	13	¿Se planifica que el despliegue del sistema de IA llegue a tener un alcance: acotado, amplio o generalizado?	Acotado (ej.: una compañía o una región pequeña)	Responsabilidad y Transparencia	Medio

5.4 Diagrama radial

Además de la calificación general, se genera un diagrama radial que muestra las calificaciones de riesgo en las tres dimensiones (Sociedad, Solución y Sistema). Esto proporciona una representación visual de los riesgos en cada área.

Gráfica de radar con base en respuestas:



5.5 Interpretación de resultados

La herramienta permite evaluar de manera integral las startups considerando aspectos clave relacionados con la IA y proporciona una calificación general y visualización de **potenciales riesgos** en las dimensiones relevantes. El nivel de riesgo está altamente correlacionado con la naturaleza e industria en la cual se aplique la herramienta.

6. Recomendaciones y limitaciones:

Sugerimos que el análisis lo realice el VC con base en la situación actual del startup y del sistema de IA, y no con base en estados futuros posibles. Esto permitirá tener una mirada más objetiva y facilitará la discusión. No obstante, si el VC decidiera invertir en la startup, la herramienta se podría aplicar en distintos momentos del tiempo y podría ayudar a monitorear el progreso en las distintas dimensiones.

Sugerimos que los resultados de la herramienta no impliquen una decisión de inversión en la startup. Es decir, niveles altos de riesgo en la herramienta no necesariamente se deben traducir en una decisión de no inversión. Proponemos que la herramienta sea un insumo para abrir una conversación informada entre VC y startup.

Recomendamos que durante el proceso de aplicación de la herramienta haya al menos un espacio de conversación entre el VC y la startup alrededor de las dimensiones y preguntas. En la experiencia del pilotaje identificamos que el debate y la reflexión pueden generar valor tanto para la startup como para el VC.

Consideramos que no es imperativo que los VC involucren personal especializado (desarrolladores, asesores o expertos en IA) para aplicar la herramienta. El ejercicio de pilotaje lo realizamos con Investment Officers (IO) de los fondos con resultados positivos. Sin embargo, un contexto general sobre sistemas de IA, sus riesgos y sus oportunidades, pueden ser de utilidad para sacar mejor provecho de este ejercicio. Recomendamos en particular estas dos guías:

BID (2020). Adopción ética y responsable de la Inteligencia Artificial en América Latina y el Caribe.

World Economic Forum (2022). Empowering AI Leadership: AI C-Suite Toolkit. WEF.

Si bien la herramienta proporciona una experiencia de usuario y arroja promedios y un resultado gráfico, su aplicación puede ser flexible tomando solo algunas de las preguntas o ajustando las opciones de respuesta y sus niveles de riesgo para reflejar la realidad del VC y/o de sus startups. No obstante, el equipo del proyecto se encuentra evaluando alternativas para la maduración y escalamiento del proyecto que ayudarán a aclarar posibles casos de uso.

Queremos seguir mejorando y evolucionando la herramienta. Esta versión 1.0 se encuentra abierta a comentarios y recomendaciones. Para mayor información contacta a vanessa@magicalstartups.com.

Contribuciones:

La coordinación del proceso de co-creación estuvo a cargo del equipo del proyecto Algoritmos Éticos, en particular Cesar Rosales, Tetsuro Narita y Carolina Carrasco (BID Lab), María Paz Hermosilla, Romina Garrido y Reinel Tabares (Universidad Adolfo Ibáñez), Juan Orlandi, Teresita Saavedra y Vanessa Neble (Magical Startups), Marcos Maldonado (Ministerio de Ciencia - Chile) y Juan Manuel Ramírez (consultor del proyecto). Agradecemos por su tiempo y dedicación a Andrea Araneda y Paola Negrín (Fen Ventures), Sebastián Ibáñez (Dadneo), Camila Petignat (The Yield Lab), Carolina Alarcón (Red Chile Global Ventures), Chris Thompson (Magical), Constanza Corvalán y Michelle Viscaíno (Simbiótica) y Diego Borquez (Charly.io). Agradecemos por su tiempo y retroalimentación al equipo de BID Lab, en particular a Cesar Buenadicha, Mara Balestrini y Gyoung Joo Choe.