



CHARLA · CARRERA EN AI SAFETY · 2026

# Seguridad en Inteligencia Artificial

Una carrera en AI Safety: una historia personal  
desde Buenos Aires

Eitan Sprejer · BAISH · [baish.com.ar](https://baish.com.ar)

QUIÉN LES HABLA

# Eitan Sprejer



● Fundador y director de BAISH

● Becario · UK AI Security Institute

● Facilitador · BlueDot Impact · 13 cohortes

● Research Project Manager Lead · Apart Research

● +80 career advising calls

Buenos Aires AI Safety Hub. Primera comunidad de AI Safety en Argentina. +300 miembros. Fundada en julio de 2024.

Licenciado en Física (UBA). Cofundador de LANAIS.

¿Estás pensando tu carrera en AI safety? Escíbeme y hablemos — me encanta dar una mano.

01

¿Por qué **AI Safety**?

La motivación, en breve.

02

Mi historia

De física en la UBA a BAISH y research. Cómo pasó.

03

Cómo construir **esta carrera**

Qué podés hacer vos.

SI TE LLEVÁS UNA SOLA COSA DE ESTA CHARLA

En *AI safety*, casi todo pasa por tu red de contactos.

Casi cada paso de mi carrera salió de una conexión. No se construye en aislamiento.

ACTO 1

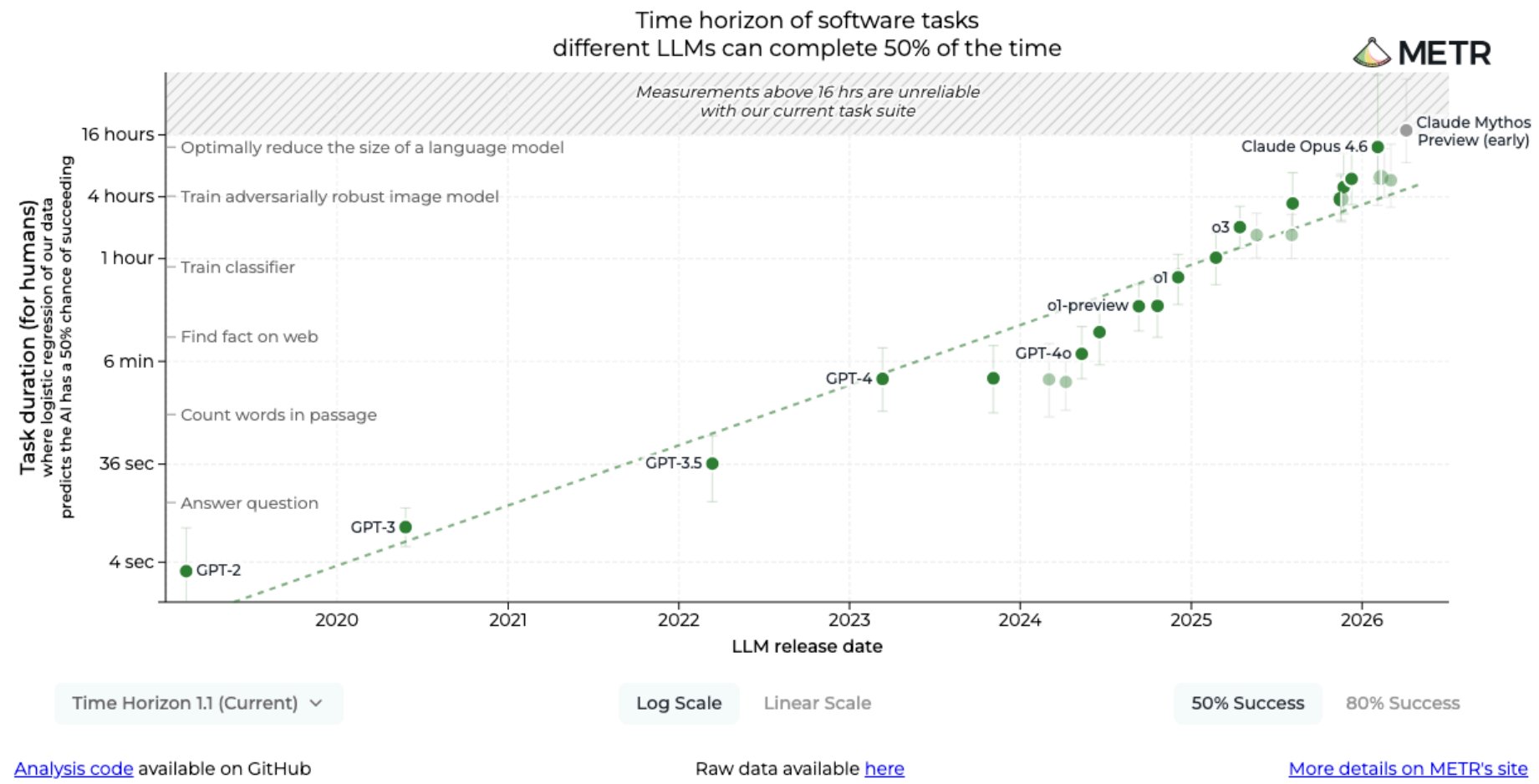
# ¿Por qué *AI Safety*?

