

Expresiones de intención: conjuntos de datos de aprendizaje automático para obtener mejores resultados en la atención médica

Lacuna Fund: nuestra voz en los datos

15 de junio de 2021

Índice

1: INTRODUCCIÓN.....	2
PROPÓSITO Y METAS DEL FONDO	2
PRINCIPIOS DEL FONDO	2
FILOSOFÍA DE LA CONCESIÓN DE SUBVENCIONES	3
2: DESCRIPCIÓN GENERAL	3
ELEGIBILIDAD ORGANIZACIONAL	3
PROCESO DE SELECCIÓN Y CRITERIOS DE EVALUACIÓN	4
TIMELINE.....	5
3: PROPÓSITO Y NECESIDAD	6
PROPÓSITO.....	6
NECESIDAD	6
4: INFORMACIÓN SOBRE EXPRESIONES DE INTENCIÓN	7



This document is licensed under a [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0) license.

1: Introducción

Propósito y metas del fondo

Lacuna Fund apoya la creación, expansión y mantenimiento de conjuntos de datos de capacitación y evaluación equitativos que permitan la aplicación sólida de herramientas de aprendizaje automático con un alto valor social.

Los objetivos del fondo son los siguientes:

- Desembolsar fondos a las instituciones para crear, expandir o mantener conjuntos de datos que subsanen las brechas y reduzcan el sesgo en los datos etiquetados utilizados para el aprendizaje automático.
- Hacer posible que las poblaciones desatendidas aprovechen los avances que ofrece la inteligencia artificial.
- Profundizar la comprensión por parte de las comunidades de aprendizaje automático y filantropía sobre cómo financiar de manera más efectiva y eficiente el desarrollo y mantenimiento de conjuntos de datos etiquetados de manera equitativa.

Principios del fondo

Los siguientes principios guiarán la gobernanza y las operaciones de Lacuna Fund.

- **Accesibilidad:** Lacuna Fund se compromete a garantizar que los conjuntos de datos creados a través de su financiación sean accesibles y beneficien a las comunidades desatendidas al servicio de los objetivos descritos anteriormente. Los conjuntos de datos y la propiedad intelectual relacionada utilizarán las licencias de datos abiertos apropiadas para maximizar el uso responsable posterior. (Consulte la política de IP para obtener detalles adicionales).
- **Equidad:** Lacuna Fund tiene como objetivo hacer que la inteligencia artificial sea más equitativa al respaldar conjuntos de datos que se crean y responden a las necesidades de aquellos con identidades subrepresentadas a nivel mundial. Estos conjuntos de datos no deben crear ni reforzar sesgos (por ejemplo, deben incluir el género y representar a las personas de color a nivel mundial), ni deben respaldar sistemas o tecnologías que sean perjudiciales.
- **Ética:** Lacuna Fund financiará la recopilación de datos de manera coherente con los estándares laborales éticos, a la vez que exigirá a los destinatarios que especifiquen los pasos que tomarán para proteger la privacidad y evitar daños en la recopilación, concesión de licencias y uso de conjuntos de datos creados con fondos de subvenciones.
- **Enfoque participativo:** Lacuna Fund se esfuerza por satisfacer las necesidades de las partes interesadas afectadas, ya que fomenta el liderazgo o el fuerte compromiso de los expertos locales, los beneficiarios y los usuarios finales en la gobernanza del fondo y en los proyectos respaldados. El fondo considerará la participación de manera consistente con nuestros principios de equidad y ética.

- **Calidad:** los datos generados por los esfuerzos financiados por Lacuna deben ser de alta calidad, y permitir aplicaciones beneficiosas en la investigación, las comunidades y la industria.
- **Impacto de transformación:** Lacuna Fund tiene como objetivo desbloquear los avances que ofrece la inteligencia artificial para las comunidades pobres y desatendidas mediante la financiación de conjuntos de datos que abordan las brechas fundamentales en la inteligencia artificial.

Filosofía de la concesión de subvenciones

Lacuna Fund valora un enfoque colaborativo y local para la creación, la expansión y el mantenimiento de datos. Reconocemos que la utilidad y el mantenimiento continuos de los datos abiertos provienen de una comunidad que invierte en esos datos.

