

Seguridad de la IA, estrategia y liderazgo comunitario

Hispanic AI Safety Institute | Programa de formación universitaria | Edición para participantes 1.1, septiembre de 2026

La pregunta que nos guía

¿Cómo podemos lograr que una IA cada vez más capaz beneficie a las personas, preservando el control humano, la rendición de cuentas y la capacidad de recuperarnos cuando los sistemas fallan?

Este programa de 25 horas prepara a estudiantes para evaluar afirmaciones sobre la IA avanzada, explicar cómo podría producirse un daño, comparar salvaguardas técnicas e institucionales y proponer una contribución concreta. Aplicarán estas habilidades a una institución ficticia que presta servicios a la comunidad y después a un problema de su elección.

A quién está dirigido

Estudiantes universitarios, incluidos quienes cursan en community colleges o se transfieren a universidades de cuatro años, de informática, ingeniería, ciencias de la salud, ciencias sociales, educación y otras disciplinas. No exige programación ni matemáticas avanzadas. El piloto está diseñado para personas de 18 años o más. Las ediciones en inglés y español comparten objetivos y criterios de evaluación. Se puede escribir y participar en cualquiera de los dos idiomas.

La convocatoria de HASI se centra en las comunidades hispanas y reconoce su diversidad de trayectorias, idiomas y experiencias. Participar no exige compartir una predicción sobre la IA ni una postura política determinada.

Qué se llevan los participantes

Un registro de afirmaciones y evidencia; una evaluación de capacidades; un mapa causal de riesgos; una propuesta de defensas por capas; y un informe breve de decisión con un plan de acción de 30 días.

Modalidad

Cinco semanas o cinco días. Cada módulo incluye 180 minutos de preparación y trabajo autónomo más una sesión facilitada de 120 minutos, con una pausa de cinco minutos. Dedicación programada total: 25 horas, incluidos 25 minutos de pausas durante el curso. La profundización opcional es adicional. Grupo recomendado: 8 participantes con una persona facilitadora; dividir cohortes mayores en grupos de 6–10.

Desarrollado por HASI a partir del currículo de Estrategia de IAG y los recursos de facilitación compartidos por BlueDot Impact. Es una adaptación independiente de HASI. No otorga certificación de BlueDot ni créditos universitarios.

Ruta de aprendizaje y horarios

Cinco módulos, cinco resultados observables

1. Futuros y decisiones. Distinguir evidencia, predicciones y valores; identificar al menos cuatro actores y sus incentivos. Entrega: nota de decisión W1.
2. Capacidades e incertidumbre. Explicar el papel del cómputo, los datos, los algoritmos y las condiciones de implementación; identificar dos límites de una afirmación sobre capacidades. Entrega: ficha de evaluación W2.
3. Vías de daño. Construir una cadena causal de al menos cuatro pasos, explicitar supuestos e identificar un punto de intervención. Entrega: mapa de riesgos W3.
4. Defensas y gobernanza. Comparar tres salvaguardas con responsables, modos de fallo, costos o tensiones y una regla de suspensión. Entrega: plan de defensa W4.
5. Contribuir responsablemente. Defender una decisión viable mediante evidencia, una alternativa y una incertidumbre explícita. Entrega: informe final W5 y plan de acción.

Opción de cinco semanas

Las semanas 1–5 corresponden a los módulos 1–5. Completar el bloque de preparación de 180 minutos antes de cada reunión semanal. Horario ilustrativo: 17:00–19:00 en la zona horaria anunciada por la institución. Entregar W1–W4 al cerrar cada sesión; revisar el informe final durante la sesión 5. Acordar las fechas reales antes de convocar.

Opción intensiva de cinco días

Los días 1–5 corresponden a los módulos 1–5. Cada día: 09:00–10:30 trabajo autónomo; 10:30–10:45 descanso adicional; 10:45–12:15 trabajo autónomo; 12:15–13:00 almuerzo; 13:00–15:00 sesión facilitada. Son seis horas transcurridas y cinco horas programadas de curso por día. No exige tareas la noche anterior. El trabajo autónomo puede realizarse en una sala de estudio supervisada.

Bloque de preparación de 180 minutos

30 min: leer la nota de HASI y el glosario, y recordar las ideas principales sin consultar el texto. 30 min: laboratorio de fuentes con la lectura asignada o la alternativa bilingüe incluida. 60 min: completar la ficha del módulo y sus preguntas de profundización. 45 min: elaborar o revisar el informe final acumulativo. 15 min: comprobar la comprensión y registrar una pregunta. Son tiempos para trabajar activamente, no estimaciones de duración de lectura. Al agotarse el tiempo, señalar lo que quedó pendiente.

Material básico y profundización

El cuaderno incluye explicaciones, fichas de evidencia ficticia y tareas completas en ambos idiomas. El Informe Internacional sobre la Seguridad de la IA ofrece un resumen oficial en español. Los demás enlaces en inglés son profundización opcional para quienes estudian en español; siempre existe una tarea equivalente de HASI. No sumar toda la carga de lecturas de BlueDot a estas 25 horas.

Participación, finalización y responsabilidades

Cómo se completa el programa

Entregar las cinco fichas, con una vía razonable de recuperación de sesiones. Participar en al menos cuatro sesiones; recuperar una ausencia con un análisis de 300 palabras o tres minutos de exposición oral y una conversación de 15 minutos con la persona facilitadora. Completar las preguntas finales. El informe debe alcanzar al menos 8/12 en la rúbrica publicada, sin ningún criterio en cero. Se permite una revisión dentro de los siete días siguientes. Se evalúa el razonamiento, no el acuerdo con HASI. Estas reglas son la propuesta de HASI para el piloto.

La constancia de HASI puede indicar título del programa, nombre, fechas y 25 horas programadas. No debe sugerir habilitación profesional, créditos académicos, inserción laboral garantizada ni una credencial de BlueDot.

Cómo trabajamos

Cuestionar afirmaciones con respeto; distinguir una experiencia personal de una generalización. Compartir el tiempo de palabra. Se puede pasar el turno, escribir o responder oralmente. Nadie debe revelar situación migratoria, experiencias traumáticas, ingresos familiares ni vivencias de daño. Usar ejemplos ficticios para temas sensibles. Las herramientas de IA son opcionales: declarar su aporte, comprobar las fuentes citadas y explicar el razonamiento propio. Nunca subir datos privados de participantes ni de la institución.

Preparación institucional

HASI prepara materiales, facilitación y retroalimentación. La institución anfitriona designa un contacto, convoca participantes, confirma accesibilidad y condiciones de reunión, y acuerda fechas y procedimientos con HASI.

El piloto propuesto es gratuito para participantes e instituciones. La capacidad de impartición y las fechas se confirman con cada institución. No se requiere una cuenta de IA pagada ni un entorno de programación. Se pueden usar materiales impresos o archivos sin conexión.

Atribución y fuentes

BlueDot permite reutilizar y adaptar sus propios materiales con identidad independiente y atribución clara. Las obras externas enlazadas conservan sus condiciones. Este paquete enlaza esas obras y aporta explicaciones y ejercicios originales de HASI, sin reproducir lecturas completas. Para una cohorte oficial de BlueDot o un acuerdo institucional interno, consultar sus orientaciones y a su equipo.

Condiciones de reutilización de BlueDot • Estrategia de IAG • Informe Internacional, resumen bilingüe