¿Cuál es el problema más importante de tu campo?

.....

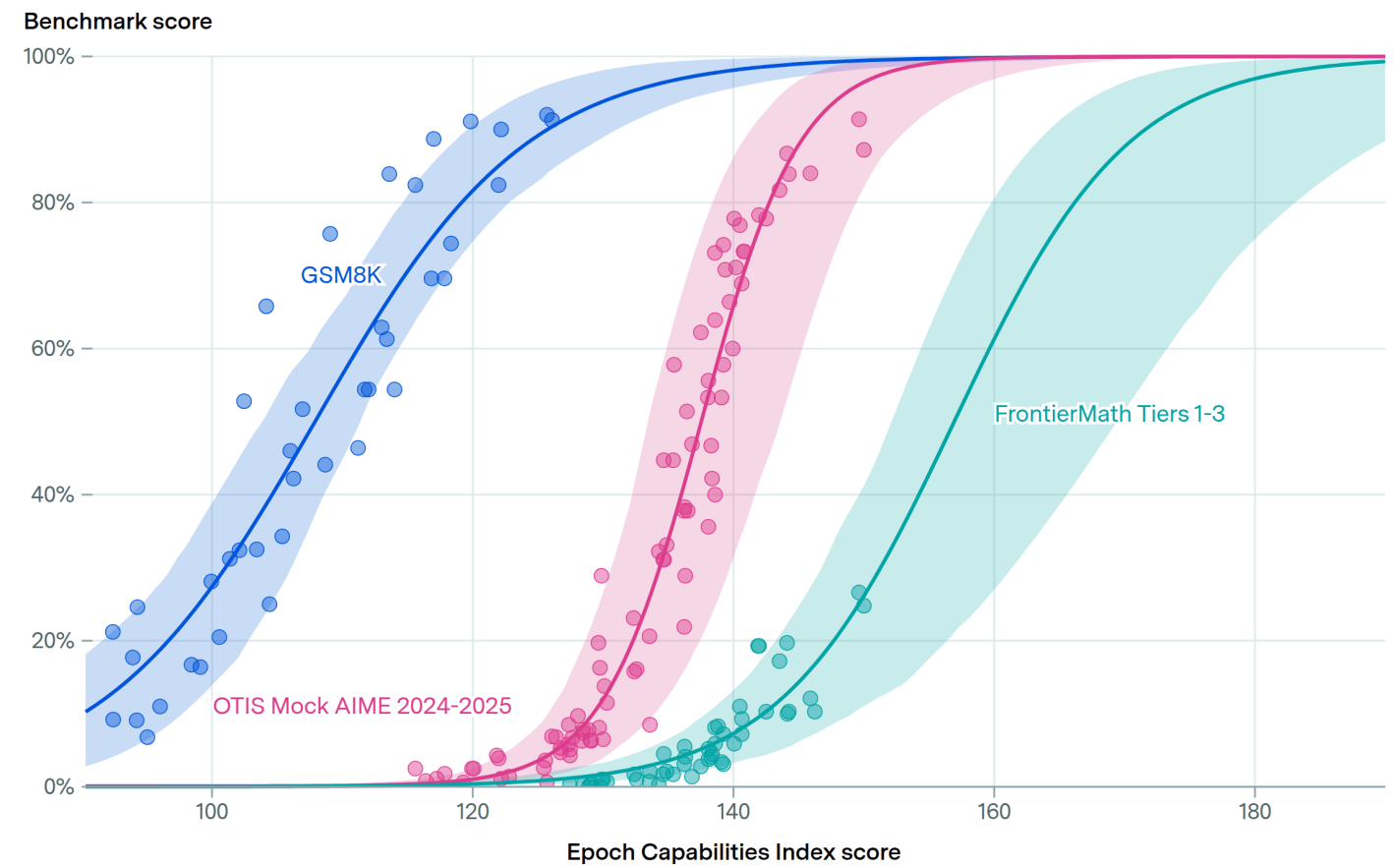
¿Por qué no estás trabajando en eso?