Lacuna Fund espera financiar conjuntos de datos que contribuyan a múltiples aplicaciones con un alto valor social, ya sea a través de investigación, innovación comercial o servicios mejorados del sector público. **Si bien la “Sección 3: Propósito y necesidad” establece las necesidades identificadas por el Panel Asesor Técnico (TAP, por sus siglas en inglés), Lacuna Fund da la bienvenida a ideas novedosas dentro del área de dominio que tienen un beneficio claramente articulado alineado con los criterios de selección que se enumeran a continuación.**

Lacuna Fund cuenta con el respaldo de The Rockefeller Foundation, Google.org y el Centro de Investigación para el Desarrollo Internacional de Canadá. Además, esta convocatoria con respecto a conjuntos de datos etiquetados de manera equitativa para obtener mejores resultados en la atención médica cuenta con el respaldo de Wellcome Trust y The Gordon and Betty Moore Foundation.

2: Descripción general

Elegibilidad organizacional

Lacuna Fund tiene como objetivo hacer que su financiación sea accesible para tantas organizaciones como sea posible en el espacio de la inteligencia artificial para el bien social y cultivar la capacidad y las organizaciones emergentes en el campo.

Para ser elegibles para recibir financiación, las organizaciones deben cumplir con lo siguiente:

- Ser una entidad sin fines de lucro, una institución de investigación, una empresa social con fines de lucro o un equipo de dichas organizaciones. Las personas deben postularse a través de un patrocinador institucional. Se anima encarecidamente a las asociaciones; sin embargo, solo el solicitante principal recibirá fondos.
- Tener una misión que respalde el bien social, ampliamente definida.
- Tener su sede o una asociación sustancial en el país o región donde se recopilarán los datos.
- Tener todas las aprobaciones nacionales o de otro tipo necesarias para realizar la investigación propuesta, así como los acuerdos o planes de uso de datos para asegurarlos. El

proceso de aprobación puede realizarse en paralelo con la solicitud de subvención, si es necesario. Los costos de aprobación, si los hubiera, son responsabilidad del solicitante.

- Tener la capacidad técnica, o la capacidad de desarrollar esta capacidad a través de una asociación descrita en la expresión de interés (EOI, por sus siglas en inglés), para llevar a cabo el etiquetado, la creación, la incorporación, la expansión o el mantenimiento de conjuntos de datos, incluida la capacidad de aplicar las mejores prácticas y los estándares establecidos en el dominio específico (por ejemplo, resultados en la atención médica) para permitir que múltiples entidades realicen análisis de inteligencia artificial/aprendizaje automático de alta calidad.

Proceso de selección y criterios de evaluación

Lacuna Fund busca escuchar a las organizaciones que estén interesadas en responder a una convocatoria de propuestas para desbloquear, crear, agregar o mejorar conjuntos de datos etiquetados que puedan respaldar resultados más equitativos en la atención médica. Lacuna Fund y sus socios realizarán una selección inicial de EOI para determinar la elegibilidad organizacional y la viabilidad del concepto original presentado en la EOI. Después de la selección inicial, un Panel Asesor Técnico de expertos en el dominio, usuarios de datos y partes interesadas evaluará las EOI en función de los criterios de selección que se describen a continuación. Lacuna Fund luego invitará a un pequeño número de organizaciones a proporcionar una propuesta completa basada en la presentación de la EOI. Los miembros del Panel Asesor Técnico no pueden presentar una EOI o una propuesta en respuesta a una solicitud de propuesta (RFP, por sus siglas en inglés) cuya revisión tienen asignada (consulte la [Política de conflicto de intereses](#) de Lacuna Fund).

