CARRERAS CON IMPACTO

RIESGOS CATASTRÓFICOS GLOBALES

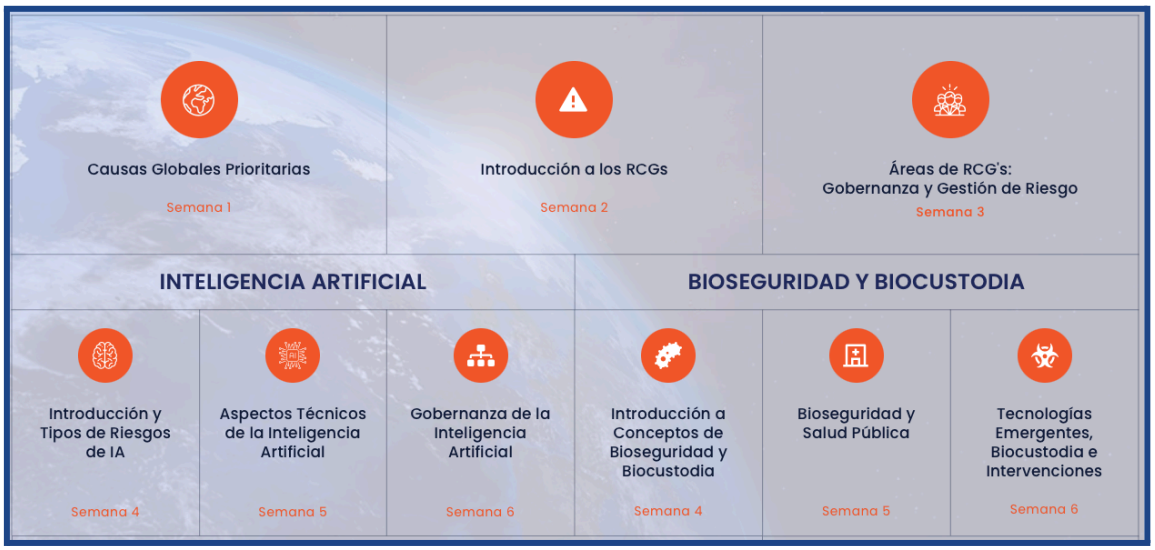
Manual de Contenidos

Programa Virtual - 2024

🌐 OBJETIVO DEL CURSO 🌐

En este programa, aprenderás sobre los riesgos catastróficos globales (RCGs) y explorarás tus opciones profesionales para contribuir en su prevención y mitigación. Las primeras tres semanas corresponden a la introducción a los RCG's, luego el programa se especializa en dos ramas: "Inteligencia Artificial" o "Bioseguridad y Biocustodia".

📖 ESTRUCTURA GENERAL 📖



⭐ INSTRUCCIONES ⭐

Cada semana de este programa aborda un tema diferente relacionado con las herramientas necesarias para tomar decisiones más informadas y acertadas sobre nuestras carreras relacionadas a los RCGs.

Estimamos que las lecturas principales llevarán alrededor de 1 a 2 horas, mientras que las discusiones en grupo tendrán una duración de 2 horas. Esto significa que durante el programa se requerirá un compromiso de tiempo de aproximadamente 4 horas por semana.

Dado que sólo habrá 6 sesiones, faltar a 1 sesión representaría una cantidad sustancial de material no cubierto, por lo que te recomendamos intentes asistir a las sesiones, incluso si no completas todas las lecturas principales. En ese caso, aún tendrás la oportunidad de escuchar a otrxs participantes que compartan su visión personal sobre los argumentos presentados en secciones que quizá no hayas leído.

Para obtener el certificado de finalización y poder postularte a la fase de mentorías de proyectos, es necesario contar con al menos un 80% de asistencia a las sesiones de discusión y haber completado el 80% de los ejercicios en la plataforma del curso.

Esperamos que, como participante, completes las lecturas con anticipación para que logres aprovechar al máximo el grupo y contribuir a una mejor experiencia tanto para tí como para tus compañerxs. Algunos de estos materiales se encuentran en Inglés; puedes traducirlos utilizando los subtítulos automáticos de Youtube o mediante la herramienta de traducción de Google, ya sea ingresando el enlace directamente en el recuadro de traducción o haciendo uso de la extensión para navegadores web.

Aunque se brindan tiempos aproximados de lectura para cada uno de los materiales requeridos, en general preferimos que te tomes tu tiempo y reflexiones sobre la lectura en lugar de apresurarte a través de ella.

🔎 CRITERIOS DE DISCUSIÓN 🔎

1. **Tomar las ideas en serio:** A menudo, las conversaciones sobre ideas funcionan como distracciones recreativas; disfrutamos discutiendo conceptos interesantes y mencionando cosas inteligentes, para luego volver a nuestra rutina habitual. Esto no

tiene nada de malo, pero creemos que en ocasiones deberíamos hacernos preguntas como: "*¿Cómo puedo saber si esta idea es verdadera?*" "Si lo es, *¿qué implicaría en cuanto a las decisiones que debo tomar en mi vida?*" "*¿En qué más podría estar equivocadx?*" Y, ampliando la perspectiva: "*¿Cuáles son mis puntos ciegos?*" "*¿Qué preguntas importantes debería estar explorando y no lo estoy haciendo?*" Tomar las ideas en serio significa buscar que nuestras visiones del mundo sean lo más completas y precisas posible, reconociendo que tener creencias bien fundamentadas nos permiten tomar mejores decisiones sobre lo que realmente importa.

2. **Las discrepancias son interesantes**: Cuando como personas reflexivas, con acceso a la misma información, llegamos a conclusiones diferentes entre sí, deberíamos sentir curiosidad al respecto. A menudo, tendemos a no mostrar esta curiosidad simplemente porque es algo tan común que lo damos por hecho. Pero si, por ejemplo, una comunidad médica discute sobre si el tratamiento A o B es más efectivo para curar una enfermedad, es fundamental investigar a fondo esa discrepancia, ya que la respuesta correcta importa y hay vidas en juego. Incluso si no se logra llegar a un acuerdo, es valioso tratar de comprender el por qué de las diferencias y en qué aspectos concretos exactamente difieren las opiniones.

3. **Opiniones sólidas, pero flexibles**: A menudo, las personas se abstienen de expresar sus opiniones porque piensan cosas como "*No soy un experto*" o "*Es difícil saberlo con certeza*". Sin embargo, durante este programa te invitamos a ser lo suficientemente valiente para compartir tus suposiciones, haciéndolo de manera clara para que la evidencia actual pueda señalar si estás en lo correcto o no. A largo plazo, esperamos que te conviertas en unx pensadorx más sólidx y comprometidx, pues consideramos que esto es más valioso que evitar errores en el corto plazo. Es fácil

caer en desacuerdos vagos y optar por "*acordar en no estar de acuerdo*" sin aprender realmente del otro. Una opinión claramente formulada facilita que otras personas identifiquen con precisión en qué coinciden o difieren contigo.

🧠 NORMAS DEL PROGRAMA 🧠

Te animamos a seguir estas normas a lo largo del programa, con el objetivo de garantizar conversaciones respetuosas y productivas en las que todxs se sientan bienvenidxs a participar.

- **Respeto:** Esperamos que todxs se traten con respeto mutuo. Esto incluye seguir normas básicas de sentido común, pero para mayor claridad, te alentamos a ser amable, evitar interrumpir a los demás, no iniciar conversaciones paralelas a la discusión principal, dirigir la discusión a las ideas y no a las personas, y abstenerte de gestos despectivos como rodar los ojos o reirte de los demás. *¡Sé amable!*
- **Constructividad:** Para fomentar una discusión constructiva, te animamos a no presentar objeciones absolutas, sino a dejar abierta la posibilidad de que exista una respuesta o solución.
- **Respuestas constructivas al desacuerdo:** Durante las discusiones, nos enfocaremos en aclarar nuestra comprensión de la lectura y debatir nuestras perspectivas, por lo que es normal que surjan desacuerdos. Te invitamos a abordar estas situaciones intentando comprender qué piensa la otra persona y por qué: *¿Se debe a una cosmovisión diferente?*, *¿Qué aspecto específico causó el desacuerdo?*, *¿Hay una diferencia en modelos o enfoques?*, *¿Podemos encontrar el núcleo del desacuerdo?*. Recuerda mantener abierta la posibilidad de estar equivocado y reconocer que pueden existir matices en la situación.