# Tareas más largas, cada pocos meses

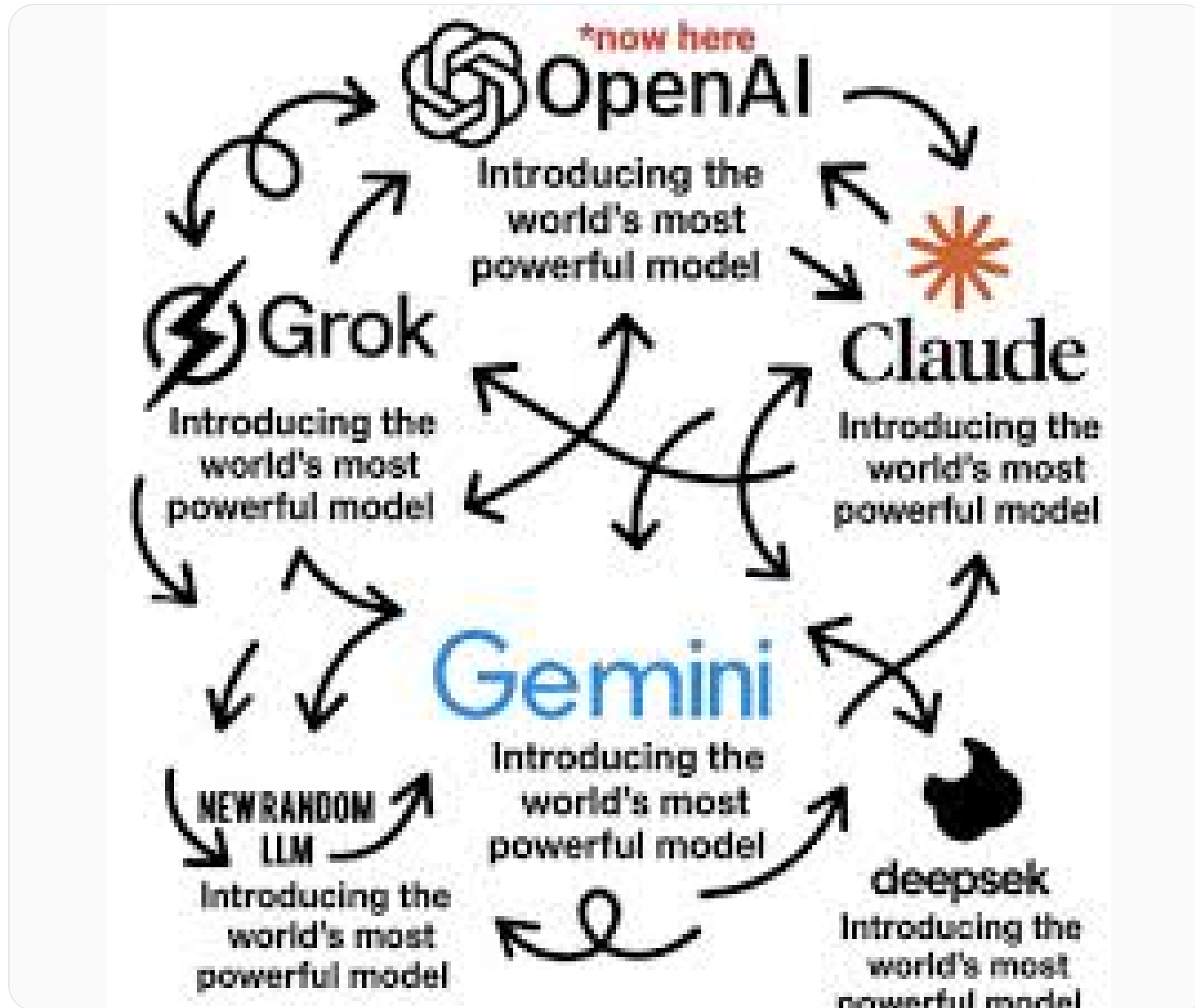


# Los benchmarks se saturan

Epoch's Capabilities Index stitches together benchmarks across a wide range of difficulties



# Por defecto, no va a salir bien



*"Safety culture and processes have taken a backseat to shiny products."*

Jan Leike, ex-líder Superalignment OpenAI, al renunciar. Mayo 2024.

# Statement on AI Risk

"Mitigar el riesgo de **extinción** por IA debería ser una prioridad global, junto a otros riesgos de escala social como las pandemias y la guerra nuclear."

Center for AI Safety, 2023. Firmado por Geoffrey Hinton, Yoshua Bengio, los CEOs de OpenAI / Anthropic / Google DeepMind, y premios Nobel y Turing.

The statement:

Mitigating the risk of **extinction from AI** should be a **global priority** alongside other societal-scale risks such as pandemics and nuclear war.

Signatories:

AI Scientists  Other Notable Figures

**Geoffrey Hinton**

Emeritus Professor of Computer Science, University of Toronto

**Yoshua Bengio**

Professor of Computer Science, U. Montreal / Mila

**Demis Hassabis**

CEO, Google DeepMind

**Sam Altman**

CEO, OpenAI

**Dario Amodei**

CEO, Anthropic

ACTO 2

# Mi historia

# ML4Good

Entré por un grupo de EA. Pero mi primer contacto *real* con AI safety fue un bootcamp de ML4Good.

[ml4good.org](https://ml4good.org)



# Fundé BAISH

Volví a Buenos Aires y arranqué BAISH como grupo universitario. Corrimos el primer curso introductorio. Julio 2024.



ACTO 2 · DÓNDE ESTÁ HOY

# BAISH hoy

**+300**

miembros

**+80**

alumni de cursos

**+10**

cohortes

**15**

equipo de voluntarios



# Cuatro task forces

## Comunidad

Scouting de talento. Networking, eventos sociales presenciales y 1:1s con personas e instituciones.

## Cursos

Technical AI Safety, Governance y Projects. Con financiamiento de BlueDot.