El Panel Asesor Técnico para esta convocatoria evaluará las EOI para determinar una lista breve de organizaciones que luego serán invitadas a presentar propuestas completas de financiación. Las selecciones para la lista breve de solicitantes invitados se basarán en el grado en que cumplan con los siguientes criterios:

- **Calidad:** la organización o el equipo que propone el proyecto demuestra una colaboración interdisciplinaria entre expertos calificados en salud, investigación, aprendizaje automático y gestión de datos. Se fomentan las asociaciones entre sistemas sanitarios con recursos suficientes y una sólida capacidad de investigación y sistemas que atienden a poblaciones marginadas.
- **Impacto de transformación:** los conjuntos de datos deberían mejorar el aprendizaje automático y permitir un beneficio social demostrable para las poblaciones desatendidas. Los ejemplos incluyen, pero no se limitan a lo siguiente: a) etiquetado, limpieza, recopilación de datos de validación o agrupación de conjuntos de datos existentes para desbloquear un valor adicional o garantizar una mayor precisión en el conjunto de datos existente; b) creación de un nuevo conjunto de datos etiquetado con un alto valor para una población desatendida o un problema relacionado con los resultados en la atención médica; c) facilitación de un conjunto de datos existente para que sea más representativo e inclusivo de raza, género, etnia, capacidad, etc.; o d) vinculación de conjuntos de datos clínicos existentes con datos sobre determinantes sociales de la salud para crear conjuntos de datos de información más sólidos.

- **Equidad:** existe una teoría convincente del cambio que demuestra cómo se aplicará el conjunto de datos para ayudar a las comunidades vulnerables y desatendidas.
- **Enfoque participativo:** los conjuntos de datos deben centrar las necesidades de las comunidades afectadas y trabajar con los socios para identificar los beneficios para la comunidad. Específicamente, los proyectos deben involucrar a pacientes, médicos y miembros de la comunidad en las decisiones de gobernanza de datos (por ejemplo, qué datos se seleccionan y cómo se utilizan). Si el conjunto de datos tiene un alcance geográfico, el equipo está ubicado predominantemente en el área respectiva o mantiene vínculos estrechos con los actores locales para garantizar el mantenimiento y el uso sostenidos del conjunto de datos por parte de la comunidad local.
- **Ética:** el proyecto puede aprobar una selección ética (por ejemplo, una junta de revisión institucional) que investiga: a) preocupaciones de privacidad; b) potencial de uso indebido posterior; c) posibles vectores de discriminación (por ejemplo, género); y d) condiciones de trabajo justas y equitativas, en caso de que haya etiquetadores pagos involucrados en el proyecto.
- **Eficiencia:** el proponente ha considerado los conjuntos de datos existentes y sugiere utilizar técnicas y herramientas eficaces de recopilación y etiquetado de datos para acelerar la recopilación, la limpieza y el intercambio de datos.
- **Factibilidad:** el proyecto es factible en relación con el presupuesto y el alcance del trabajo propuesto. Si bien no se espera que la EOI incluya un presupuesto detallado, sí esperamos que las EOI incluyan una estimación amplia del presupuesto general esperado para completar el proyecto.
- **Accesibilidad:** el conjunto de datos será ampliamente accesible bajo licencia de código abierto. Si esto no es posible, se presenta un caso convincente para licencias más restrictivas a fin de proteger la privacidad o evitar daños, junto con un mecanismo para proporcionar acceso bajo la licencia propuesta.
- **Sostenibilidad:** el proyecto tiene un plan para garantizar la sostenibilidad y el mantenimiento futuro del conjunto de datos, por ejemplo, por una comunidad dedicada o un grupo de partes interesadas (con fines de lucro o sin fines de lucro) y un modelo de gobernanza sólido para el conjunto de datos abiertos.

Timeline

Convocatoria de EOI hecha pública en el sitio web de Lacuna Fund	15 de junio de 2021
Fecha límite de preguntas y respuestas Envíe sus preguntas a secretariat@lacunafund.org	25 de junio de 2021
Publicación de respuestas	2 de julio de 2021
Vencimiento de solicitudes	15 de julio de 2021

Período de preguntas y respuestas: todas las preguntas relacionadas con la EOI deben enviarse a secretariat@lacunafund.org con “Pregunta sobre equidad y salud 2021” en la línea de asunto. Las preguntas enviadas antes del 25 de junio serán desidentificadas y respondidas públicamente el 2 de julio en el sitio web de Lacuna Fund en un documento publicado en la [página “Postularse”](#).