Te recordamos también que la institución cuenta con un **Código de Conducta**, que deberás leer detenidamente antes de comenzar el programa. Este código está diseñado para promover un ambiente inclusivo, seguro y respetuoso para todos los participantes.

☎ CONTACTO ☎

Si tienes dudas con el contenido, necesitas asistencia con el contenido o tienes dudas del seguimiento, favor de contactar a tu facilitador/a a través de la plataforma oficial, o bien, al medio de contacto proporcionado.

Si necesitas asistencia, no puedes asistir a una sesión o requieres cambios de horarios, favor de escribir a la siguiente dirección:

miguel.alvarado@carrerasconimpacto.org

Se brinda atención de Martes a Sábado, por lo que si envías una solicitud en fin de semana, deberías obtener respuesta hasta el día Martes.

📚 TEMARIO 📚

🌱 **TRONCO COMÚN** **Sesión 1 y 2**

Objetivo

Conocer los conceptos de "Riesgos Catastróficos Globales" (RCGs) y asociados. En esta parte del curso explorarán los conceptos fundamentales, clasificaciones y priorización de estos riesgos.

SESIÓN 1.
🌐 Causas Globales Prioritarias 🌐

Objetivos de Sesión.

El objetivo de esta sesión es familiarizarse con la "Mentalidad de Scout" y su relevancia a lo largo del curso. Además, busca profundizar en los criterios para priorizar causas globales y su aplicación en la resolución de problemas críticos, así como identificar nuestra ventaja comparativa para determinar la forma más efectiva de contribuir a estas soluciones.

📢 Contenido obligatorio 📢

- **The Scout Mindset** **(12 min - *Inglés con subtítulos en español*)**
  Esta lectura pretende hacer que identifiques 2 tipos de mentalidades: Mentalidad de Scout y Mentalidad de Soldado. Con ello, se busca que se razonen las problemáticas y posibles soluciones a estas con una mentalidad en búsqueda de la verdad, evitando y/o reconociendo sesgos o razonamientos "defensivos".
  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:
  A. ¿Cuál es la diferencia entre la mentalidad de Scout y la de Soldado?
  B. ¿Qué se entiende por "Razonamiento Motivado"?

C. ¿De qué manera podría afectar la falta de mentalidad de Scout en la resolución de una problemática?

- **Criterios de priorización de causas** **(15 min - *Inglés*)**

  El objetivo de esta lectura es que puedas identificar los tres criterios principales para evaluar causas prioritarias y aplicarlos en la comparación de problemáticas globales.

  ***Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas*:**

  A. ¿Cuáles son los 3 criterios de priorización de causas?
  B. Menciona al menos una organización que aplique estos criterios.
  C. ¿Las evaluaciones son cualitativas, cuantitativas o ambas?
  D. ¿Cómo se debe incluir el "Fit Personal?" en la evaluación?

- **Prospecting for gold (Criterios y Ventaja comparativa) - Sección "Working Together" en adelante** **(15 min - *Inglés*)**

  El objetivo de esta lectura es reconocer la importancia de colaborar con la sociedad para construir soluciones a problemáticas globales, aportando nuestros conocimientos y habilidades en aquellas áreas donde tenemos una ventaja comparativa.

  ***Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas*:**

  A. ¿A qué se refiere el concepto de ventaja comparativa?
  B. ¿Qué se entiende por conocimiento agregado?
  C. ¿Cómo puede contribuir el reconocimiento de la ventaja comparativa de cada individuo a la resolución de problemáticas globales complejas?

- **Is Civilization on the brink of collapse?** **(10 min - *Inglés*)**

  Se busca reconocer la historia de la humanidad a través de escenarios de "colapso" vividos por civilizaciones pasadas y analizar las posibles causas de futuros colapsos.

  ***Al finalizar este contenido, deberás ser capaz de responder las siguientes preguntas*:**

A. ¿En qué momentos de la historia se ha considerado que las civilizaciones han "colapsado"?
B. ¿Cuáles son las posibles causas del próximo colapso de nuestra civilización en la actualidad y cómo se pueden clasificar? Proporciona al menos un ejemplo.
C. ¿Es posible recuperarse de un colapso de esta naturaleza?

📝 Ejercicios previos a la sesión 📝

1. Explora la página de contenido previo a la primera sesión en la plataforma y haz el llenado de la **encuesta inicial** del curso.
2. Prepara tu **cámara y micrófono** para la sesión de discusión. Recuerda que es importante que estos estén disponibles para una mejor interacción. Sólo en casos excepcionales se permitirá no hacer uso de su cámara y micrófono.
3. Prepara una **breve presentación personal** de 1-2 minutos para compartir durante la sesión. No es necesario usar una presentación en PowerPoint, solo estar listo para hablar de temas como tu nombre, intereses, actividades, datos sociodemográficos, hobbies, entre otros.
4. Escribe un pequeño párrafo con tus palabras sobre lo que has entendido de la "Mentalidad de Scout". Si tienes algún ejemplo en el que la hayas aplicado o pudieras haberla aplicado, compártelo.
5. Haz una lista de los puntos clave de cada criterio para priorizar causas, ya que se discutirán con tus compañeros durante la sesión.

💬 Contenido Sugerido 💬

- **What are the most important moral problems of our time? (12 min - *Inglés con subtítulos al español*)**
- **All Possible Views About Humanity's Future Are Wild (20 min - Inglés)**
- **Holden Karnofsky: On the most important century (1h 20 min a velocidad 2X - *Inglés*)**

👀 Organizaciones y Páginas Sugeridas 👀

- **Probably Good**
  Es una organización sin fines de lucro que ofrece asesoramiento personalizado en la elección de carrera, utilizando criterios de evaluación de impacto para identificar el mejor rol profesional según el perfil de cada individuo. Brindan orientación en inglés, adaptada al perfil profesional y ubicación geográfica de cada persona. Además, cuentan con una bolsa de trabajo enfocada en carreras de alto impacto, así como contenido y notas relacionadas.
- **80,000 H**
  Es una organización sin fines de lucro que proporciona guías, asesoramiento, contenido, podcasts, investigaciones y una bolsa de trabajo. Su objetivo es ayudar a las personas a tomar decisiones de carrera más informadas, al tiempo que promueven el conocimiento sobre causas prioritarias de alto impacto.
- **Successif**
  Es una organización que ofrece contenido y asesoramiento profesional en inglés, dirigida especialmente a personas en un estado intermedio de su carrera, como quienes cursan un posgrado o tienen cierto nivel de experiencia laboral.

- **High Impact Professionals**
  Esta organización ayuda a los profesionales a unirse a un directorio público de talento para organizaciones que reclutan, facilitando también la búsqueda de candidatos prometedores para organizaciones de "Alto Impacto"
- **Libros Gratuitos**
  A través de este **enlace**, puedes solicitar de manera gratuita varios libros relacionados con temas de este curso tal como criterios de priorización, riesgos catastróficos globales, entre otros.

🗣️ Dinámica de Sesión 🗣️

- ➔ Presentación de la organización, programa y facilitador.
- ➔ Presentación de compañerxs de curso.
- ➔ Discusión sobre concepto de "Mentalidad de Scout" seguida de un ejercicio de calibración.
- ➔ Exposición y análisis de ejemplos de problemáticas globales reales para ilustrar los "Criterios de Priorización de Causas".
- ➔ Explicación de los conceptos de "Ventaja Comparativa".
- ➔ Ejercicio práctico para aplicar los criterios de priorización y evaluar la ventaja comparativa de cada uno.

SESIÓN 2.
🌎 Introducción a Riesgos Catastróficos Globales ☣️

Objetivos de Sesión.

El objetivo de esta sesión es familiarizarse con el concepto de Riesgos Catastróficos Globales (RCGs), identificar las principales causas que pueden desencadenarlos, especialmente las de origen antropogénico, y reconocerlas como prioritarias. También se busca comprender las diferentes capas de estos riesgos, explorar posibles intervenciones para su mitigación y su posible gobernanza.