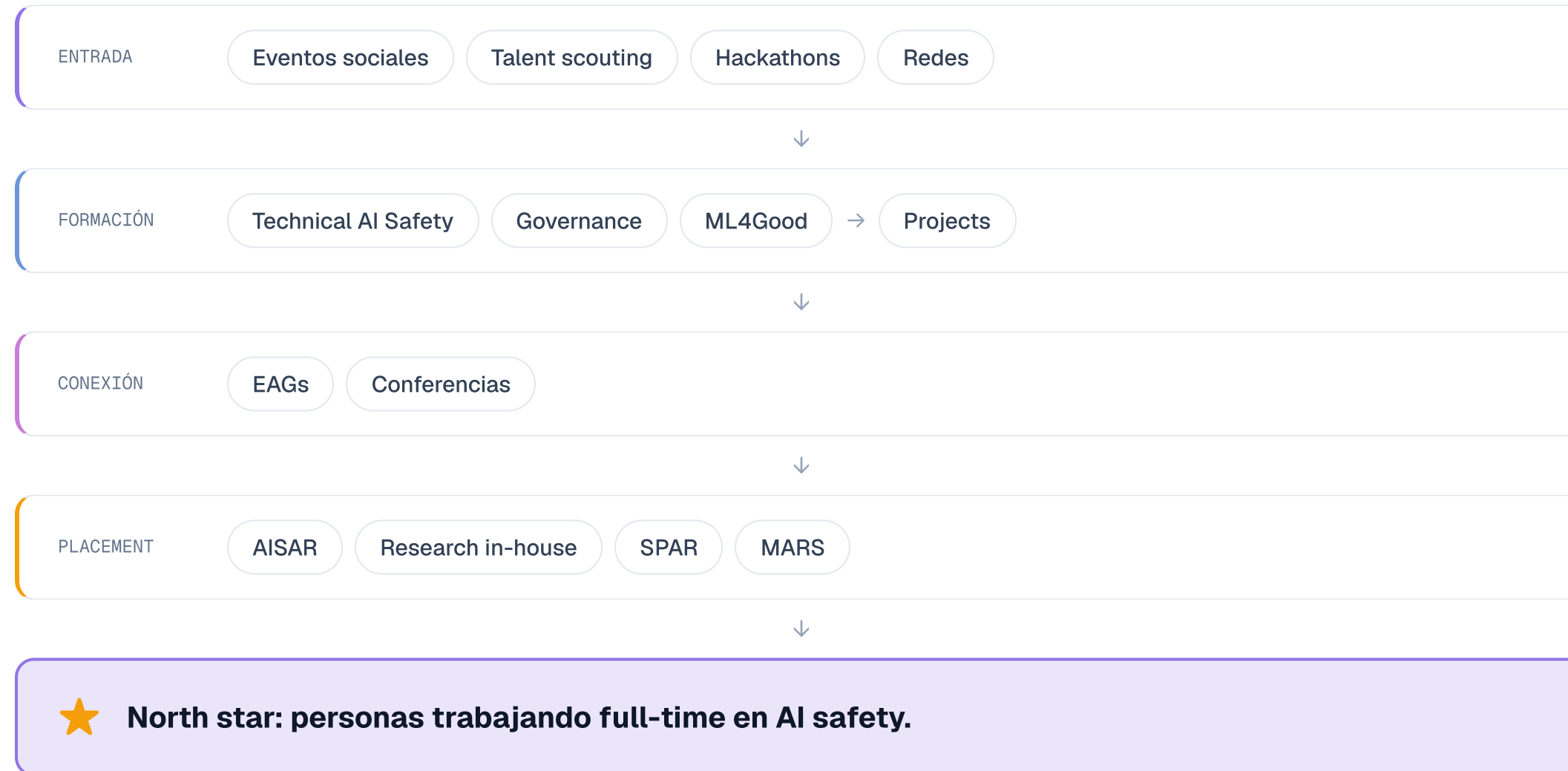
## Comunicaciones

Contenido y marca ([Instagram](#)). Divulgación seria de safety.

## Research

Armando una división de investigación propia. Y AISAR, nuestro programa de pasantías de research.

# Un pipeline de talento



# AISAR

Un **programa de research**: pasantías pagas para producir investigación real en AI safety.