3: Propósito y necesidad

Propósito

El propósito de esta convocatoria de EOI es identificar proyectos para presentar propuestas completas a fin de desarrollar conjuntos de datos de capacitación y evaluación abiertos y accesibles para aplicaciones de aprendizaje automático (ML, por sus siglas en inglés) que abordan las inequidades en los resultados en la atención médica. Una mayor implementación de modelos de ML tiene el potencial de proporcionar mejor información a los médicos en el punto de atención al paciente. Al proporcionar más datos e información a los médicos, dichos profesionales pueden tomar mejores decisiones sobre los diagnósticos de los pacientes y las opciones de tratamiento, al tiempo que comprenden los posibles resultados y el costo de cada uno. El valor del aprendizaje automático en la atención médica es su capacidad para procesar enormes conjuntos de datos más allá del alcance de la capacidad humana y luego convertir de manera confiable el análisis de esos datos en conocimientos clínicos que ayudan a los médicos a planificar y brindar atención, lo que en definitiva conduce a mejores resultados, menores costos atención y mayor satisfacción del paciente.

Ha surgido una preocupación importante con la implementación de ML con respecto al sesgo en los conjuntos de datos que informan los algoritmos en los que se basan estos sistemas. Los conjuntos de datos de atención médica a menudo carecen de una representación amplia entre los grupos demográficos y socioeconómicos. Cuando se aborda esta representación, los datos también pueden no ser totalmente precisos, lo que da como resultado una comprensión incorrecta de las diferencias entre esos grupos. Finalmente, los datos actuales pueden pasar por alto importantes datos socioeconómicos, ambientales y de otra índole que pueden determinar los resultados de salud. La mejora de estos conjuntos de datos en esta línea, por lo tanto, puede mejorar los modelos de aprendizaje automático y los resultados de salud.

Necesidad

Lacuna Fund busca EOI de organizaciones calificadas para desarrollar conjuntos de datos de capacitación y evaluación abiertos y accesibles para aplicaciones de ML que aborden las inequidades en los resultados en la atención médica. El alcance del esfuerzo es el siguiente:

Alcance geográfico: dentro de los EE. UU. y en los países de bajos y medianos ingresos (LMIC, por sus siglas en inglés), con 1 o 2 propuestas y entre el 30 % y el 70 % del grupo que se financiará en cada una de estas áreas geográficas, según la cantidad de conjuntos de datos que se financien.

Alcance sustantivo: los proyectos deben centrarse en conjuntos de datos que puedan utilizarse para abordar una disparidad de salud en una o más de las siguientes áreas: cáncer, enfermedades infecciosas o enfermedades crónicas. Además, los solicitantes pueden argumentar por qué un conjunto de datos en otra área podría reducir significativamente las desigualdades. En cualquier caso, se anima a los solicitantes a considerar si el conjunto de datos que proponen podría utilizarse para abordar otros problemas, además del caso de uso previsto.

Alcance de diversidad: nuestro objetivo es apoyar la creación, el aumento o la incorporación de conjuntos de datos que sean representativos de las poblaciones afectadas y, por lo tanto, estén menos sesgados y tengan más probabilidades de generar resultados de salud equitativos. Dadas las desigualdades históricas relacionadas con la raza en los EE. UU., estamos interesados en conjuntos de datos que podrían ayudar a reducir las disparidades raciales en los resultados en la atención médica en dicho país, y conjuntos de datos que pueden mitigar las desigualdades en los resultados en la atención médica relacionados con la identidad en los LMIC (por ejemplo, etnia, afiliación tribal, género, etc.). Además, sabemos que la falta de diversidad en términos de sexualidad, edad, capacidad, ubicación geográfica e incluso el tipo de entorno de atención (por ejemplo, sistema de salud comunitario frente a un gran sistema médico académico; pacientes hospitalizados frente a pacientes ambulatorios), puede hacer que un conjunto de datos sea menos representativo y conduce a resultados injustos para subconjuntos de una población. Alentamos a los solicitantes a considerar la posibilidad de adoptar un enfoque interseccional al incluir datos para múltiples grupos desatendidos, o explicar por qué un cierto tipo de diversidad es importante para garantizar resultados equitativos para un caso de uso específico.