📢 Contenido obligatorio 📢

- **Global Catastrophic Risks: An Impact-Focused Overview (15 min - *Inglés*)**
  Esta lectura tiene como propósito definir el concepto de Riesgos Catastróficos Globales (RCGs) y explicar cómo estos riesgos se relacionan con cada criterio de priorización de causas para ser reconocidos como causas prioritarias. Igualmente, comprender por qué es relevante la conservación de la humanidad.
  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:
  A. ¿Qué son los riesgos catastróficos globales?
  B. ¿Qué origen pueden tener?
  C. ¿Por qué se considera que estos riesgos son desatendidos, impactantes y pueden ser tratados?
- **Riesgos Catastróficos Globales: Ciencia, Tecnología, Política y Derechos Humanos (20 min - *Español*)**
  El objetivo de esta lectura es comprender los dos posibles orígenes de los Riesgos Catastróficos Globales (RCGs), reconociendo sus posibles consecuencias y las formas de actuar para mitigar estos riesgos. Se pondrá especial énfasis en los

riesgos antropogénicos, destacando su potencial para ser tratados.

Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

A. ¿Cuál puede ser el origen de los RCGs?
B. Proporciona ejemplos de riesgos antropogénicos y naturales.
C. ¿Cuáles son los principales tipos de riesgos asociados con tecnologías emergentes, como la Inteligencia Artificial y la Biotecnología?

- **Capas de RCGs y Clasificación de riesgos (Pg. 271 - 278, 30 min - *Inglés*)**

  Con esta lectura se pretende identificar cuáles son las capas de defensa ante la extinción de la humanidad, las clasificaciones de los riesgos de acuerdo a su origen, mecanismo de escalada y resiliencia.

  Al finalizar esta lectura, deberás ser capaz de clasificar adecuadamente los distintos RCGs de acuerdo con su origen, mecanismo de escalada y resiliencia.

- **Gobernanza y Gestión de Riesgos en Latinoamérica (Pg. 5 - 14, 20 min - *Español*)**

  El objetivo de esta lectura es comprender cómo se gestionan los Riesgos Catastróficos Globales (RCGs) a nivel mundial, con un enfoque particular en América Latina. También se busca reconocer el nivel de riesgo actual en los países de la región y explorar propuestas para su mitigación.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

  A. ¿Cómo se gestiona el riesgo de RCGs a nivel mundial?
  B. ¿Qué institución(es) han estado principalmente involucradas en la gestión de estos riesgos?
  C. ¿Cuáles son algunas propuestas para su mitigación en América Latina? Menciona al menos tres.

📝 Ejercicios previos a la sesión 📝

1. Identifica los **siguientes conceptos**:
    a. Riesgos Catastróficos Globales (RCGs)
    b. Clasificación de RCGs y capas de gestión del riesgo.
    c. Identifica los criterios de priorización que se pueden utilizar dentro del área de los RCGs.

💬 Contenido Sugerido 💬

- **Resumen y notas del precipicio (Capítulos 3 y 4, 25 min - *Inglés*)**
- **Riesgos desconocidos (11 min - *Inglés*)**
- **2024 Doomsday Clock Statement (10 min - *Inglés*)**
  El propósito de explorar esta página es comprender qué es el reloj de Doomsday, qué representa y quiénes son las personas que deciden su hora. Además, se espera reconocer cómo ha evolucionado a lo largo del tiempo desde su creación y los riesgos principales que representa actualmente.
  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:
    A. ¿Qué representa el reloj de Doomsday?
    B. ¿Quién decide la hora en la que nos encontramos?
    C. ¿A cuánto tiempo estamos de la medianoche y cuáles son los principales riesgos que han contribuido a esta situación en 2024?

Áreas de RCGs

Con la lectura de los siguientes materiales se espera que se reconozca el tipo de RCGs, sus posibles causas y consecuencias.

Igualmente, se espera que se puedan reconocer posibles estrategias de mitigación de cada uno.

- **Riesgos naturales**:
  - **Supervolcanes, Asteroides (Capítulo 4, 8 min - *Inglés*)**
  - **Cambio Climático (5 min - *Inglés*)**
- **Inteligencia Artificial (8 min - *Inglés*)**
- **Riesgo Biológico (9 min - *Inglés con subtítulos al Español*)**
- **Riesgos Nucleares (Capítulo 4, 8 min - *Inglés*)**
- **ERALS (min 00:00:30 a 00:09:40, 10 min - *Inglés*)**

👀 Organizaciones y Páginas Sugeridas 👀

- **Observatorio de Riesgos Catastróficos Globales (ORCGs)**
  El observatorio es una organización de diplomacia científica cuyo objetivo es impulsar la gestión de riesgos catastróficos globales en América Latina y España. Para lograr nuestra misión, conectamos a tomadores de decisiones con expertos y elaboramos publicaciones basadas en evidencia.
- **Blue Dot Impact**
  Cursos enfocados en dar herramientas y conocimientos principalmente en riesgos catastróficos globales de origen antropogénico: Bioseguridad y Biocustodia, Alineamiento y gobernanza de Inteligencia Artificial.
- **Centro de Estudios de Riesgos Existenciales (CSER, Cambridge)**
  Centro de estudios perteneciente a la Universidad de Cambridge dedicada al análisis de riesgos existenciales y desarrollo de proyectos e iniciativas en el área.

🗣️ Dinámica de Sesión 🗣️

- ➔ Descripción de lo que son los Riesgos Catastróficos Globales (RCGs), sus posibles orígenes y los escenarios que podrían generarlos.
- ➔ Análisis de las distintas capas de los RCGs, identificando cómo se escalan y qué mecanismos existen para su mitigación.
- ➔ Ejercicio de Clasificación de Riesgos: Actividad de clasificación de los riesgos según su impacto y probabilidad, y discusión de propuestas de medidas de mitigación para cada uno.
- ➔ Análisis de RCGs en contexto global y regional para su gobernanza.

Ejercicios a entregar en Plataforma

- ➔ Ejercicios realizados en clase de acuerdo a los Ejercicios prácticos de sesión
  - ◆ ***Nota:*** Recuerda hacer una copia de estos templetes.

TRONCO DE ESPECIALIDAD I

🦠 **BIOSEGURIDAD Y BIOCUSTODIA** **Sesión 3 a 6**

Objetivo

El objetivo de esta especialidad es identificar los Riesgos Catastróficos Globales (RCGs) de origen biológico, reconociendo tanto los riesgos naturales como los antropogénicos, junto con sus posibles consecuencias y estrategias de mitigación. Este enfoque permitirá comprender los riesgos y las formas de enfrentarlos, especialmente ante los nuevos desafíos que surgen con la innovación en técnicas y disciplinas biotecnológicas, como la biología sintética, la síntesis de ADN, la secuenciación de nueva generación, y su intersección con la Inteligencia Artificial, entre otras.

🦠 BIO - SESIÓN 3 🦠

👩‍🔬 Introducción al riesgo biológico y conceptos clave 👩‍🔬

Objetivos de Sesión.

Esta sesión tiene como objetivo conocer los conceptos claves para el entendimiento de RCGs de origen biológico, su costo efectividad, historia y estrategias que se han propuesto para su mitigación.

📢 Contenido obligatorio 📢

- **Conceptos Clave (Pg. 2-5, 5 min - *Inglés*)**
  Este contenido busca aclarar los conceptos más relevantes sobre Bioseguridad, Biocustodia, Bioterrorismo, Biocrimen, Riesgo Catastrófico Global de origen biológico, entre otros relevantes. Estos conceptos servirán como base para la comprensión de las siguientes lecturas y análisis de casos.

Al finalizar esta lectura, deberás ser capaz de identificar correctamente el término con sus conceptos.

- **Germ Warfare (10 min - *Inglés*)**

  El objetivo de esta lectura es reconocer los sucesos más importantes en la historia relacionados a actos de bioterrorismo, programas de armas biológicas e historia de la convención de armas biológicas.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

  A. ¿Desde cuándo se han utilizado armas biológicas?
  B. ¿Cómo han sofisticado este tipo de armas?
  C. ¿Cuál es el origen y objetivo de la convención de armas biológicas?