6 → 15

scholars (1ª cohorte → 2ª)

3

papers aceptados

# Paso a paso — casi todo, una [conexión](#)

ML4Good

Mi primer contacto real.

BAISH

Lo fundé.

TA en ML4Good

Me contactaron ellos. [← conexión](#)

EAG London '25

Conocí a Austin → su proyecto MATS. [← conexión](#)

Facilitador de BlueDot

Por aplicar.

AISAR

Beca de research (2º semestre '25). [← referencias](#)

MARS → UK

Cambridge me llevó al Reino Unido.

Cursos de BlueDot en BAISH

Oficiales. De conocer a Josh en una conferencia en UK. [← conexión](#)

EAG SF '26

Dos oportunidades salieron de acá ↓

UK AISI

Co-grantee con Austin. [← conexión](#)

Apart

Conocí a Jaime. [← conexión](#)

# Mis co-directores, también

Todos terminamos con roles full-time en el ecosistema. En todos los casos, por una conexión.

Eitan                  Apart Research · UK AISI

Charly                 Redwood Research · Equistamp

Luca                    80,000 Hours

# Casi ningún paso pasó en aislamiento.

La lección: sé una parte activa de tu comunidad — y de la comunidad global. Eso implica ir a conferencias, meterte en los espacios donde pasa.

ACTO 3

# Cómo construir esta carrera

# Cuatro campos

## Técnico

Alignment, interpretabilidad, control, evals.

## Governance

Política, regulación, coordinación entre actores.

## Estrategia

Vista de pájaro: qué priorizar, dónde invertir tiempo y plata.

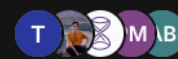

## Field-building

Creer la capacidad del campo. La palanca meta — **acá hay muchísimo por hacer.**

# Necesitamos más generalistas

El campo invierte muchísimo en capacidad técnica, pero poco en generalistas: gente que ejecuta, conecta y hace estrategia. Ahí hay impacto enorme por tomar.

## AI Safety's Biggest Talent Gap Isn't Researchers. It's Generalists.

 Topaz, Agustin Covarrubias , Kairos, Parv Mahajan, Alexandra Bates  
7 min read · Apr 13