Las consideraciones adicionales incluyen lo siguiente:

Conjuntos de datos existentes frente a nuevos: vemos un gran valor en desbloquear, aumentar o agregar conjuntos de datos existentes y también estamos abiertos a propuestas para crear nuevos conjuntos de datos, como conjuntos de datos de referencia para probar la generalización de algoritmos y las diferencias en la calidad de la predicción para poblaciones desatendidas. Lo más importante es que queremos asegurarnos de que el conjunto de datos subyacente no tenga fallas fundamentales para evitar la perpetuación del sesgo. Todos los enfoques que se describen a continuación resultan de interés:

- Agrupar datos existentes de sistemas de salud, compañías de seguros de salud o intermediarios de datos de salud para que los investigadores puedan revelar y corregir sesgos algorítmicos (por ejemplo, MIMIC, PhysioNet).
- Subsancar las brechas/hacer que los conjuntos de datos existentes sean más representativos (por ejemplo, imágenes de cáncer de piel, radiografías de pulmón para detectar el COVID-19, etc.).
- Vincular conjuntos de datos clínicos existentes con datos sobre determinantes sociales de la salud para crear conjuntos de datos informativos más sólidos.
- Limpiar los conjuntos de datos existentes para garantizar la precisión de los datos sobre raza, etnia, género, discapacidad, etc.

Tipo de conjunto de datos: los solicitantes seleccionados propondrán la creación, expansión o incorporación de conjuntos de datos que se puedan utilizar para abordar múltiples preguntas de investigación y corregir un sesgo o una disparidad.

4: Información sobre expresiones de intención

Las expresiones de intención solo se aceptarán a través del portal de solicitud disponible en www.lacunafund.org/apply. Una descripción de las preguntas de solicitud está disponible a continuación solo para información. *Limite su expresión de interés a 4 páginas sin incluir referencias, con márgenes de 2,5 cm y una fuente mínima de 11 puntos. No se pueden revisar los apéndices o el material narrativo de la propuesta de más de 4 páginas.*

La siguiente información es obligatoria:

- Información del solicitante
 - Nombre de las organizaciones participantes
 - Descripciones y calificaciones de las organizaciones participantes: brinde una breve descripción de la misión de las organizaciones participantes, los servicios prestados y los distritos electorales atendidos; cómo cumplen con los criterios de elegibilidad articulados anteriormente; y las calificaciones únicas del solicitante para realizar el trabajo propuesto, incluida la experiencia en el desarrollo y el intercambio de conjuntos de datos de salud/atención médica.
- Resumen del propósito de la EOI
 - Identificación de problemas y soluciones propuestas: describa la disparidad o inequidad en la salud, quién se ve más afectado por este problema y la solución propuesta (etiquetado, incorporación, creación, aumento o mantenimiento del conjunto de datos).
 - Metodología general: proporcione una breve descripción general de los pasos propuestos (y los supuestos clave) para desarrollar e implementar el conjunto de soluciones. Explique los permisos vigentes o los pasos que tomará para asegurar los acuerdos de uso de datos.
 - Resultados y beneficios esperados después de la implementación del proyecto: explique cómo el proyecto propuesto contribuirá a lograr el impacto deseado. Si corresponde, describa cómo los productos podrían motivar caminos múltiples y duraderos de investigación o aplicación comercial.
- Plazo y presupuesto generales para la ejecución del proyecto:
 - Plazo: comparta un plazo amplio para completar los pasos incluidos en la Metodología general anterior, incluido el número total de meses necesarios para completarlos.
 - Presupuesto: proporcione una descripción general amplia del presupuesto esperado para completar los pasos incluidos en la Metodología general anterior. Los presupuestos deben presentarse en dólares estadounidenses. El fondo total disponible es de aproximadamente USD 800 000. Anticipamos presupuestos propuestos en el rango de USD 10 000 a 100 000 para proyectos pequeños a medianos y hasta USD 500 000 para proyectos grandes y complejos. Prevemos poder financiar de 1 a 2 grandes proyectos o una mayor cantidad de proyectos más pequeños.