- **¿Cómo crear un arma biológica por accidente? (25 min a velocidad 2X - *Español*)**

  Este contenido tiene como principal objetivo comprender los beneficios y riesgos que posee el avance de la biología sintética.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

  A. Describir el concepto de "Uso Dual".
  B. Comprender las ventajas en innovación para solución de problemáticas que trae la biología sintética.
  C. Relacionar los beneficios de la biología sintética con posibles riesgos y estrategias de mitigación.

- **Cost-effectiveness of biosecurity (10 min - *Inglés*)**

  Con esta lectura se busca comprender el término de "Costo-efectividad" aplicado a la gestión de riesgos biológicos y el nivel de impacto que un riesgo de este tipo podría tener en vidas humanas y nivel económico. Igualmente, explora antecedentes que funcionan como un punto de partida para comprender estas estimaciones.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

A. ¿El trabajo en bioseguridad y biocustodia puede considerarse "Costo-Efectivo?¿Por qué?
B. ¿Cuáles son los distintos alcances en tiempo y personas afectadas que diferencian entre un: incidente, evento, desastre o riesgo existencial?
C. ¿Cuáles son los costos por vida estimados de acuerdo a cada modelo?

- **80,000 Hours problem profiles: Global catastrophic biological risks (20 min - *Inglés*)**
  Explorarás los enfoques y estrategias necesarias para prevenir pandemias catastróficas, abordando los riesgos biológicos a nivel global. Aprenderás a identificar las causas más impactantes para reducir la probabilidad de pandemias devastadoras y comprender los criterios para priorizar proyectos con el mayor potencial de mitigación.
  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:
  A. ¿Qué factores hacen que una pandemia sea catastrófica?
  B. ¿Cuáles son las estrategias más efectivas para prevenir pandemias catastróficas?

Ejercicio previo a la sesión 📝

1. Identifica los conceptos clave de la sesión: Bioseguridad, Biocustodia, Biodefensa, Bioterrorismo y Biocrimen.
2. Investiga uno de los casos que más te haya interesado de casos de bioterrorismo o armas biológicas para discutirlo en la sesión.

Contenido Sugerido 💬

- Does Biotechnology Pose New Catastrophic Risks? (20 min)
- A Short History of Biological Warfare: From Pre-History to the 21st Century (60 min)

- What Are the Odds H5N1 Is Worse Than COVID-19?(15 min)

🗣️ Dinámica de Sesión

- ➔ Quiz de conceptos clave relacionados al tema.
- ➔ Visualización de Cronología e Historia de armas biológicas y ataques bioterroristas.
- ➔ Discusión de casos investigados y relación con actores actuales.
- ➔ Alcances actuales e innovación de la biotecnología.
- ➔ Debate de ventajas y riesgos de las innovaciones en biotecnología.
- ➔ Mención de riesgos emergentes en la biotecnología como fuente de riesgo para futuras pandemias.
- ➔ Mención de estrategias para prevención de pandemias.

📑 **Contenido de Sesión:** Ejercicios prácticos de sesión

Nota: Recuerda hacer una copia de estos templetes.

🦠 BIO - SESIÓN 4.

🧬 Biotecnología y su intersección con tecnologías emergentes.

Objetivos de Sesión.

En esta sesión, el objetivo es profundizar en los Riesgos Catastróficos Globales (RCGs) de origen biológico, con un enfoque particular en aquellos relacionados con fuentes antropogénicas y tecnologías emergentes. Se abordarán temas clave como la biocustodia, la biodefensa, y cómo estas áreas se entrelazan con tecnologías emergentes, como la Inteligencia Artificial, ampliando la visión de los riesgos y oportunidades que presentan en un contexto global.

Contenido Obligatorio 📢

- **Global Catastrophic Risks Chapter 20 - Biotechnology and Biosecurity | Nick Bostrom** (50 mins.)
  Este capítulo explora la biotecnología, sus aplicaciones y los riesgos asociados. Se centra en la historia del bioterrorismo, el desarrollo de armas biológicas y las formas de gestionar los riesgos actuales para maximizar los beneficios que ofrece esta tecnología.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

  A. ¿Cuáles son los beneficios y desafíos de la biotecnología?
  B. Definir: Bioseguridad, biocustodia, biodefensa, bioterrorismo e investigación de uso dual.
  C. Reconocer escenarios históricos que han representado riesgos biológicos y posibles escenarios futuros.

- **Report launch: examining risks at the intersection of AI and bio** (15 min).
  Este informe explora cómo la Inteligencia Artificial (IA) puede amplificar los riesgos biotecnológicos, incluyendo el desarrollo de armas biológicas y herramientas biológicas avanzadas. Además, analiza los puntos de intervención y estrategias para mitigar estos riesgos sin frenar la innovación.
  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

A. ¿Cómo puede la IA acelerar los riesgos en biotecnología?
B. ¿Cuáles son las estrategias clave para mitigar estos riesgos?
C. ¿Qué mecanismos de gobernanza pueden implementarse para proteger la biocustodia en un mundo con IA?

- **Preventing the Misuse of DNA Synthesis** (15min)
  Este informe explora los riesgos asociados a la síntesis de ADN, incluyendo su potencial uso indebido en la creación de patógenos peligrosos. Examina los enfoques actuales de screening, desafíos y propone recomendaciones para reducir estos riesgos.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

A. ¿Cómo puede la síntesis de ADN ser utilizada de forma indebida?
B. ¿Qué medidas de mitigación existen para prevenir este uso indebido?
C. ¿Cuáles son las políticas recomendadas para la reducción de este riesgo?

- **Information Hazards in Biotechnology | Greg Lewis** (25 mins.)
  Este artículo explora los peligros de la difusión de información biotecnológica que podría ser mal utilizada, como en casos de patógenos potencialmente pandémicos o toxinas. Discute la dificultad de equilibrar el avance científico con la seguridad y los desafíos del uso dual de la investigación.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:
  A. ¿Qué es el peligro que puede traer la información en biotecnología?
  B. ¿Cómo se puede gestionar el mal uso de la información biológica?
  C. ¿Cuáles son ejemplos clave que destacan estos peligros?

Ejercicio previo a la sesión 📝

1. Piensa en las motivaciones de un grupo (estatal o no estatal) para un ataque a través de un agente biológico.
2. Investiga los alcances de "AlphaFold".
3. Piensa en estrategias para mitigar los riesgos de difundir información que puede tener uso dual.

Contenido Sugerido 💬

- The New Bioweapons: How Synthetic Biology Could Destabilize the World (12 min)
- The Biological Weapons Convention: An Introduction. (40min)
- Dual Use Research of Concern in the Life Sciences: Current Issues and Controversies (10 min)
- How AI Can Help Prevent Biosecurity Disasters (15 min)

🗣️ Dinámica de Sesión

- ➔ Presentación de los alcances de la biotecnología y futuras tendencias.
- ➔ Discusión de tendencias de Inteligencia Artificial (IA) como herramienta para el desarrollo de armas biológicas. Enfoque en caso de AlphaFold.
- ➔ Debate de investigación de uso dual y de motivaciones de grupos para uso de armas biológicas.
- ➔ Propuesta y estrategias para la mitigación de riesgos biológicos de origen antropogénico (Énfasis en convención de armas biológicas).

📑 **Contenido de Sesión:** Ejercicios prácticos de sesión

Nota: Recuerda hacer una copia de estos templetes.

🦠 BIO - SESIÓN 5.

🏥 Bioseguridad y Salud Pública

Objetivos de Sesión.

En esta sesión, se profundizará en temas clave relacionados con la bioseguridad y la respuesta ante pandemias. Se mencionan áreas fundamentales para mejorar la bioseguridad, monitoreo, diagnóstico y tratamientos, así como los desafíos en escalabilidad y robustez. Además, se analizarán ejemplos exitosos de contención de pandemias y se evaluan la efectividad de las intervenciones no farmacéuticas para reducir la transmisión del COVID-19. Igualmente, se presentan ideas de proyectos para mejorar la resiliencia global ante riesgos biológicos.