180

25

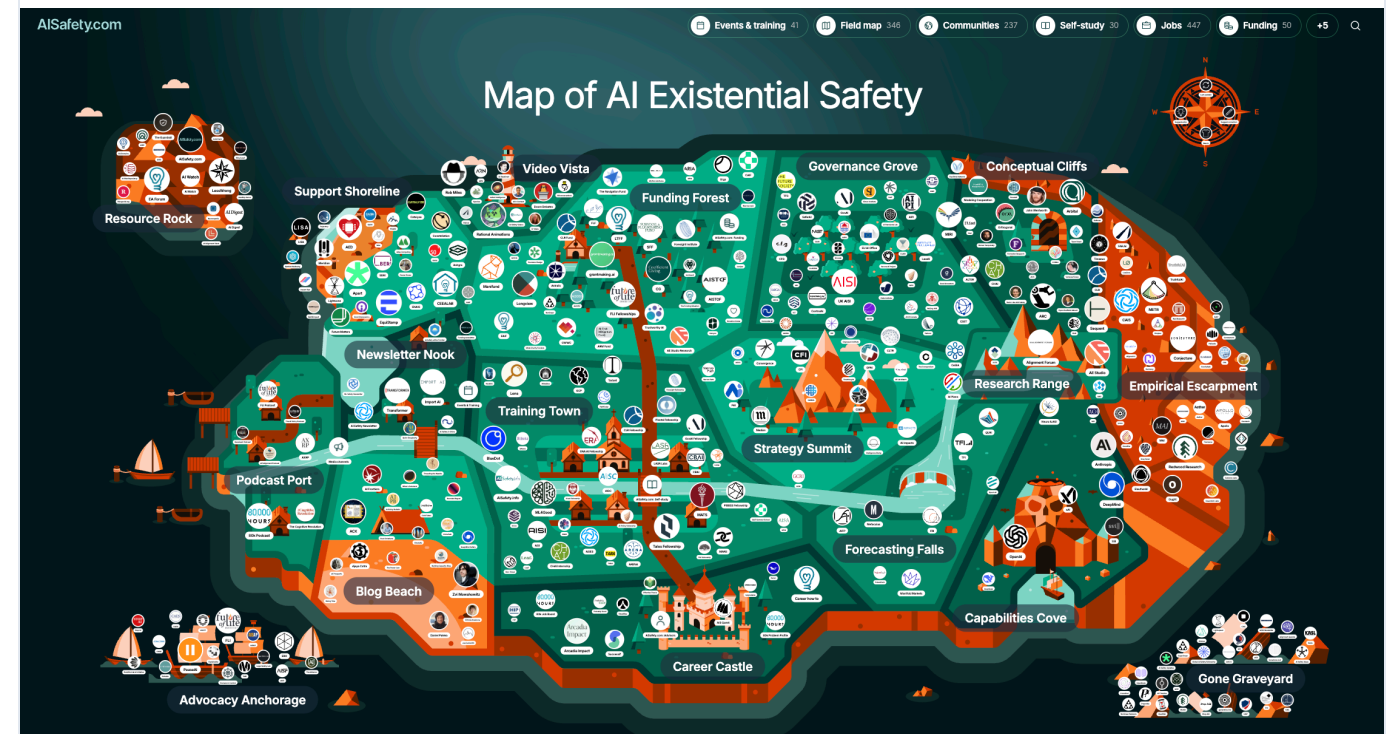
AI safety Career choice Opportunities to take action Announcements and updates 5 more

This post was cross posted to [LessWrong](#)

**TL;DR:** One of the largest talent gaps in AI safety is competent generalists: program managers, fieldbuilders, operators, org leaders, chiefs of staff, founders. Ambitious, competent junior people could develop the skills to fill these roles, but there are no good pathways for them to gain skills, experience, and credentials. Instead, they're incentivized to pursue legible technical and policy fellowships and then become full-time researchers, even if that's not a good fit for their skills. The ecosystem needs to make generalist careers more legible and accessible.

**Kairos** and **Constellation** are announcing the **Generator Residency** as a first step. Apply [here](#) by April 27.

Epistemic status: Fairly confident, based on 2 years running AI safety talent programs, direct hiring experience, and conversations with ~30 senior org leaders across the ecosystem in the past 6 months.



# Por dónde empezar

Un curso de BlueDot es el punto de partida más claro. Pero hay muchas formas — y más gente debería aprovecharlas.

- **Un curso de BlueDot Impact.** El punto de partida más claro. [bluedot.org](https://bluedot.org)
- **Un hackathon de Apart Research.** Producís research real en un fin de semana. [apartresearch.com](https://apartresearch.com)
- **Eventos de tu comunidad local.** Y si no existe una, armá los tuyos.
- **Eventos virtuales** , como este.
- **Pedí 1:1s** a gente que admirás. Mucha gente dice que sí.
- **Andá a conferencias** (EAG) y aplicá a cursos y oportunidades (ej. AISAR).

# No lo hagas en aislamiento

- **Animate a hacer cosas.** Si ves una necesidad en el campo, proponé una solución — eso trae muchos frutos. (Así nació BAISH.)
- **Hacelo con más gente.** Conectá, ofrecé ayuda, armá comunidad.
- **Mostrá lo que hacés.** Publicá tu research, tené tu web, subí blog posts.
- **Sé un conector.** Faltan en safety — y es de lo más valioso que podés aportar.



Preguntas · Gracias

[baish.com.ar](http://baish.com.ar) · [eitan@baish.com.ar](mailto:eitan@baish.com.ar)