Contenido Obligatorio 📢

- **A Framework for Technical Progress on Biosecurity ((10 mins.)**
  Este artículo ofrece un marco para mejorar la bioseguridad y la preparación ante pandemias mediante seis áreas clave: monitoreo,

forense, barreras, diagnóstico, profilácticos y terapéuticos. En cada área, se buscan cinco metas: aumentar la velocidad, reducir costos, mejorar la generalización, robustez y escalabilidad.

Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

A. ¿Cuáles son las áreas clave para mejorar la bioseguridad y la preparación ante pandemias?
B. ¿Cómo puede la innovación biotecnológica acelerar la prevención y mitigación de riesgos?
C. ¿Qué desafíos aborda el marco en términos de escalabilidad y robustez?

- **Emerging COVID-19 success story: Vietnam's commitment to containment (15 mins.)**

  Este artículo destaca las medidas de Vietnam para detectar y contener las infecciones desde el inicio de la pandemia. Estrategias clave incluyeron cierres fronterizos inmediatos, un rastreo de contactos extensivo, cuarentenas centralizadas y comunicación gubernamental clara y coherente.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

A. ¿Qué estrategias implementó Vietnam para contener la COVID-19?
B. ¿Cómo contribuyó la comunicación del gobierno al éxito de la respuesta?
C. ¿Qué factores diferenciaron la respuesta de Vietnam de otros países?

- **COVID-19: examining the effectiveness of non-pharmaceutical interventions (Pg. 3-5, 15 min)**

  Este informe examina la eficacia de varias intervenciones no farmacéuticas (INF) en la reducción de la transmisión de COVID-19, incluyendo el uso de mascarillas, distanciamiento social y restricciones de movilidad. Se destaca la importancia de combinar diferentes medidas para lograr un mayor impacto en la contención del virus.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

A. ¿Qué INF fueron más efectivas para reducir la transmisión de COVID-19?

B. ¿Cómo influyó la combinación de INF en los resultados?
C. ¿Qué factores afectaron la efectividad de estas intervenciones?

- **Project Ideas in Biosecurity for EAs (10 min)**
  Este artículo proporciona una lista de ideas de proyectos en bioseguridad para altruistas efectivos, destacando oportunidades para mejorar la resiliencia global frente a riesgos biológicos.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

A. ¿Cuáles son algunos proyectos clave para mejorar la bioseguridad?
B. ¿Cómo pueden los altruistas efectivos contribuir a la mitigación de riesgos biológicos?

Ejercicio previo a la sesión 📝

1. Revisión de al menos 3 estrategias tomadas en su país para contención de la pandemia por COVID-19.
2. Revisa al menos 2 ideas de proyectos de la lista de "Project Ideas in Biosecurity for EAs" e investiga un poco más acerca de su potencial para discutir en la sesión.

Contenido Sugerido 💬

- **Interventions to Reduce Risk for Pathogen Spillover and Early Disease Spread to Prevent Outbreaks, Epidemics, and Pandemics (20 min)**

🗣️ **Dinámica de Sesión**

➔ Discusión del progreso en biocustodia y bioseguridad.
➔ Análisis retrospectivo de casos de países y su manejo de COVID-19.
➔ Presentación y análisis de estrategias para futura prevención y preparación para pandemias.
➔ Preparación de Pitch: Ideación de nueva idea para mitigación de riesgos.

➔ Explicación de ejercicios a hacer previo a la sesión 6.

📑 **Contenido de Sesión:** Ejercicios prácticos de sesión

Nota: Recuerda hacer una copia de estos templetes.

🦠 BIO – SESIÓN 6.

📲Discusión de Caminos Profesionales, ¿Qué sigue?

Objetivos de Sesión.

En esta sesión, contarás con las herramientas y referentes necesarios para identificar y proponer al menos tres caminos profesionales o proyectos enfocados en la mitigación de riesgos catastróficos globales de origen biológico, aprovechando tu carrera profesional. Utilizarás el Modelo de Factores Ponderados (WFM) y otros enfoques para evaluar y priorizar tus opciones, enmarcados en criterios de prioridades globales e impacto. Al finalizar, estarás equipado para tomar decisiones informadas y alineadas con el impacto que deseas generar en bioseguridad y/o áreas relacionadas.

Contenido Obligatorio 📢

- **Aborda tu decisión de carrera estratégicamente (10 min)**
  La lectura explora el cómo tomar decisiones en tu carrera de forma estratégica para maximizar tu impacto en el mundo. Su contenido te introducirá al método "SELF" en la toma de decisiones profesionales analizando cada factor por separado.
  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:
  - ¿Qué es el método SELF y por qué es útil aplicarlo para enfocar nuestra carrera en causas prioritarias?
  - ¿A que se refiere el factor "Leverage" del método y cómo aplicarlo en la evaluación de puestos de trabajo?

- ¿Qué es el factor "Fit" y cómo aplicarlo para identificar tus intereses y fortalezas?

- **Modelo de Factores Ponderados (15min)**

  El modelo de factores ponderados (WFM) es una herramienta utilizada para evaluar y comparar opciones basadas en varios criterios ponderados según su importancia. El proceso incluye generar una lista de criterios, asignarles pesos según su relevancia, y puntuar las alternativas bajo cada criterio. Esta herramienta funge como una herramienta valiosa para la toma de decisiones.

  Al finalizar esta lectura, deberás ser capaz de reconocer lo que es el modelo y aplicarlo efectivamente en el contexto de decisiones de vida profesional.

- **¿Cómo elegir una causa a la cual apoyar? (11 min)**

  El video muestra una perspectiva general de cómo seleccionar una causa para generar el mayor impacto posible. Su contenido te ayudará a comprender a qué nos referimos con "impacto" y qué consideraciones debes tomar al comparar distintas causas.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

  - ¿Cuáles son los criterios a considerar al momento de comparar distintas causas?
  - Además de los criterios de priorización, ¿qué otros parámetros debemos tomar en cuenta para impactar en una causa?
  - ¿Cómo influyen parámetros como el dinero, el tiempo y el capital social en la priorización de causas?

- **Ideas de proyectos en bioseguridad, biocustodia y preparación ante futuras pandemias (10min)**

  Presenta diversas ideas de proyectos en bioseguridad, desde iniciativas para mejorar la preparación ante pandemias hasta el desarrollo de nuevas tecnologías de detección y control de patógenos.

Al finalizar esta lectura, deberás ser capaz de identificar y proponer ideas de proyectos en bioseguridad, entendiendo cómo estas iniciativas pueden mitigar los riesgos biológicos globales.

Ejercicio previo a la sesión 📝

1. Anota los **puntos más relevantes de las lecturas**, con el fin de discutirlos con tus compañeros durante la sesión.
    a. Enfócate en reflexionar cómo puedes aplicar los conceptos presentados en tú desarrollo profesional individual

2. Previo a la sesión, genera una **comparativa** de tus **posibles caminos profesionales** impactantes usando este templete. Tu facilitador se encargará de explicar los pasos a seguir al final de la sesión 5.
    a. Este ejercicio aplica los principios del modelo de factor ponderado. Te recomendamos revisar este recurso para comprender más sobre el mismo.
    b. Te recomendamos que revises los perfiles profesionales de 80,000 horas y Probably Good, que han sido curados por expertos en el área para determinar los perfiles profesionales con un mayor impacto en la actualidad.

3. Desarrolla un **plan de acción detallado** a mediano plazo con base en en este templete. Tu facilitador se encargará de explicar los pasos a seguir al final de la sesión 5.

Contenido Sugerido 💬

- **Ajuste personal (10 min)**
  Esta lectura complementa el contenido visto en la lectura anterior, enfocándose específicamente en cómo evaluar el factor "Fit" en tu

carrera. Este contenido te permitirá identificar lo que haces bien y en lo que podrías mejorar al momento de elegir roles impactantes.

Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

- ¿Cómo puedo tomar en cuenta mi motivación sin caer en el cliché de "encontrar mi pasión"?
- ¿Por qué debo considerar mis límites y necesidades al momento de elegir un camino profesional?
- ¿Cómo puedo identificar fortalezas y habilidades personales que pueden ser útiles en un puesto laboral?

- **Pregúntate qué pasaría si no hicieras este trabajo (12 min)**

  Esta lectura nos adentra a la comprensión del pensamiento contrafactual a partir de la reflexión de qué sucedería si no realizáramos determinada acción. Esperamos que su contenido te permita comparar roles e identificar aquellos donde podrías hacer una mayor diferencia.

👀Organizaciones y Páginas sugeridas

- **Base de Datos: Organizaciones.**

  Organizaciones clave para la bioseguridad, biocustodia, prevención y preparación para pandemias o áreas relacionadas.

- **Effective Thesis**

  Ideas de proyectos de investigación para desarrollo de tesis en RCGs o área considerada de "Alto Impacto" a nivel global.

🗣️ Dinámica de Sesión

- En esta sesión primero se presentará y explicará lo que son las herramientas de toma de decisión aplicadas a la vida profesional.
- Posteriormente se dará espacio a cada uno de presentar los posibles caminos profesionales a escoger, sus resultados y el plan de acción del más adecuado.

> ***Nota***: Recuerda relacionar estos caminos a ponderar con las propuestas de proyectos que podrías presentar para el programa de mentoría.

📑 **Contenido de Sesión:** Ejercicios prácticos de sesión

Nota: Recuerda hacer una copia de estos templetes y **subirlo a la plataforma**.

TRONCOS DE ESPECIALIDAD: II

🤖INTELIGENCIA ARTIFICIAL — Sesión 3 a 6

Objetivo

El objetivo de este curso es identificar los Riesgos Catastróficos Globales (RCGs) asociados con la Inteligencia Artificial (IA). Se pretende que los participantes reconozcan el impacto de esta tecnología emergente en la innovación y las posibles amenazas que conlleva. Además, se busca que los estudiantes comprendan los desafíos y oportunidades relacionados con la gobernanza, la ética y el alineamiento de la IA. Al finalizar el curso, los participantes habrán desarrollado las habilidades y conocimientos necesarios para proponer proyectos orientados a la mitigación de estos riesgos.

💻 IA - SESIÓN 3.

🖥️Introducción a Riesgos de Inteligencia Artificial

Objetivos de Sesión.

En esta sesión, obtendrás una comprensión profunda sobre el impacto de la inteligencia artificial (IA) como una tecnología de propósito general y su papel en el presente y futuro disruptivo. Exploraremos los riesgos asociados a la IA, desde los usos maliciosos hasta los desafíos en su gobernanza, y analizaremos las metodologías para prever su evolución. Al finalizar, estarás equipado para identificar los principales riesgos catastróficos relacionados con la IA y las oportunidades para mitigarlos, así como para comprender cómo estos avances podrían transformar

diversos sectores, desde la seguridad hasta la economía global.

Contenido Obligatorio 📢

- **AI and impact of General Purpose Technologies (20min)**
  El objetivo de esta lectura es comprender qué son las tecnologías de propósito general, identificar ejemplos históricos de tecnologías revolucionarias, y analizar las principales revoluciones tecnológicas que han transformado a la humanidad y las que se anticipan en el futuro. Además, se busca entender la inteligencia artificial como una tecnología general y su potencial impacto.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

A. ¿Qué son las tecnologías de propósito general y cuáles son sus principales características?
B. ¿Qué ejemplos históricos se consideran "tecnologías revolucionarias" y por qué?
C. ¿Cuáles han sido las principales revoluciones tecnológicas en la historia de la humanidad y cuáles se proyectan para el futuro?
D. ¿Qué significa que la inteligencia artificial sea considerada una "tecnología general"?

- **Inteligencia Artificial y sus posibles impactos en la sociedad. (Min. 00:01:31 a 00:52:00) (30min a velocidad 2X).**
  El principal objetivo de este contenido es analizar el papel actual de la IA y su futuro disruptivo, explorando tanto sus capacidades presentes como las esperadas a corto y largo plazo. Además, se clasificarán los riesgos asociados a la IA y se abordarán los desafíos relacionados con su gobernanza, proporcionando una visión integral de cómo manejar dichos riesgos.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

  A. ¿Cómo se pueden clasificar los riesgos asociados a la inteligencia artificial?
  B. ¿Qué son los riesgos adversarios y cuál es un ejemplo de ellos?
  C. ¿Qué son los riesgos estructurales y cuál es un ejemplo de ellos?

D. ¿Qué actores podrían hacer un uso indebido de la inteligencia artificial?
E. ¿Cuáles son los principales desafíos en la gobernanza de la IA y qué soluciones se proponen?

- **An Overview of Catastrophic AI Risks. (Pg. 4 - 19, 25 min)**
  El objetivo de esta lectura es comprender cómo la inteligencia artificial (IA) presenta riesgos de uso malicioso. Se exploran los potenciales usos indebidos de la IA en diversos contextos, los actores que podrían aprovecharse de estas tecnologías para fines adversos, y los desafíos asociados a la regulación y la mitigación de estos riesgos de forma introductoria.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:
  A. ¿Cuáles son los principales riesgos relacionados con los usos maliciosos de la inteligencia artificial?
  B. ¿Qué tipos de actores podrían utilizar la inteligencia artificial con fines maliciosos?
  C. ¿Cómo podrían usarse las tecnologías de IA para desarrollar armas autónomas u otros sistemas peligrosos?
  D. ¿Cuáles son los ejemplos de ataques cibernéticos facilitados por IA y cómo podrían evolucionar?
  E. ¿Qué mecanismos de gobernanza se proponen para mitigar los riesgos de uso indebido de la IA?

- **AI Timelines: Where the Arguments, and the "Experts," Stand (15min)**
  Esta lectura tiene como objetivo presentar los distintos métodos de estimación de predicciones sobre el desarrollo y los impactos de la inteligencia artificial. Se analizan las metodologías empleadas en la previsión de IA, las principales incertidumbres que persisten, los posibles escenarios futuros, y las implicaciones de estos avances para la sociedad y la gobernanza.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

A. ¿Qué métodos se utilizan para realizar pronósticos sobre el avance de la IA?
B. ¿Cómo podrían los avances en IA afectar el mundo laboral, la seguridad y la economía en las próximas décadas?
C. ¿Para qué año se piensa que la IA sea considerada una "Tecnología Revolucionaria"? ¿Qué implicaciones tendría esto?
D. ¿Para qué año se espera tener una IA general y qué riesgos e incertidumbres pueden surgir de ello?

Ejercicio previo a la sesión 📝

1. Realiza las lecturas correspondientes e identifica los conceptos clave de acuerdo a las preguntas que debes ser capaz de responder.

2. Escoge uno de los posibles riesgos y haz una investigación de máximo 30 min al respecto con el objetivo de hacer una lluvia de ideas de posibles formas de prevención del riesgo. Esto se compartirá con tus compañeros en la sesión.

Contenido Sugerido 💬

- **An Overview of Catastrophic AI Risks. (Resto de contenido)**
- **AI Explained (Youtube Channel)**
- **What risks does AI pose? (25 min)**

👀Organizaciones y Páginas sugeridas

- **AI SAFETY: Landscape Map.**
  Se encuentran las organizaciones, compañías, instituciones, proyectos, fuentes de financiamiento, etc. relevantes asociadas a los avances y seguridad de la Inteligencia Artificial.

- **Effective Thesis: IA**

Ideas de proyectos de investigación para desarrollo de tesis en IA, RCGs o área considerada de "Alto Impacto" a nivel global.

🗣️ Dinámica de Sesión

- ➔ Quiz de conceptos clave relacionados al tema.
- ➔ Presentación y aclaración de conceptos relevantes.
- ➔ Debate de ventajas y riesgos de Inteligencia Artificial (IA).
- ➔ Ejercicio de línea de tiempo de tecnologías revolucionarias.
- ➔ Ejercicio: Clasificación de riesgos por tipo de riesgo.
- ➔ Presentación con compañeros de posibles usos maliciosos de la IA y estrategias para su mitigación.

📑 **Contenido de Sesión:** Ejercicios prácticos de sesión

Nota: Recuerda hacer una copia de estos templetes.

💻 IA - SESIÓN 4.

🔒 Seguridad y problema de alineamiento de la IA

Objetivos de Sesión.

En esta sesión, comprenderás de forma integral los principales desafíos y enfoques relacionados con el alineamiento y la seguridad de la inteligencia artificial (IA) avanzada. Exploramos los riesgos asociados al desarrollo de IA general, los problemas inherentes al alineamiento de la IA con los valores humanos, y las metodologías clave para ello. Además, comprenderás cómo se utilizan herramientas para garantizar que los sistemas de IA operen de manera segura y transparente.

Contenido Obligatorio 📢

- **Intro to AI Safety (20 min)**
  Este contenido lo que busca es identificar los riesgos de acuerdo a los que se creen que pueden suceder en el corto plazo o en el largo plazo.

Igualmente habla sobre las tendencias de la Inteligencia Artificial y lo que sucedería al alcanzar una IA General.

Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

A. ¿Cuál es la tendencia de avance de la IA?
B. ¿Cuál es el problema del alineamiento? Da un ejemplo.
C. ¿Cuál es la diferencia entre la optimización y el conseguir un resultado al utilizar la IA?
D. ¿Por qué se considera que la IA General es peligrosa por *default*?
E. ¿Cómo el problema de alineamiento de la IA podría afectar situaciones más complejas? Da un ejemplo.

- **What is AI alignment & Safety? (15 min)**

  Esta lectura tiene como objetivo explicar qué significa el alineamiento de la IA, por qué es crucial para la seguridad de las IA avanzadas y los desafíos asociados con lograr que los sistemas de IA realicen acciones acordes con los valores humanos. También se exploran ejemplos de fallas de alineamiento y las posibles consecuencias a largo plazo si no se aborda adecuadamente este problema.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

  A. ¿Qué es el alineamiento de la IA y por qué es importante?
  B. ¿Qué ejemplos existen de problemas de alineamiento en sistemas de IA actuales?
  C. ¿Cómo puede un sistema de IA fallar en alinearse con los valores humanos?
  D. ¿Qué riesgos podría traer el problema del alineamiento en el futuro a medida que los sistemas de IA se vuelvan más poderosos?

- **The Need For Work On Technical AI Alignment (20 min)**

  Esta lectura tiene como objetivo ofrecer una introducción al problema del alineamiento de la IA, explicando los conceptos clave, las razones por las cuales es difícil alinear sistemas avanzados de IA con los intereses humanos y los enfoques actuales que buscan resolver este desafío.

Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

E. ¿Por qué es importante aprender sobre el problema del alineamiento de la IA?
F. ¿Cuáles son las principales dificultades para alinear sistemas avanzados de IA con los intereses humanos?
G. ¿Qué enfoques se están considerando actualmente para resolver el problema del alineamiento?(Ej. Sistemas de recompensas, limitación de capacidades, etc.)
H. ¿Qué consecuencias podría tener no resolver el problema del alineamiento a largo plazo?

- **International Scientific Report on the Safety of Advanced AI**
  Pg. 34 - 41; 78 - 83 (25 min)
  Esta lectura abarca los enfoques y metodologías utilizados para evaluar el desempeño y la alineación de los sistemas de IA avanzada. Se analizan los *benchmarks* como herramientas clave para medir la capacidad de la IA, junto con otros métodos como auditorías y pruebas adversariales. También se discuten los retos asociados con la transparencia de los modelos, la interpretación de resultados y las limitaciones que presentan los actuales sistemas de IA en cuanto a su alineamiento con los intereses humanos.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

  F. ¿Qué métodos se utilizan para evaluar el alineamiento de la IA?
  G. ¿Qué son los *benchmarks* y cuál es su rol en la evaluación de los sistemas de IA?
  H. ¿Cuáles son los principales desafíos para asegurar la transparencia y comprensión de los sistemas de IA?
  I. ¿Cómo las auditorías y las pruebas adversariales contribuyen a la evaluación de la seguridad de la IA?

Ejercicio previo a la sesión 📝

Revisa el siguiente caso "Emergent tool use from multi-agent

interaction" de Open AI y contesta lo siguiente:

- ¿Cuál crees que fue el resultado pretendido a nivel humano para el sistema?
- ¿Cuál era la meta especificada para el modelo?
- ¿Qué hizo realmente el modelo?
- ¿Considerarías que el modelo tiene un problema de alineamiento?¿Por qué?
- ¿Con qué intenciones o "valores" deberíamos alinear los sistemas de inteligencia artificial? ¿Cómo manejarías a diferentes partes interesadas que desean alinear los sistemas de IA con intenciones o valores diferentes?

Contenido Sugerido 💬

- **But what is a Neural Network? - 20 min**
- **AI Safety.World**

👀**Organizaciones y Páginas sugeridas**

- **AI Safety.World**
- **Blue Dot Impact: AI Safety Fundamentals.**
  Cursos diseñados para el desarrollo de tecnologías que intervengan en desarrollar IA segura.

🗣️ **Dinámica de Sesión**

➔ Explicación de los conceptos clave.
➔ Análisis de casos de problema de alineamiento: Debate de ejercicio realizado previo a la sesión.
➔ Ejercicio de calibración / Gráfica.
➔ Revisión de estrategias para mitigación de riesgos y evaluación de alineamiento.
➔ Ejercicio de propuestas de benchmarks para análisis de capacidades de IA.

📑 **Contenido de Sesión:** Ejercicios prácticos de sesión

Nota: Recuerda hacer una copia de estos templetes.

💻 IA – SESIÓN 5.

🏛️Gobernanza de la IA

Objetivos de Sesión.

En esta sesión, explorarás la gobernanza de los riesgos asociados con la inteligencia artificial (IA), basándote en lecciones de estudios históricos, marcos regulatorios emergentes y enfoques globales. Estudiarás los desafíos regulatorios que surgen de las capacidades inesperadas de los modelos de IA avanzada, así como las estrategias propuestas para establecer estándares de seguridad.

Contenido Obligatorio 📢

- **Historical case studies of technology governance and international agreements (compilation – various authors) (30min)**

  Los estudios de casos históricos sobre la gobernanza de tecnologías analizan lecciones de acuerdos internacionales y la regulación de tecnologías poderosas, como las armas nucleares y la electricidad. Este contenido busca identificar paralelismos con la gobernanza de la IA, destacando los riesgos, desafíos regulatorios y estrategias de coordinación global exitosas.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

  A. ¿Cómo informan los acuerdos de control de armas sobre la gobernanza de la IA?
  B. ¿Qué desafíos surgen al regular tecnologías de uso general?
  C. ¿Qué lecciones del tratado de no proliferación nuclear podrían aplicarse a la IA?

- **Why governing AI is our opportunity to shape the long term future? (20min)**
  Este contenido apoya a comprender los desafíos del futuro a largo plazo que posee la inteligencia artificial y la importancia de su gobernanza.

- **Frontier AI Regulation: Managing Emerging Risks to Public Safety. (Pg. 6 -21, 20 min)**
  Este contenido busca explorar los retos regulatorios que presentan los modelos de IA, principalmente aquellos capaces de generar riesgos severos para la seguridad pública. Se discuten los problemas de capacidades inesperadas, la seguridad en el despliegue y la proliferación de estos modelos.
  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:
  D. ¿Cuáles son los principales problemas regulatorios que plantea el desarrollo de IA?
  E. ¿Qué mecanismos se proponen para desarrollar estándares de seguridad en la IA?
  F. ¿Qué rol deberían tener las autoridades gubernamentales y reguladoras en la gobernanza de estos modelos?

- **Envisioning a Global Regime Complex to Govern Artificial Intelligence (Pág. 1 - 12, Introduction - Setting Standards (15 min)).**
  Esta lectura examina los desafíos de establecer un marco de gobernanza global para la IA. Se propone un enfoque de régimen complejo que incluye varias instituciones para abordar temas como la comprensión científica, la armonización de estándares, el acceso equitativo a los beneficios de la IA y la promoción de la seguridad colectiva frente a los riesgos emergentes de la IA.

  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:
  A. ¿Cuáles son las funciones clave necesarias para la gobernanza global de la IA?
  B. ¿Qué implica el concepto de régimen complejo en la gobernanza de la IA?

C. ¿Cuáles son los principales desafíos para armonizar los estándares internacionales de IA?
D. ¿Cómo se puede promover la seguridad colectiva ante la proliferación de la IA?
E. ¿Qué modelos institucionales se proponen para compartir los beneficios de la IA con países de bajos y medianos ingresos?

Ejercicio previo a la sesión 📝

- Investiga al menos una ley del país de tu preferencia y haz una breve revisión de la misma, sus alcances, limitaciones y formas de implementarla. No pases más de 20 min en esta actividad. Intenta hacer notas para compartir tus conocimientos con tus compañeros.

  NOTA: En la semana habrá un foro en donde se tendrá que publicar la que has seleccionado, intenta no repetir con tus compañeros. Puedes consultar fuentes como esta.

- Hacer revisión del foro y roles ofrecidos para el juego de roles de la sesión (Ej. Científico,)

Contenido Sugerido 💬

- **Recent U.S. efforts on AI Policy**
- **AI Index Report: Policy and Governance (2024)**
- **Managing extreme AI risks amid rapid progress**
- **Model Evaluations for Extreme Risks**
- **AI Safety.World**

👀Organizaciones y Páginas sugeridas

- **Epoch**

  Epoch AI es un instituto de investigación que investiga tendencias y preguntas clave que darán forma a la trayectoria y la gobernanza de la IA.

- **Centre of the Governance of AI**

Centro de investigación sobre las amenazas que los sistemas de IA de propósito general pueden representar para la seguridad. Buscan comprender los riesgos que plantean hoy, al mismo tiempo que miramos hacia los riesgos más extremos que podrían plantear en el futuro.

- **Blue Dot Impact: AI Governance.**
  Cursos diseñados para el desarrollo de marcos normativos que intervengan en desarrollar IA segura.

🗣️ Dinámica de Sesión

- ➔ Explicación de conceptos clave.
- ➔ Discusión sobre la mejor manera de gobernar IA (Ej. Instituciones nuevas, nuevas responsabilidades a existentes, etc).
- ➔ Juego de roles: representación de intereses de cada parte y acuerdos para gobernanza de IA.
- ➔ Presentación breve de ley de gobernanza consultada de IA
- ➔ Explicación de Ejercicio para sesión 6: Caminos profesionales.

📑 **Contenido de Sesión:** Ejercicios prácticos de sesión

Nota: Recuerda hacer una copia de estos templetes.

💻 IA – SESIÓN 6.

📲Discusión de Caminos Profesionales, ¿Qué sigue?

Objetivos de Sesión.

En esta sesión, contarás con las herramientas y referentes necesarios para identificar y proponer al menos tres caminos profesionales o proyectos enfocados en la mitigación de riesgos catastróficos globales asociados a Inteligencia Artificial aprovechando tu carrera profesional. Utilizarás el Modelo de Factores Ponderados

(WFM) y otros enfoques para evaluar y priorizar tus opciones, enmarcados en criterios de prioridades globales e impacto. Al finalizar, estarás equipado para tomar decisiones informadas y alineadas con el impacto que deseas generar en ésta área.

Contenido Obligatorio 📢

- **Aborda tu decisión de carrera estratégicamente (10 min)**
  La lectura explora el cómo tomar decisiones en tu carrera de forma estratégica para maximizar tu impacto en el mundo. Su contenido te introducirá al método "SELF" en la toma de decisiones profesionales analizando cada factor por separado.
  ***Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas*:**
  - ¿Qué es el método SELF y por qué es útil aplicarlo para enfocar nuestra carrera en causas prioritarias?
  - ¿A que se refiere el factor "Leverage" del método y cómo aplicarlo en la evaluación de puestos de trabajo?
  - ¿Qué es el factor "Fit" y cómo aplicarlo para identificar tus intereses y fortalezas?

- **Modelo de Factores Ponderados (15min)**
  El modelo de factores ponderados (WFM) es una herramienta utilizada para evaluar y comparar opciones basadas en varios criterios ponderados según su importancia. El proceso incluye generar una lista de criterios, asignarles pesos según su relevancia, y puntuar las alternativas bajo cada criterio. Esta herramienta funge como una herramienta valiosa para la toma de decisiones.

  Al finalizar esta lectura, deberás ser capaz de reconocer lo que es el modelo y aplicarlo efectivamente en el contexto de decisiones de vida profesional.

- **¿Cómo elegir una causa a la cual apoyar? (11 min)**
  El video muestra una perspectiva general de cómo seleccionar una causa para generar el mayor impacto posible. Su contenido te ayudará a comprender a qué nos referimos con "impacto" y qué consideraciones

debes tomar al comparar distintas causas.

Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:

- ¿Cuáles son los criterios a considerar al momento de comparar distintas causas?
- Además de los criterios de priorización, ¿qué otros parámetros debemos tomar en cuenta para impactar en una causa?
- ¿Cómo influyen parámetros como el dinero, el tiempo y el capital social en la priorización de causas?

- **AI governance and strategy: a list of research agendas and work that could be done.**
  Presenta diversas ideas de proyectos y la agenda en IA que pueden ser útiles como referentes al camino a desarrollar.

  Al finalizar esta lectura, deberás ser capaz de identificar y proponer ideas de proyectos en IA, entendiendo cómo estas iniciativas pueden apoyar a la reducción de los riesgos asociados a esta tecnología desde un ámbito global.

Ejercicio previo a la sesión 📝

4. Anota los **puntos más relevantes de las lecturas**, con el fin de discutirlos con tus compañeros durante la sesión.
   a. Enfócate en reflexionar cómo puedes aplicar los conceptos presentados en tú desarrollo profesional individual

5. Previo a la sesión, genera una **comparativa** de tus **posibles caminos profesionales** impactantes usando este templete. Tu facilitador se encargará de explicar los pasos a seguir al final de la sesión 5.
   a. Este ejercicio aplica los principios del modelo de factor ponderado. Te recomendamos revisar este recurso para

comprender más sobre el mismo.

b. Te recomendamos que revises los perfiles profesionales de 80,000 horas y Probably Good, que han sido curados por expertos en el área para determinar los perfiles profesionales con un mayor impacto en la actualidad.

6. Desarrolla un **plan de acción detallado** a mediano plazo con base en en este templete. Tu facilitador se encargará de explicar los pasos a seguir al final de la sesión 5.

Contenido Sugerido 💬

- **Ajuste personal (10 min)**
  Esta lectura complementa el contenido visto en la lectura anterior, enfocándose específicamente en cómo evaluar el factor "Fit" en tu carrera. Este contenido te permitirá identificar lo que haces bien y en lo que podrías mejorar al momento de elegir roles impactantes.
  Al finalizar esta lectura, deberás ser capaz de responder las siguientes preguntas:
  - ¿Cómo puedo tomar en cuenta mi motivación sin caer en el cliché de "encontrar mi pasión"?
  - ¿Por qué debo considerar mis límites y necesidades al momento de elegir un camino profesional?
  - ¿Cómo puedo identificar fortalezas y habilidades personales que pueden ser útiles en un puesto laboral?

- **Pregúntate qué pasaría si no hicieras este trabajo (12 min)**
  Esta lectura nos adentra a la comprensión del pensamiento contrafactual a partir de la reflexión de qué sucedería si no realizáramos determinada acción. Esperamos que su contenido te permita comparar roles e identificar aquellos donde podrías hacer una mayor diferencia.

👀Organizaciones y Páginas sugeridas

- **Base de Datos: Organizaciones.**
  Organizaciones clave para la bioseguridad, biocustodia, prevención y preparación para pandemias o áreas relacionadas.

- **Effective Thesis**
  Ideas de proyectos de investigación para desarrollo de tesis en RCGs o área considerada de "Alto Impacto" a nivel global.

🗣️ Dinámica de Sesión

- En esta sesión primero se presentará y explicará lo que son las herramientas de toma de decisión aplicadas a la vida profesional.
- Posteriormente se dará espacio a cada uno de presentar los posibles caminos profesionales a escoger, sus resultados y el plan de acción del más adecuado.
  Nota: Recuerda relacionar estos caminos a ponderar con las propuestas de proyectos que podrías presentar para el programa de mentoría.

📑 Contenido de Sesión: Ejercicios prácticos de sesión

Nota: Recuerda hacer una copia de estos templetes y **subirlo a la plataforma**.